

PITTSBURGH SUPERCOMPUTING CENTER

300 S. CRAIG STREET

PITTSBURGH, PENNSYLVANIA 15213



PITTSBURGH
SUPERCOMPUTING
CENTER

PEOPLE. SCIENCE. COLLABORATION.

FALL 2015

PITTSBURGH SUPERCOMPUTING CENTER

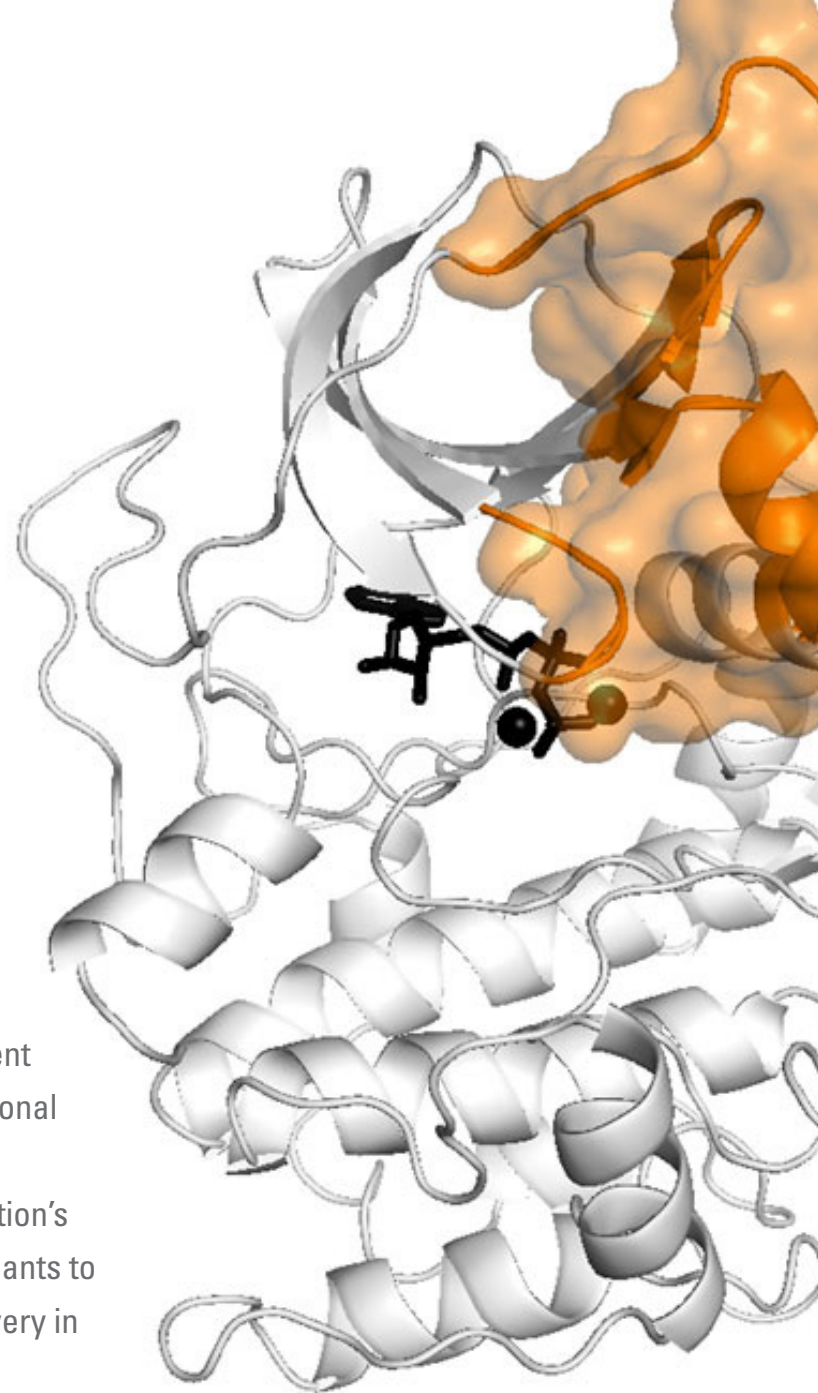
PEOPLE. SCIENCE. COLLABORATION.

FALL 2015

PEOPLE. SCIENCE. COLLABORATION. | FALL 2015

PSC.EDU

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.



FROM THE DIRECTORS

Welcome to the Fall 2015 edition of People. Science. Collaboration.

This summer we marked a significant transition at PSC: We retired the Blacklight system, and in October will begin constructing the new, National Science Foundation (NSF)-funded Bridges system (p. 12). In the interim, we have replaced Blacklight in the Data Exacell (DXC) accelerated development pilot project with the Greenfield system, to serve the needs of our users until Bridges goes into production.

Bridges, however, is about much more than sustaining the scientific communities that have found Blacklight's data-oriented architecture to be so useful for their work. Bridges will incorporate lessons learned from Blacklight and the DXC/Data Supercell projects to serve an expanding community of researchers, both traditional and new to HPC, requiring very large memory, sophisticated analytics and, above all, an unsurpassed level of flexibility to support the needs of "Big Data" research projects. We're proud, once again, to be leading in the effort to bring HPC to a wider circle of investigators.



Successful projects—among the last performed on the NSF-funded Blacklight—include building a virtual poker player that took on four of the best Texas Hold'Em players in the world (p. 4); testing the myriad of decision-making theories to better understand how people make choices (p. 8); and assembling and identifying the active genes in the economically important whiteleg shrimp, the number one human food source from the sea (p. 18). And the D.E. Shaw Research Anton supercomputer hosted at PSC has once again helped overturn preconceptions about the function of complex biomolecules, giving a new perspective on an enzyme that we'd thought to be "understood" (p. 14).

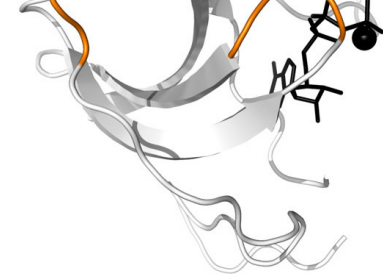
In addition to serving users across the nation and the world, PSC continues to pursue independent research. Our Public Health and Networking groups have reported success in uncovering advantages to increasing the size of vaccine vials in West Africa (p. 23) and have won renewed funding to improve network performance (p. 22), respectively. Our participation in the National Center for Multiscale Modeling of Biological Systems

is generating exciting results regarding the subcellular neurostructure of the brain in multiple species (stay tuned for our upcoming Spring 2016 issue). And our largest research effort, the DXC, continues to expand its capabilities (p. 20).

Membership in XSEDE continues to be of great importance to PSC, with PSC staff providing leadership in user support and collaboration, networking and elsewhere. As a resource within the NSF's XSEDE network of HPC sites, Bridges will serve an even broader community than Blacklight's.

We would like to acknowledge all our funders, especially the NSF, the NIH and the Commonwealth of Pennsylvania (p. 10). We would like to hear any feedback you have, on our work or this publication. You can send any comments or suggestions via our feedback page at psc.edu/index.php/contact/607. You can also contribute to PSC's nonprofit, academic mission at psc.edu/donate.

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



4 LIFE IS NOT CHESS

♠♥ ♦♣ ARTIFICIAL INTELLIGENCE / MACHINE LEARNING

8 Understanding CHOICE

 DECISION MAKING / PSYCHOLOGY

10 PSC'S IMPACT ON PENNSYLVANIA





12 BRIDGE TO THE
FUTURE


DATA ANALYTICS

A SHIFT
IN PER
SPEC
14 TIVE

 BIOCHEMISTRY/PHARMACOLOGY

20 PSC AND ITS PARTNERS
IN BRIEF

DXC EXPANDS SCOPE
ANSYS PARTNERSHIP RENEWED
2015 MARC INTERNS COMPLETE SUMMER PROJECTS
PSC BUILDING EARLY WARNING TOOL FOR FIXING INTERNET TRAFFIC JAMS
CHANGING VACCINE VIAL SIZE CAN IMPROVE SUPPLY: PSC/HOPKINS STUDY

24 PSC RESOURCES FOR DATA-DRIVEN SCIENCE

18 SECURING
THE CHAIN

 GENOMICS/MARICULTURE

LIFE IS NOT CHESS



COUNTED AS A LOSS OR A
“STATISTICAL TIE,” POKER MATCHUP
POINTS TO BETTER DECISION MAKING

ARTIFICIAL INTELLIGENCE / MACHINE LEARNING



Life is not a chess game.

The world is not a chessboard, with every piece visible. Life is more like a hand of poker, according to Tuomas Sandholm of Carnegie Mellon University's School of Computer Science. Other players have cards we can't see and try to trick us. Could our decisions be better if we leveraged artificial intelligence?

With graduate students Noam Brown and Sam Ganzfried, Sandholm has used PSC's Blacklight to build an artificial poker player. This program, Claudico, recently lost a squeaker of a tournament against top human poker players. Lessons learned from that matchup promise to transform how we navigate a world with adversaries and incomplete information.

WEIGHTING DECISIONS

Sandholm is quick to say his team did not actually write a poker-playing program. "We really didn't write a program called 'Claudico,'" he says. "The algorithms we've developed for solving incomplete information games are general-purpose." Claudico emerged from those algorithms as output, given only the rules of the game as input.

The algorithms try to approximate "game-theory-optimal" play, subject to computational limitations. Heads-up, no-limit Texas Hold'em poker, in which players can bid as much as they like on a given hand, contains some 10^{161} situations (called information sets). That's far more than there are atoms in the Universe and way beyond any foreseeable computing capability.

"The first step is creating an abstraction," Sandholm explains. "The algorithm takes the rules of the game and outputs a smaller game that's strategically similar." The algorithm treats similar hands as

identical; for example, possibly equating two Jacks with two Queens. But as the game progresses, this "rounding off" error amplifies. The CMU researchers countered this problem using Blacklight, whose large cache-coherent memory allowed a much finer-grained abstract than otherwise possible. In computing the strategy for Claudico, the researchers routinely used an enormous eight Terabytes of RAM—4,000 times as much as in a top-line computer tablet.

The team's earlier AI, Tartanian7, dominated the 2014 Annual Computer Poker Competition. Claudico, by comparison, performs an even finer-grained abstraction and uses two different abstractions, depending on whether it is the first or the second mover in the game. Also, its algorithm runs on Blacklight 24/7, in parallel with its lighter, real-time thinking, which takes place on a commodity server.

Claudico could beat the tar out of Tartanian7. But was it ready, like Deep Blue in chess in 1996 or Watson in *Jeopardy!* in 2011, to battle humanity's best?

“PLAYING A MARTIAN”

Claudico comes up with some strategies that humans find to be downright alien.

“Playing Claudico is like playing a Martian,” Sandholm says. In particular, it likes to “pass” on the first move—meet the other player’s bid without raising it or folding. Pros denigrate that as a rookie move, calling it “limping.” But Claudico owned that strategy— its name is Latin for “I limp.”

“Humans learn how to play poker in two ways,” Sandholm notes. “One is that they play a lot of poker against other humans; the other is that they read books about how to play poker—but who wrote those books?” Possibly, the researchers thought, humans had evolved into a point in the strategy space that Claudico could beat.

At the Rivers Casino in Pittsburgh from April 24 through May 8, 2015, and after 80,000 hands against four of the top-10 ranked poker players in the world, Claudico fell a little short of that goal. But it came close. Out of some 170 million virtual dollars wagered, Claudico wound up just \$732,713 behind the humans. Statistically, the contest was a tie.

“Claudico was a very strong opponent,” said Doug Polk, one of the contestants and the number-one ranked poker player in the world. “It’s extremely aggressive.”

Polk noted that at times Claudico would risk tens of thousands of dollars to win hundreds. He said that human players often hesitate to take such risks, and pros succeed by being willing to lose in the short term with a strategy that will pay off in the long run.

“Claudico just takes that to the next level,” he adds.

NOT JUST A POKER PLAYER

Ultimately, Claudico is about more than poker. Sandholm’s group has conducted AI research in several incomplete-information fields:

- The FCC periodically runs “spectrum auctions,” in which tens of billions of dollars can be bid for little-used radio frequencies. An AI could help participants bid more rationally.
- Allocation of air marshals, detection dog teams, and other resources might be more efficient if an AI could weigh the odds of a terror threat at a given place and time.
- A Claudico-like AI could help avoid committing cybersecurity resources at moments when hacking attempts are unlikely.



Pictured top to bottom: Poker experts Dong Kim, Doug Polk, Bjorn Li and Jason Les

Medical decision making may represent the most exciting opportunity, Sandholm says. “Most medical treatment today is myopic,” he explains. “We throw one treatment at a problem at a time.” An incomplete-information game solving AI might help doctors design multi-step treatment plans with better outcomes.

A series of pharmaceutical “nudges,” for example, might steer an HIV infection to a less life-threatening state. Similarly, therapies could push cancer-cell populations toward less malignancy or populations of bacteria away from antibiotic resistance.

“It’s not specific to any particular disease,” much less any particular “game,” Sandholm adds. “That’s a big vision and I’m very excited about it.”



PSC staff helped Tuomas Sandholm’s work with the support of programs within NSF’s XSEDE network of supercomputing centers: Extended Collaborative Support Service; Novel and Innovative Projects Program.



BRAINS VS. ARTIFICIAL INTELLIGENCE

Be sure to tweet @WinBigRivers and @SCSatCMU using #BrainsvsAI



APRIL 24 - MAY 8 | 11AM - 10PM

80,000 HANDS NO-LIMIT HOLD 'EM

Each hand starts with each player having 200 big blinds. One big blind is \$100, and one small blind is \$50.

Hands Dealt: 80,000/80,000

BRAINS : \$732,713

CLAUDICO : (\$732,713)

DOUGLAS POLK : \$213,671

CLAUDICO : (\$213,671)

BJORN LI : \$529,033

CLAUDICO : (\$529,033)

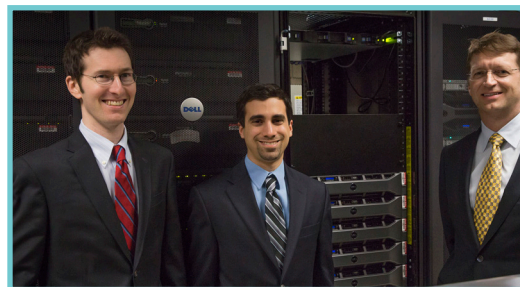
DONG KIM : \$70,491

CLAUDICO : (\$70,491)

JASON LES : (\$80,482)

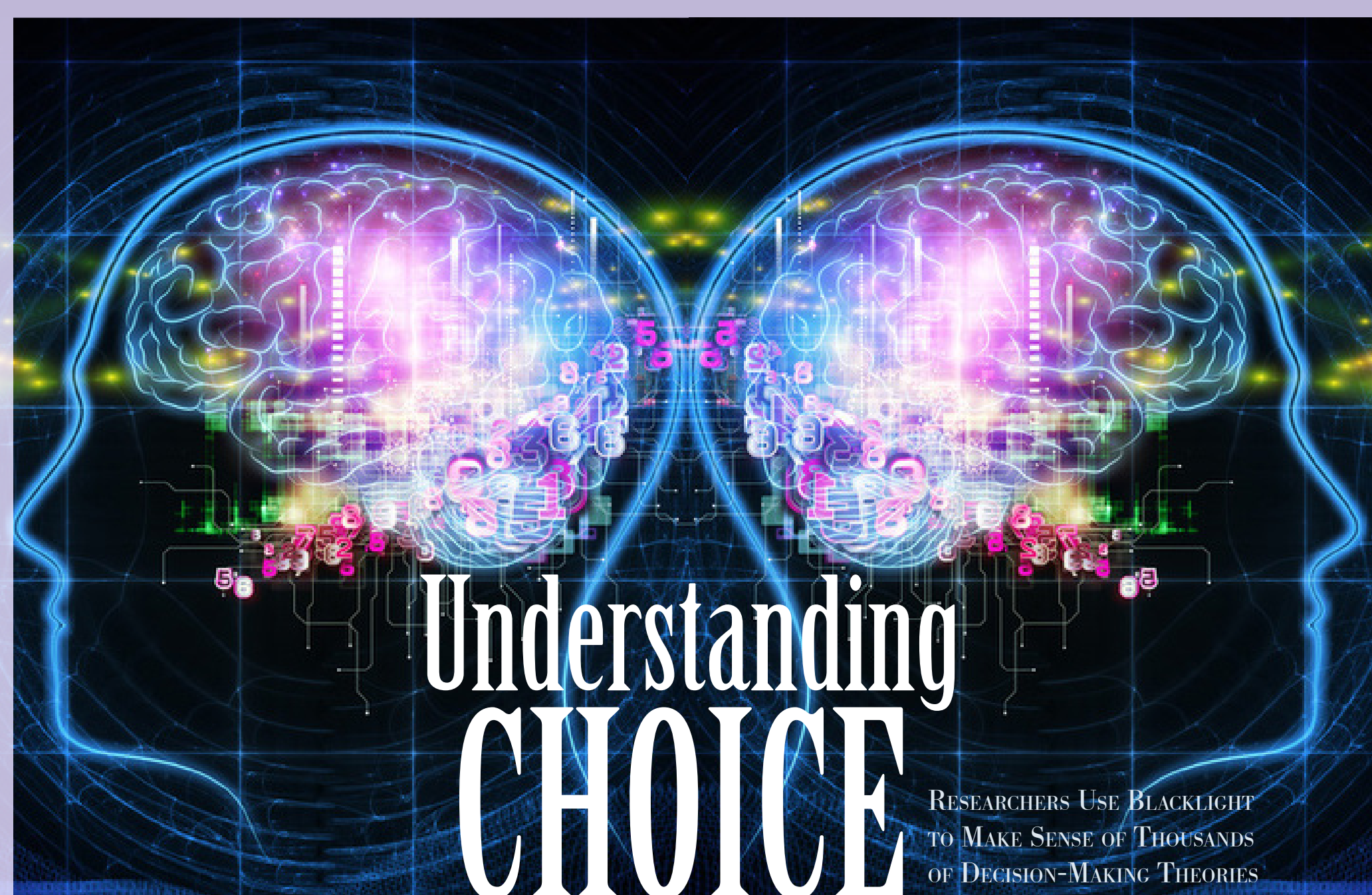
CLAUDICO : \$80,482

Parentheses indicate a negative number.



In the final tally, Claudico wound up a little over \$700,000 behind the humans. But it was a “statistical tie.”

Pictured Sam Ganzfried (l), Roan Boreen, and Tuomas Sandholm have created a set of rules for decoding “incomplete information” games.



Understanding CHOICE

RESEARCHERS USE BLACKLIGHT
TO MAKE SENSE OF THOUSANDS
OF DECISION-MAKING THEORIES



PSC staff helped Michel Regenwetter's work with the support of programs within NSF's XSEDE network of supercomputing centers: Extended Collaborative Support Service; Novel and Innovative Projects Program.



DECISION MAKING / PSYCHOLOGY

WHY IT'S IMPORTANT

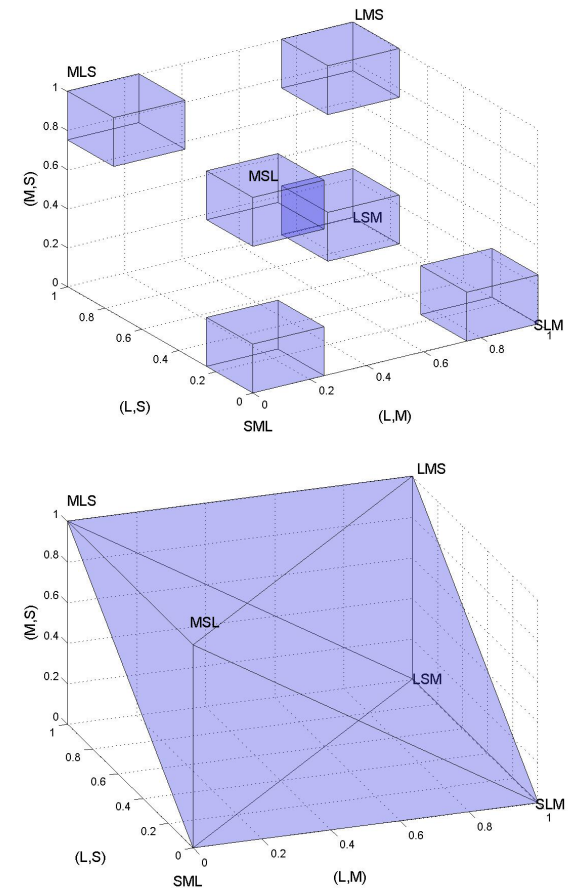
Understanding how human beings make decisions is critical in fields like cybersecurity, public health, elections and governance, and economics. How often do people make rational choices, weighing all the options? How often do they use mental shortcuts, short-circuiting good choices? Do our choices vary because we make mistakes, or are our minds split between different priorities? The answers to these questions matter whether we're asking why people don't choose healthy lifestyles, why they don't come out to vote in elections or why they end-run security measures. The field suffers in particular from a large number of theoretical models that have not been sufficiently tested.

“There are basically two reasons we need supercomputing. One is that there are a lot of theories out there that we need to test, and most don't actually explain variability of behavior. The other is that the statistical methods we use are very complex, and so computationally expensive. We're scaling up the speed at which the research is being done by two to three orders of magnitude.”

—Michel Regenwetter, University of Illinois at Urbana-Champaign

HOW PSC HELPED

Michel Regenwetter and colleagues at the University of Illinois at Urbana-Champaign believe that understanding decision making will mean statistically testing the different theories against real people making decisions—and against each other. With help from PSC staff, and utilizing PSC's Blacklight system, the group was able to test tens of thousands of theories against the data at once—thousands more than the largest previous analyses. With new colleagues in the neurosciences, Regenwetter plans to test whether people who tend to make rational decisions have different patterns of brain activity than those who make decisions by shortcut.



In one possible explanation for decision making (top), the decider has a fixed preference between three choices (L, M and S) but makes mistakes. Each shaded box represents a different preference between the three, its size determined by the decider's rate of error. In another explanation (bottom), the decider's preference isn't so set, and can change, so that the possible choices smear out to include a much larger possible set of choices. Data from people making decisions in experiments can help researchers compare the two explanations' ability to predict those decisions.

PSC'S IMPACT ON PENNSYLVANIA

PSC provides a significant economic benefit to the Commonwealth of Pennsylvania. In addition to supporting research, education and business activities throughout the Commonwealth, PSC consistently returns many times the state's investment in federal funding, economic stimulation and workforce development.

ECONOMIC IMPACT*



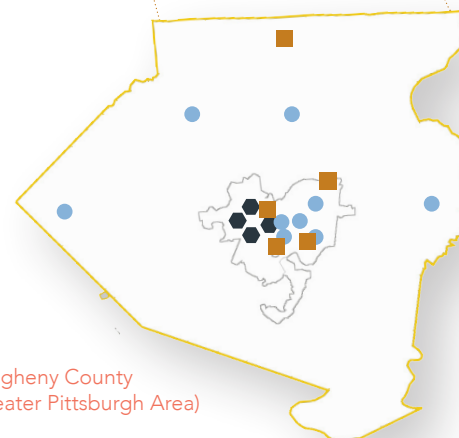
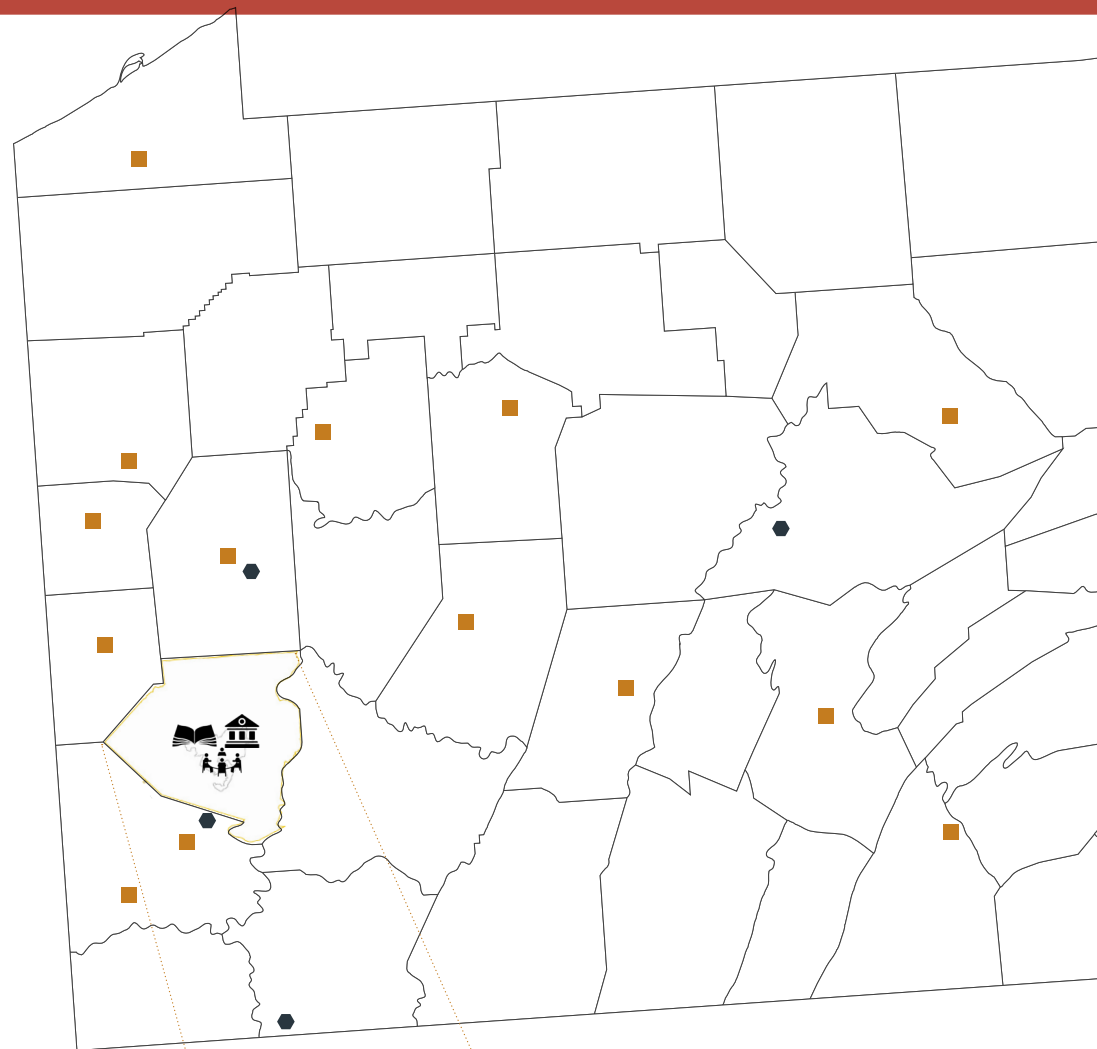
\$1M

State investment in PSC

\$60M

Federal funding for PSC users (\$30M)
and PSC/Research partners (\$30M)

*24 months, July 2013 – June 2015



Allegheny County
(Greater Pittsburgh Area)

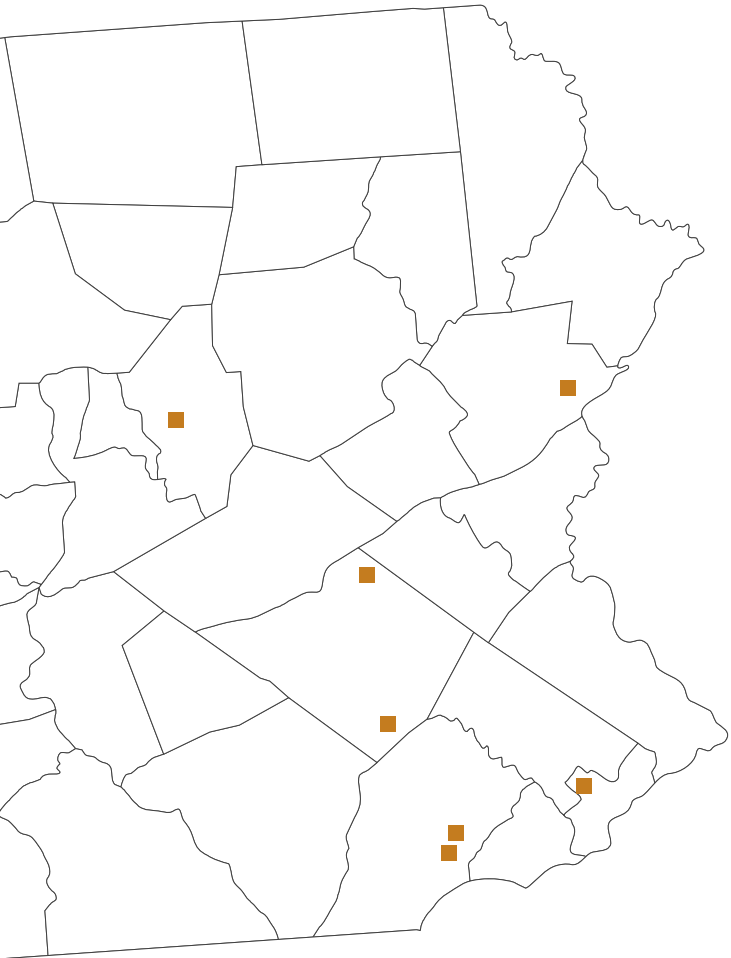


PSC BEST Program provides classroom material or full bioinformatics curricula to prepare high school students for 21st-century careers.

- Central Catholic High School
- Ellwood City Area School District
- Fort Cherry School District
- Frazier School District
- Our Lady of The Sacred Heart High School
- Oakland Catholic High School
- PA Cyber Charter Schools
- Pittsburgh Public Schools
including SciTech Academy
- Plum Borough School District
- Winchester-Thurston High School

PITTSBURGH SUPERCOMPUTING CENTER

RESEARCH PARTNERS AND USERS OF PSC RESOURCES



27 INSTITUTIONS OF HIGHER EDUCATION

Bloomsburg University
 Bryn Mawr College
 Bucknell University
 Carlow College*
 Carnegie Mellon University
 Drexel University
 Duquesne University
 Edinboro University
 Geneva College*
 Grove City College
 Hartford Graduate Center
 Indiana University
 Juniata College*
 LaRoche College

Lehigh University
 The Pennsylvania State University
 Philadelphia University
 Point Park University*
 St. Francis University*
 Swarthmore College
 Temple University
 University of Pennsylvania
 University of Pittsburgh
 Washington and Jefferson College*
 West Virginia University
 Westminster College*
 Wilkes University
**3ROX members only*

13 COMPANIES AND UNIVERSITIES

Ansys Inc.
 Bechtel Bettis
 Carnegie Science Center
 Children's Hospital (PGH)
 KINBER
 PNC Financial Corporation
 SpiralGen Inc.
 UPMC
 Westinghouse Electric Company

OTHER

Allegheny Intermediate Unit
 Beaver Valley Intermediate Unit
 National Energy Technology Laboratory
 West Virginia Network (WVNET)

PA STEM WORKFORCE

75

Current PSC employees



230

560

Employed by
 PSC since 1986

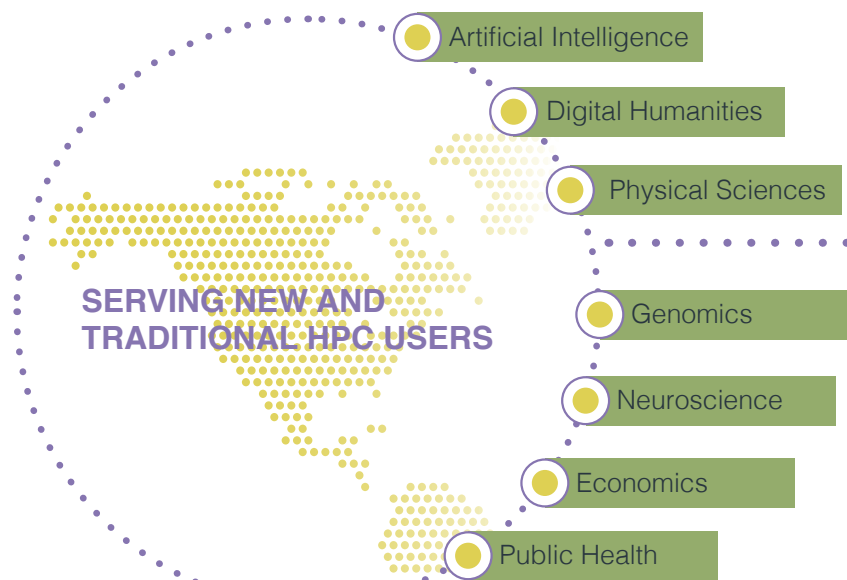
PSC interns

Since its inception, PSC has seen past employees and interns go on to work at some of Pennsylvania's most prestigious companies and universities, such as:

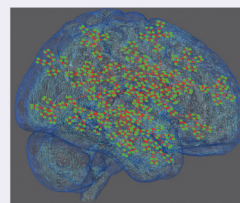
Carnegie Mellon University
 University of Pennsylvania
 University of Pittsburgh
 IBM
 Medrad
 Respironics
 Vocollect

Westinghouse
 Bank of New York Mellon
 Comcast
 Giant Eagle
 Google
 NetApp
 PNC

BRIDGE TO THE FUTURE

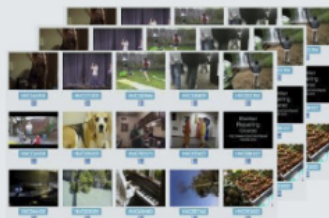


MEETING RESEARCHERS' NEEDS



- Expand research beyond the laptop
- Bring the power of HPC to Big Data
- Apply the power of HPC without the programming
- Enable rigorous engineering design optimization
- Support popular languages and environments such as Python, R, Hadoop, and Spark
- Accommodate a wide range of popular software

POTENTIAL EXAMPLES



- Large-scale genome assembly
- Financial market policy and analysis
- Social media trend analysis
- Agent-based disease spread modeling
- Diverse literary source analysis

Bridges Will Bring High-Performance Computing and Data Analysis to New Fields

Early next year, PSC's newest supercomputer, Bridges, will go online. The \$9.65-million, NSF-funded Bridges represents the culmination of lessons learned from PSC's Blacklight supercomputer and the Data Exacell pilot project. Bridges will offer unprecedented flexibility along with vastly more power.

"We designed Bridges to benefit new communities and transform research by integrating high-performance computing with Big Data and introducing new features for ease of use."

—Nick Nystrom,
Bridges Principal Investigator
and Project Director, PSC

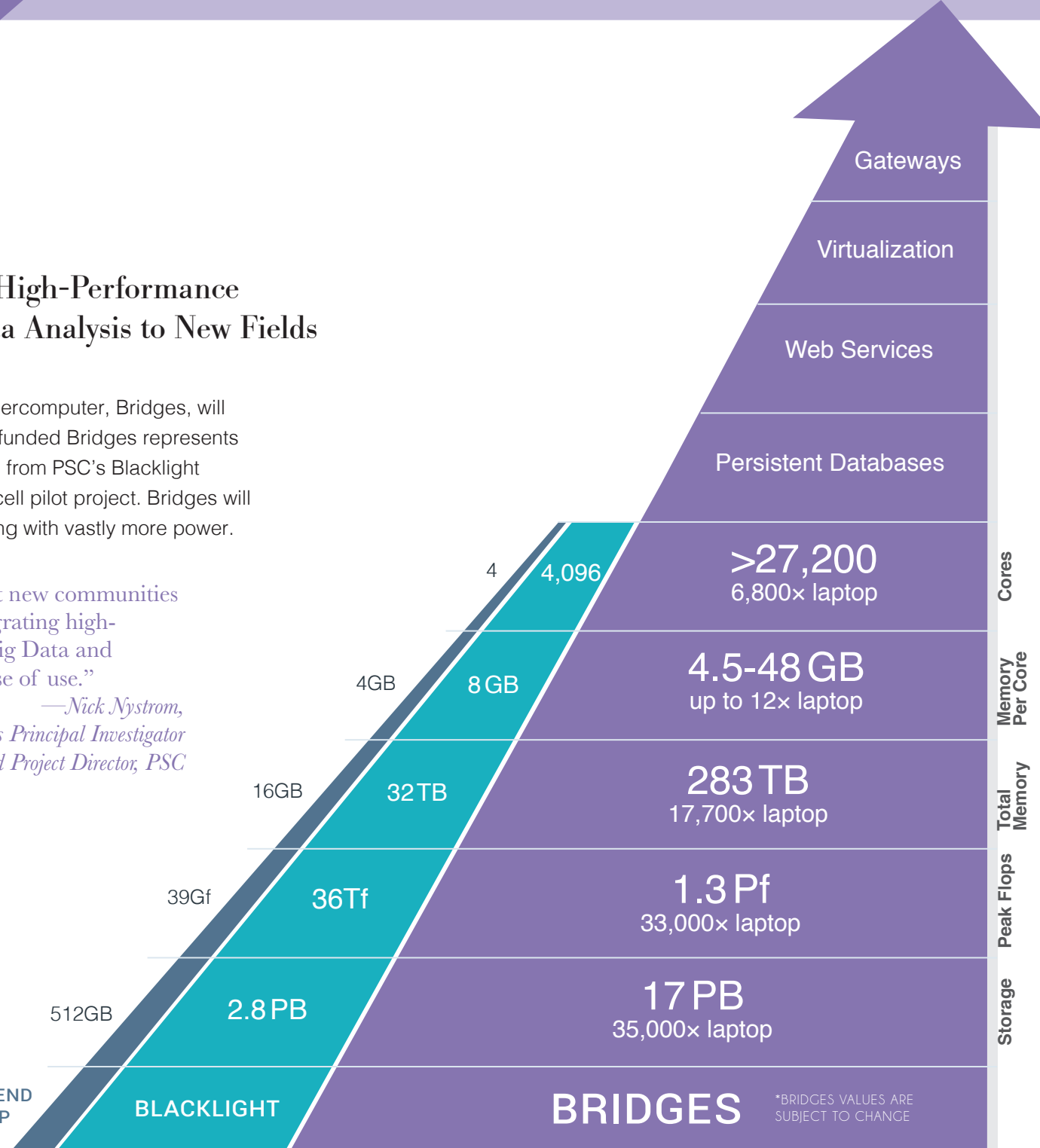


HIGH-END
LAPTOP

BLACKLIGHT

BRIDGES

*BRIDGES VALUES ARE
SUBJECT TO CHANGE





A SHIFT IN PERSPECTIVE

Anton Simulations Upend Picture of How Proteins Work

 BIOCHEMISTRY/PHARMACOLOGY

We've all seen the optical illusion: At first, it's clearly a picture of a white vase against a black background. But then your perspective changes and it becomes two black faces against a white background. In an instant, your understanding of the image is upended.

Using simulations on the Anton supercomputer at PSC, researchers at the University of California, San Diego (UCSD), have had a similar experience that may upend how scientists understand protein structure. The research may open new opportunities for correcting faulty protein function in diseases as diverse as cancer, diabetes, neurological disease, inflammatory disease and more.

Researchers led by Susan Taylor at UCSD have simulated the motion of protein kinase A (PKA). Their simulations, reported in the *Proceedings of the National Academy of Sciences USA*, revealed how the seemingly obvious sub-structures of the protein—“domains”—did not play the central role in function and movement that researchers had expected.

Rather, “communities” of PKA’s amino acid components that *do* move and act together have their own boundaries that don’t seem to relate to the ones researchers’ eyes had defined. This finding questions whether domains are actually relevant to the protein’s function and suggests that, to understand how proteins work, movement is as important as structure.

THE VERY MODEL OF A PROTEIN KINASE

The UCSD researchers chose PKA in part because previous research had provided a great amount of information about the protein, including its function and its three-dimensional structure. PKA is also medically important, acting in blood sugar regulation and diabetes, and has also served as a model for a family of signaling proteins called protein kinases that touch virtually every biological function in health and disease.

Protein structure, researchers once assumed, dictated function in a very straightforward way. Using X-ray crystallography and other methods,

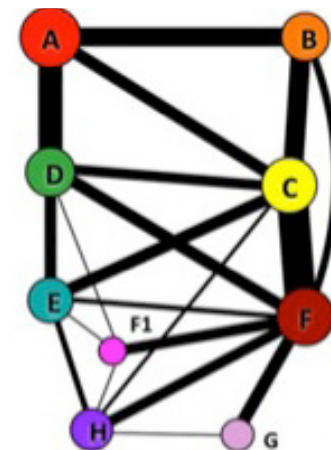
The Anton simulations suggest that PKA has nine distinct communities (A through H). The size of the circles shows how big each community is physically, the width of the lines connecting communities show how strongly each one’s movements affect another.

researchers had produced static “snapshots” of proteins that show how a chain of amino acids folds to make obvious smaller-scale structures—domains—that then form the larger protein. The assumption was that these domains would carry out parts of the protein’s functions.

In addition to understanding how domains worked, researchers wanted to understand a related phenomenon called *allostery*. When researchers change some specific amino acid in the chain of amino acids that makes up PKA, it could turn off an action carried out on the opposite side of the protein.

Somehow the change was transmitted across the protein, and researchers suspected that the domains somehow communicated with each other. The real puzzle was how these changes often made no visible difference in the protein’s structure.

“I mean, the kinase activity is broken in these mutants,” says Alexandr Kornev, a project scientist at UCSD and collaborator in the work. “You have the structure, it



looks the same, but it behaves completely differently.”

The difference had to be in how these very similar structures moved differently. The Anton supercomputer at PSC, the UCSD group reasoned, might help shed light on the mystery.

ANTON: MICROSECOND SIMULATIONS POINT TO NEW PROTEIN ORGANIZATION

Anton is a special-purpose supercomputer hard-wired to simulate the movements of large molecules using molecular dynamics. Unlike general-purpose supercomputers, Anton simulates large molecules at rates of microseconds per day. The Anton machine hosted at PSC was developed and made available without cost by D. E. Shaw Research.

“On Anton, our simulation took about a day of full-time use of the machine,” says Christopher McClendon, a postdoctoral fellow in Taylor’s laboratory and first author in the study. “On another resource it would take about five months to a year.”

In the case of PKA, Anton allowed the researchers to push their simulations to 5

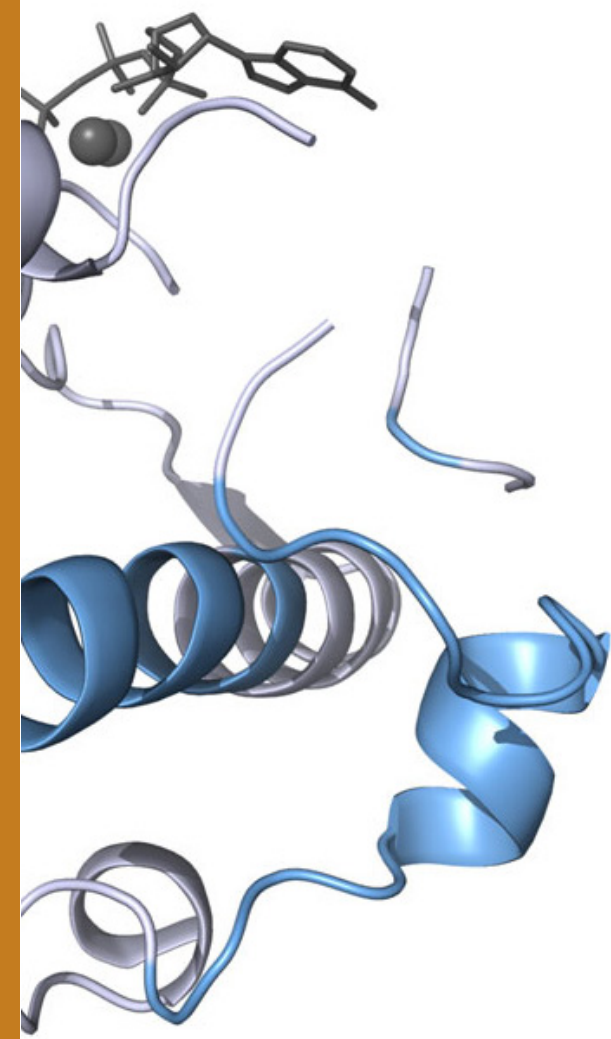
microseconds—a timeframe that revealed a surprising set of behaviors.

COMMUNITIES: A NEW STRUCTURAL CONCEPT FOR PROTEINS

The group’s Anton simulations showed that groups of amino acids in PKA tended to move in concert as the protein changed its shape. The shocker, though, was that the communities identified by the simulations and the domains identified from the static structures were not one and the same. The communities had their own boundaries, which didn’t seem to honor the “obvious” boundaries of the domains.

“We discovered eight communities of amino acids,” says Taylor, the group’s principal investigator. “What was so striking was that each community didn’t segregate by the structural elements we had previously understood but by functional units.”

More interesting, these communities interacted with each other to varying extents. Some affect each other strongly; others, weakly or not at all. Here, it seemed, was a potential explanation for allostery. An amino-acid substitution that makes no obvious change in the protein’s static



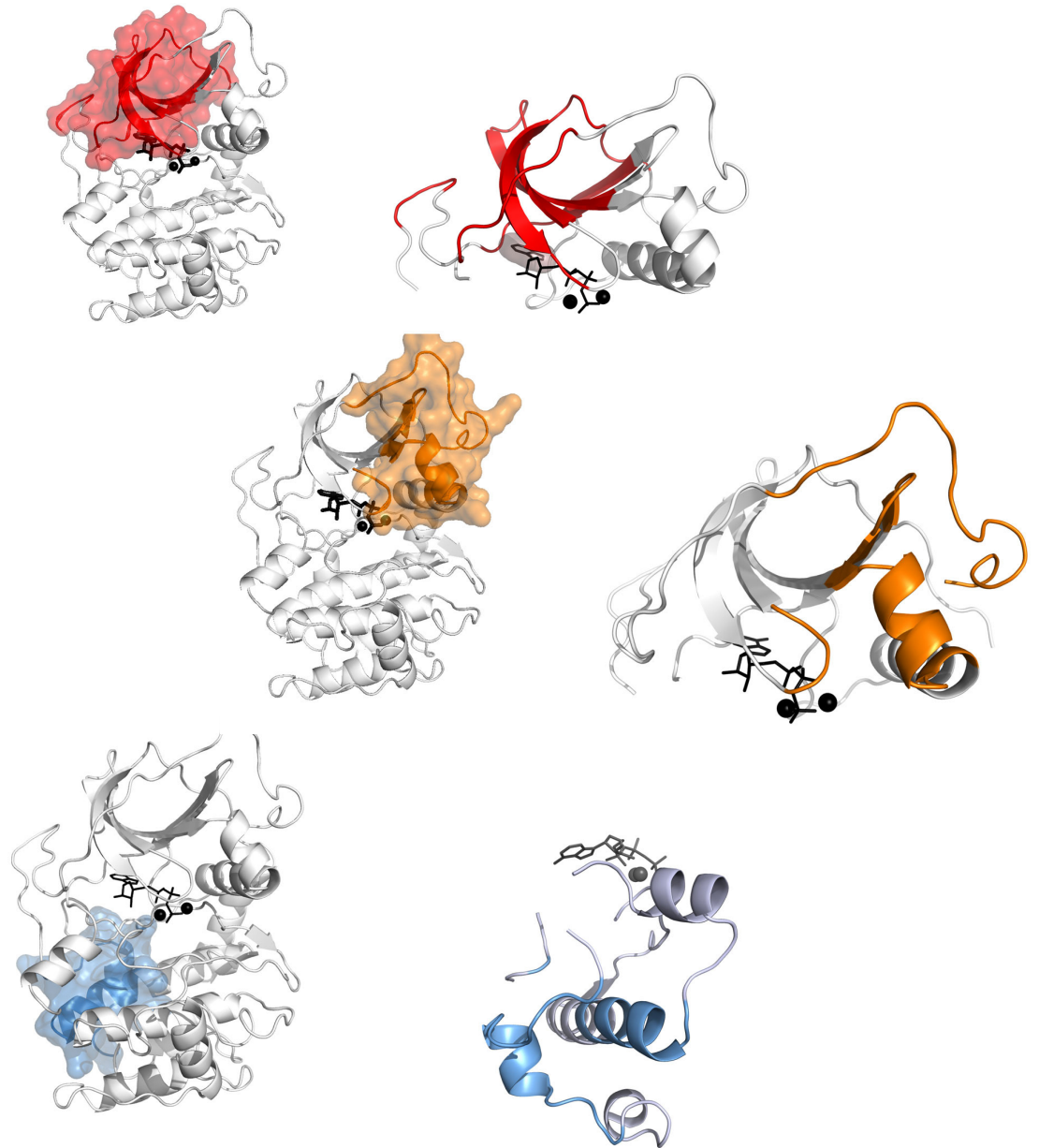
shape nevertheless transforms how that amino acid's community moves. Indirectly, changing that movement in turn modifies the movements of other communities, sometimes on the other side of the protein.

"This is really important in drug discovery and provides an alternative route to identifying novel therapeutics," McClendon says.

For example, a drug that affects an amino acid in a certain community far from the active site might be able to shut down an unwanted function carried out by that community without interfering with desired functions.

"If you can make more subtle changes distant from the active site, it might not have as many side effects as if you just totally gum up the works," he adds.

"This finding opens the door to a mechanistic understanding of many interactions and mutations that were completely obscure," says Kornev. "The communities ... look very clear and logical, knowing how the kinase works. That gives us confidence that we are onto something very promising."



Researchers had believed that domains, the “obvious” substructures of a protein, would define its functional units. But in long-timeframe modeling of PKA by Anton, “communities” of amino acids move and act together, splitting the “kinase domain” in unexpected ways—three of the new communities are marked by colors, above.

SECURING THE CHAIN

Shrimp Transcriptome Offers Secure Future Food Source, Better Understanding of Ocean Food Chain



GENOMICS/MARICULTURE

WHY IT'S IMPORTANT

Litopenaeus vannamei, the whiteleg shrimp, is the number one human food source from the sea. Over a million tons are harvested annually—a huge consumption volume that may not be sustainable. Mariculture, or “sea farming,” could help secure this harvest, but can be economically risky. Because invertebrates like shrimp have primitive immune systems, they can't be vaccinated against diseases that can wipe out a farm colony.

Noushin Ghaffari of Texas A&M AgriLife Genomics and Bioinformatics and colleagues in a multinational collaboration are working to determine the whiteleg shrimp's full genetic sequence. This knowledge can suggest ways of protecting this and related human food species of shrimp from disease. It will also help scientists understand the genetics of krill, a tiny crustacean vital to ocean food webs and forming the main diet of baleen whales. The collaboration is led by Michael F. Griscitello at the College of Veterinary Medicine and Biomedical Sciences, Texas A&M University, and Rogerio R. Sotelo-Mundo at the Centro de Investigación en Alimentación y Desarrollo, Mexico.

HOW PSC HELPED

As a first step in understanding the genomic structure of the whiteleg shrimp, Ghaffari and her collaborators created a novel assembly of the “transcriptome” from several organs in a male whiteleg shrimp. The transcriptome is the sequence of only the active genes in a given tissue. A subset of the full genome, the transcriptome offered a quick way to identify the proteins and nucleic acids the shrimp needs to carry out its life functions.

But even the “smaller” task of assembling the transcriptome posed a huge computational challenge, requiring the team to assemble and match the sequences of about 400 million nucleic acid fragments. PSC staff helped the team leverage the PSC Blacklight supercomputer's very large memory to perform this task, adapting the gold-standard transcriptome assembly software, Trinity, to Blacklight in the process. The team also annotated the transcriptome assembly—creating essentially the equivalent of a legend in a roadmap. This identified over 87,000 candidate genes that will offer insights into invertebrate immunity. The investigators published their results in *Nature Scientific Reports* in November 2014.



Top to bottom: pieces of the carapace of a whiteleg shrimp; a healthy whiteleg shrimp as harvested; and a whiteleg shrimp killed by the Taura syndrome virus (TSV). Herman Gunawan, Bangka, Indonesia.



PSC staff helped Noushin Ghaffari's work with the support of programs within NSF's XSEDE network of supercomputing centers: Extended Collaborative Support Service; Novel and Innovative Projects Program.



“The whiteleg shrimp has been one of the fastest growing aquaculture species, and also has very good potential for establishing food security for the U.S. and its neighbors, including Mexico.”—Noushin Ghaffari, Texas A&M University



PSC AND ITS PARTNERS IN BRIEF

DXC EXPANDS SCOPE

PSC's Data Exacell (DXC) pilot project has increased the community of researchers who can use the system and the scope of projects it can serve by adding new hardware and software resources. The DXC team added data storage to the system and launched dedicated servers to run Hadoop, the popular data-intensive software environment used in many scientific and business applications.

The DXC team has tested the new storage hardware to identify and eliminate bottlenecks at various stages along the data pipeline. The hardware runs PSC's SLASH2 file system, which helps remote users manage and analyze large datasets with speeds higher than those on the users' own hard drives.

In August, PSC staff used the new Hadoop capability to teach a national workshop on Big Data analysis. Hadoop is one of a number of experiments DXC has undertaken to allow researchers to bring previously-successful software solutions to HPC systems with minimal re-tooling and maximal reproducibility.

DXC's mission is to develop and test new hardware and software technologies, as well as system

architectures, to support data-intensive research. These advances will be used at PSC and will be made available as Data Infrastructure Building Blocks to benefit the National Science Foundation's user community.

ANSYS PARTNERSHIP RENEWED

PSC's partnership with Canonsburg, Pa.-based engineering simulation software company ANSYS Inc. has been renewed for another year. Under the continued partnership, ANSYS will provide PSC and its academic users with a no-cost license for ANSYS's suite of mechanical, computational fluid dynamics and multiphysics software. The software, installed and tested on PSC's Blacklight supercomputer, will eventually be installed on the upcoming Bridges system (See p.12).

Availability of ANSYS software on PSC's high-performance computing resources has enabled researchers to tackle computationally-intensive problems that otherwise take an inordinate amount of time on a local computer or that require more memory than is available locally. The partnership has also provided ANSYS with supercomputer time for joint software testing between PSC and ANSYS development and application engineering teams.

Under the renewed partnership, academic researchers at U.S. universities who wish to run ANSYS simulations at PSC will continue to be able to obtain time on PSC systems through the NSF XSEDE's proposal review and allocation process. All universities utilizing these resources must be preauthorized by the partnership group at ANSYS, Inc. All pre- and post-processing must be performed on local computers, and all research must conform to the terms of the license agreement, under which all research must be non-proprietary, public and published/publishable with no restrictions on public availability.

2015 MARC INTERNS COMPLETE SUMMER PROJECTS

Seven summer interns from seven institutions of higher learning have completed bioinformatics projects funded by the fifth year of PSC's National Institutes of Health (NIH) Minority Access to Research Careers (MARC) program grant.

A collaboration with minority-serving partner institutions, MARC is a special technology transfer and outreach program, begun at PSC in 2001, to increase minority participation in biomedical research. Most recently supported by a 2011, five-year NIH grant, MARC includes:

- An intense two-week Summer Institute in bioinformatics at PSC, open to faculty and graduate students at minority-serving institutions. The session focuses on preparing faculty to teach a semester-long bioinformatics course with a focus on cutting-edge technologies like Next-Generation Sequencing.
- An eight-week research internship at PSC for students that have completed bioinformatics training on their local campus.
- The development of a model curriculum for Bioinformatics including related course materials for the fields of Biology, Computational Science and Mathematics. Elements from these courses are incorporated into the workshop as they are completed.
- Assistance in establishing and strengthening the Bioinformatics programs at two minority-serving campuses each year, including teaching assistance for newly established courses.

MARC intern projects this year included computer investigations of how human populations in different parts of the world vary genetically, with an eye toward identifying disease vulnerability; how genes in commercial chicken breeds affect development of fatty tissues; and generating a knowledge base for retroviruses—a family of viruses distantly related to HIV—in birds.

cont'd

PSC AND ITS PARTNERS IN BRIEF

cont'd

2015 PSC MARC Interns

Antonio Camacho Flores

Universidad Metropolitana, San Juan, Puerto Rico

Nicholas Torvell Cook

Tennessee State University, Nashville

Jimmaline Hardy

Morehouse School of Medicine, Atlanta

Okoye M. Echezona

University of Texas, El Paso

Ivan L. Jiminez Ruiz

University of Puerto Rico, Piedras Campus

Selena M. Rodriguez Riviera

University of Puerto Rico, Medical Science Campus

Samantha Octavia Whitaker

North Carolina A&T State University, Greensboro

PSC BUILDING EARLY WARNING TOOL FOR FIXING INTERNET TRAFFIC JAMS

A \$300,000 NSF grant will enable software engineers at PSC to build the first practical tool for warning individual users and their network administrators when their connection has developed a problem

that will slow or halt data flow. The one-year project will build a tool called XSight, intended to become a standard part of the toolbox for maintaining network connections for all operating systems. XSight will



build on PSC's Web10G, a set of software tools for obtaining connection data.

"XSight will represent a proactive approach toward resolving network problems," says PSC's Chris Rapier, principal investigator in the project. "It's different from other approaches to measure network performance because it takes advantage of Web10G's ability to collect data on individual data transfers, both in the network and in an application's interaction with the network."

Previous attempts to measure network performance, while useful for administrators trying to recognize systemic problems, were unable to provide information either timely or detailed enough to help individual users with networking problems. By making use of the data available from Web10G, XSight will overcome these limitations, providing an automated warning system that detects data slowdowns and their causes before they become acute.

XSight will exploit Web10G's ability to extract data about network connections from TCP/IP, the set of protocols for transferring data that underlie the entire Internet. PSC developed Web10G in collaboration with the National Center for Supercomputing Applications.

The XSight project will be funded through the NSF's Early-concept Grants for Exploratory Research (EAGER) program. EAGER supports "high risk/high payoff" early stage exploratory work on "untested but potentially transformative research ideas or approaches."

CHANGING VACCINE VIAL SIZE CAN IMPROVE SUPPLY: PSC/HOPKINS STUDY

Changing vaccine vial size – the total number of doses a single vaccine vial contains – can improve immunization rates and decrease costs, according to a paper recently published in the journal *Vaccine*.

Researchers at PSC and the International Vaccine Access Center at the Johns Hopkins Bloomberg School of Public Health used their supply-chain simulation tool HERMES to estimate the impact of various vaccine vial sizes on the vaccine supply chain, including vaccine availability and costs.

HERMES is a software platform created by the HERMES Logistics Modeling Team that allows public health decision makers to rapidly and safely test ideas for improving the delivery of medical supplies.

The researchers convened stakeholders from various backgrounds to explore the implications of vaccine container choices on access, cost, and safety. In a HERMES-generated model of the West African nation of Benin, vial-size decisions had far-reaching and reverberating impacts throughout the vaccine supply chain. For example, increasing vial size reduced total costs by as much as \$0.25 per dose—a remarkable impact considering that millions of doses of vaccines are administered globally each year.

"Computational modeling can play a critical role in assessing new technologies and alternative products, whose effects can be difficult or impossible to predict without a tool like HERMES," says Leila Haidari, HERMES team coordinator, PSC public health applications manager and lead author in the study. "Vaccine supply chains are complex, dynamic systems that can make great use of modeling to inform policy decisions."

PSC Resources for Data-Driven Science

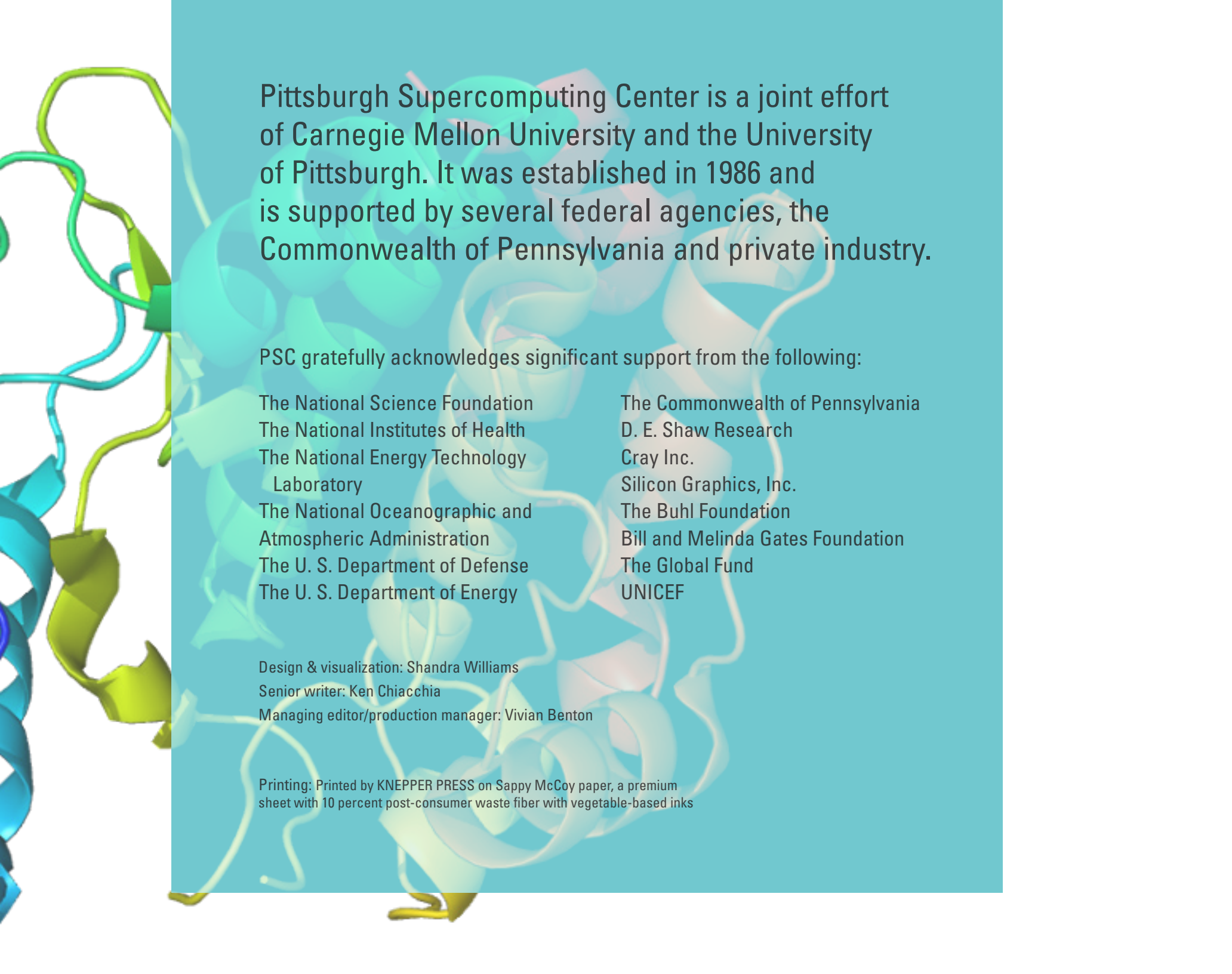
Blacklight is an SGI Altix® UV1000 supercomputer designed for memory-limited scientific applications in fields as different as biology, chemistry, cosmology, machine learning and economics. Funded by the National Science Foundation (NSF), Blacklight carries this mission out with partitions with as much as 16 terabytes of coherent shared memory. *Note: As of August 15, Blacklight was decommissioned. During the transition period to **Bridges**, PSC's new supercomputer, PSC provides **Greenfield**, a new resource that, like Blacklight, features large shared memory.*

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 other computing resources, 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, go to WWW.PSC.EDU



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

The Global Fund

UNICEF

Design & visualization: Shandra Williams

Senior writer: Ken Chiacchia

Managing editor/production manager: Vivian Benton

Printing: Printed by KNEPPER PRESS on Sappy McCoy paper, a premium sheet with 10 percent post-consumer waste fiber with vegetable-based inks