Multiyear Simulation Study Provides Breakthrough in Membrane Protein Research

Laura Wolf | Argonne National Laboratory

“Molecular machines,” composed of protein components, consume energy to perform specific biological functions. The concerted actions of the proteins trigger many of the critical activities that occur in living cells. However, like any machine, the components can break (through various mutations), and then the proteins fail to perform their functions correctly.

It’s known that malfunctioning proteins can result in a host of diseases, but pinpointing when and how a malfunction occurs is a significant challenge. Very few functional states of molecular machines are determined by experimentalists working in wet laboratories. Therefore, more structure-function information is needed to develop an understanding of disease processes and to design novel therapeutic agents.

The research team of Benoît Roux, a professor in the University of Chicago’s Department of Biochemistry and Molecular Biology and a senior scientist in Argonne National Laboratory’s Center for Nanoscale Materials, relies on an integrative approach to discover and define the basic mechanisms of biomolecular systems—an approach that relies on theory, modeling, and running large-scale simulations on some of the fastest open science supercomputers in the world.

Computers have already changed the landscape of biology in considerable ways; modeling and simulation tools are routinely used to fill in knowledge gaps from experiments, helping design and define research studies. Petascale supercomputing provides a window into something else entirely: the ability to calculate all the interactions occurring between the atoms and molecules in a biomolecular system,
such as a molecular machine, and to visualize the motion that emerges.

The Breakthrough

Roux’s team recently concluded a three-year Innovative and Novel Computational Impact on Theory and Experiment (INCITE) project at the Argonne Leadership Computing Facility (ALCF), a US Department of Energy (DOE) Office of Science User Facility, to understand how P-type ATPase ion pumps—an important class of membrane transport proteins—operate. Over the past decade, Roux and his collaborators, Avisek Das, Mikolai Fajer, and Yilin Meng, have been developing new computational approaches to simulate virtual models of biomolecular systems with unprecedented accuracy.

The team exploits state-of-the-art developments in molecular dynamics (MD) and protein modeling. The MD simulation approach, frequently used in computational physics and chemistry, calculates the motions of all the atoms in a given molecular system over time—information that’s impossible to access experimentally. In biology, large-scale MD simulations provide a perspective to understand how a biologically important molecular machine functions.

For several years, Roux’s research has been focused on the membrane proteins that control the bidirectional flow of material and information in a cell. Now, in a major breakthrough, he and his team have described the complete transport cycle in atomic detail of a large calcium pump called sarco/endoplasmic reticulum calcium ATPase, or SERCA, which plays an important role in normal muscle contraction. This membrane protein uses the energy from ATP hydrolysis to transport calcium ions against their concentration gradient and, importantly, its malfunction causes cardiac and skeletal muscle diseases.

Roux and his team wanted to understand how SERCA functions in a membrane, so they set out to build a complete atomistic picture of the pump in action. Das, a postdoctoral research fellow in Roux’s lab, did this by obtaining all the transition pathways for the entire transport cycle using an approach called the string method—essentially capturing a “molecular movie” of the transport process, frame by frame, of how different protein components and parts within the proteins communicated with each other (see Figure 1).

The Science

A membrane protein, like all protein molecules, consists of a long chain of amino acids. Once fully formed, it folds into a highly specific conformation that enables it to perform its biological function. Membrane proteins change shape and go through many conformational “states” to perform their functions.

“From a scientific standpoint, membrane proteins such as the calcium pump are very interesting because they undergo complex changes in their three-dimensional conformations,” said Roux. “Ultimately, a better understanding may have a great impact on human health.”

Experimentalists understand the structural details of proteins’ stable conformational states but very little about the process by which a protein changes from one conformational state to another. “Only computer simulation can explore the interactions that occur during these structural transitions,” said Roux.

Intermediate conformations along these transitions could potentially provide the essential information needed for the discovery of novel therapeutic agent design. (Drugs are essentially molecules that counteract the effect of bad mutations to help recover the normal functions of the protein.) Because membrane proteins regulate many aspects of cell physiology, they can serve as possible diagnostic tools or therapeutic targets.

Roux and his team are trying to obtain detailed knowledge about all the relevant conformational states that occur during SERCA’s transport cycle. In years one and two of the study, Roux’s team identified two of the conformation transition pathways needed to describe the cycle. Last year, the project shifted focus to the three remaining pathways.

The ALCF Advantage

As is the case for much of the domain science research being conducted on DOE leadership supercomputer systems today, biomolecular science relies on advances in methodology as well as in software and hardware technologies. The usefulness of Roux’s simulations hinges on the accuracy of the modeling parameters and on the efficiency of the MD algorithm enabling the adequate sampling of motions.

Computational science teams can spend years refining their application code to do what they need it to do, which is often to simulate a particular physical phenomenon at the necessary space and time scales. Code advancements can push the
simulation capabilities and take advantage of the machine's features, such as high processor counts or advanced chips, to evolve the system for longer and longer periods of time.

Roux and his team used a premier MD simulation code, called NAMD, that combines two advanced algorithms—the swarm-of-trajectory string method and multidimensional umbrella sampling.

NAMD, which was first developed at the University of Illinois at Urbana-Champaign by Klaus Schulten and Laxmikant Kale, is a program used to carry out classical simulations of biomolecular systems. It's based on the Charm++ parallel programming system and runtime library, which provides infrastructure for implementing highly scalable parallel applications. When combined with a machine-specific communication library (such as PAMI, available on Blue Gene/Q), the string method can achieve extreme scalability on leadership-class supercomputers.

ALCF staff provided maintenance and support for NAMD software and helped coordinate and monitor the jobs running on Mira, ALCF’s 10-Pflops IBM Blue Gene/Q.

ALCF computational scientist Wei Jiang has been actively collaborating with Roux’s team since 2012, as part of Mira’s Early Science Program. Jiang worked with IBM’s system software team on early stage porting and optimization of NAMD on the Blue Gene/Q architecture. He’s also one of the core developers of NAMD’s multiple copy algorithm, which is the foundation for multiple INCITE projects that use NAMD.

Jiang, who has a background in computational biology, considers the recent work a significant
breakthrough. “Only in the third year of the project did we begin to see real progress,” he said. “The first and second year of an INCITE project is often accumulated experience.”

The computations Roux and his team ran for this breakthrough work will serve as a roadmap for simulating and visualizing the basic mechanisms of biomolecular systems going forward. By studying experimentally well-characterized systems of increasing size and complexity within a unified theoretical framework, Roux’s approach offers a new route for addressing fundamental biological questions.

Acknowledgments
An award of computer time was provided by the US Department of Energy’s Innovative and Novel Computational Impact on Theory and Experiment (INCITE) program. This research used resources of the Argonne Leadership Computing Facility, which is a US DOE Office of Science User Facility supported under Contract DE-AC02-06CH11357.

Laura Wolf is a science writer and editor for Argonne National Laboratory. Her interests include science communication, supercomputing, and new media art. Wolf received a BA in political science from the University of Cincinnati and an MA in journalism from Columbia College Chicago. Contact her at lwolf@anl.gov.