# Oak Ridge Leadership Computing Facility Snapshot
# The Week of June 14, 2010

**Supercomputers Simulate the Molecular Machines that Replicate and Repair DNA**
*Scientists model replisome components to understand their role in health and disease*

Imagine you are an astronaut. A piece of space junk has cut a gash into the side of the space station, and you have been tasked with repairing the damage. Your spacesuit is equipped with a clamp, which you open, slide onto a tether connecting you to the space station, and close. Then you slide to the far end of the gash and begin applying composite material to fill the holes. You slide along the gash making repairs until you are done.

DNA replication, modification, and repair happen in a similar way, reveals a biomedical simulation run on the world's fastest supercomputer. Ivaylo Ivanov of Georgia State University, John Tainer of the Scripps Research Institute, and J. Andrew McCammon of the University of California–San Diego, used Jaguar, a Cray XT high-performance computing system at Oak Ridge National Laboratory (ORNL), to elucidate the mechanism by which accessory proteins called sliding clamps are loaded onto DNA strands and coordinate enzymes that enable gene repair or replication. They share their findings, which inspire a new approach for attacking diverse diseases, in the May 10 issue of the *Journal of the American Chemical Society.*

"This research has direct bearing on understanding the molecular basis of genetic integrity and the loss of this integrity in cancer and degenerative diseases," says Ivanov, whose investigation was supported by the Howard Hughes Medical Institute and the National Science Foundation's Center for Theoretical Biological Physics.

The paper focuses on the clamp-loading cycle in eukaryotes, or organisms whose genetic material is enclosed in a nuclear membrane (as opposed to bacteria and viruses, whose genes are not compartmentalized). The researchers reveal that a "clamp loader" called replication factor C places a doughnut-shaped "sliding clamp" called proliferating cell nuclear antigen (PCNA) onto DNA. The clamp loader first binds to the clamp to activate its opening with energy from adenosine triphosphate (ATP). Protein secondary-structure elements called beta sheets, at the junctures of the clamp's three subunits, separate at one juncture. A complex made up of the open clamp and the clamp loader then encircles primer-template DNA, which is double stranded in one region and single stranded in another. Next, in a process fueled by ATP hydrolysis, the clamp closes. The clamp is now free to slide along the DNA strand and coordinate enzymes needed for replication and repair.

"Sliding clamps and clamp loaders are part of the replisome—the molecular machinery responsible for the faithful duplication of the genetic material during cell division," explains Ivanov. "The replisome is very complex and dynamic, with interchanging parts. It's an incredibly challenging system to understand." Simulating just a few of its constituent parts—the clamp/clamp loader assembly—required a system of more than 300,000 atoms. "To make

progress simulating the system in a reasonable amount of time, we needed access to large-scale computing."

In 2009 the researchers were awarded 2.6 million processor hours through INCITE, the Innovative and Novel Computational Impact on Theory and Experiment program, which provides pioneering scientists and engineers with access to the Department of Energy's leadership computing facilities at Oak Ridge and Argonne national laboratories. They ran the NAMD molecular dynamics code on Jaguar's XT4 component, which calculates at a speed of 0.263 petaflop, or quadrillion floating point operations, per second. Subsequent calculations were conducted with Jaguar's XT5 component, which has a peak performance of 2.332 petaflops and ranks #1 on the TOP500 list of the world's fastest computers. To further the investigation, in 2010 the researchers received a 2-year allocation of 4 million processor hours on Jaguar XT5.

**Master coordinator**

In DNA replication the clamp slides along a strand of genetic material made up of repeated building blocks called nucleotides, which each consist of a base, a five-carbon sugar, and phosphate groups. Nucleotides differ only in the type of base they carry, so bases are what determine the genetic message. Enzymes called polymerases catalyze the formation of a new DNA strand from an existing DNA template. To do so they first associate with sliding clamps. (Polymerases can catalyze strand extension in the absence of the clamp, but in that case, after a short burst of synthesis the polymerase falls off the DNA. The role of the clamp is to prevent such dissociation and make sure replication continues uninterrupted for thousands of nucleotide incorporations.) Polymerases iteratively add one of four bases until they have strung together thousands of bases. Unique sequences of bases encode the blueprints of life forms from swinepox virus and *Salmonella* bacteria to rabbits and redwood trees.

In DNA repair the sliding clamp serves as the master coordinator of the cellular response to genetic damage. A number of proteins, such as cell cycle checkpoint inhibitors or DNA repair enzymes, attach themselves to the clamp to perform their functions. In this capacity the role of the clamp is to orchestrate a variety of DNA modification processes by recruiting crucial players to the replication fork, a structure in which double-stranded DNA gives rise to single-stranded prongs that serve as templates for making new DNA. Given this dual function of PCNA in both replication and repair, it is not surprising that the clamp has been implicated in a number of diseases accompanied by excessive replication and unchecked cell growth (such as cancer). PCNA modifications are key in deciding the fate of the replication fork and ultimately determine both tumor progression and the outcome of anticancer treatments. Therefore, PCNA has been used as a diagnostic and prognostic tool (biomarker) in cancer progression.

Most studies of DNA replication have focused on polymerases. "Instead of just focusing on polymerase, we can interfere with many different components within this complex machinery," Ivanov says of the replisome. "That may allow new drug targets to be developed for hyperproliferative diseases such as cancer."

An improved understanding of the replisome may make it possible to exploit differences among organisms as diverse as viruses, bacteria, plants, and animals. Although clamp loaders from the different kingdoms of life share many architectural features, significant mechanistic differences exist between the various clamp-loading machines, specifically in the ways ATP is used. Drugs targeted to the clamp loader could selectively inhibit replication of viral DNA in diseases such as chickenpox, herpes, and AIDS without interfering with DNA replication in normal human cells. Similarly, in processes with increased DNA replication, such as cancer, inhibiting clamp loading might produce therapeutic effects without unwanted side effects.

What's next? Ivanov and colleagues would like to study the mechanisms of alternative clamps such as the PCNA-related protein complex 9-1-1. This complex activates a checkpoint cascade that signals the cell to arrest division upon detection of DNA damage. The therapeutic prospects based on the fundamental research fuel Ivanov's enthusiasm, which is palpable. "I want to have an idea about the entire clamp-loading cycle including all the intermediates, and I would like to know how ATP is utilized during the clamp-loading cycle," he says of future goals. "There has been some very exciting experimental work, and we want to incorporate all the available experimental information into our models."

**Nuclear Theorists Pin Down the Proton-Halo State in Fluorine-17**

*Oak Ridge Leadership Computing Facility contributes to theoretical physics*

A halo may be difficult to acquire in terms of virtue, but it can also be tough to calculate in terms of physics. Gaute Hagen from ORNL, Morten Hjorth-Jensen from the University of Oslo, and Thomas Papenbrock from the University of Tennessee, Knoxville (UTK), have managed to do just that, however, and report their findings in the article "Ab-initio computation of the 17F proton-halo state and resonances in A = 17 nuclei," published in May in *Physical Review Letters.*

"The halo state in the atomic nucleus fluorine-17 is characterized by an excited proton that is orbiting in an appreciable distance around oxygen-16. This spatially extended nuclear state has so far eluded a direct computation from scratch," says Papenbrock. "The first-principles computation of the proton halo in fluorine-17 is a daunting challenge for nuclear theory due to the presence of seventeen strongly interacting particles and a weak binding of the state. It demands applications of novel techniques for open quantum systems and leadership computational resources." The researchers' calculations are based on the strong nuclear interaction, which in turn is rooted in fundamental properties of the interactions between quarks and gluons. They reproduced the tiny separation energy of the halo (about 105 keV, compared to 12,127 keV for the ground state in oxygen-16) and demonstrated that an appreciable contribution to its binding energy comes from the coupling to the particle continuum.

A halo nucleus differs from the more traditional nuclei because it has one or more nucleons (protons or neutrons) that are only weakly bound to the nuclear core. Consequently, they drift far away from it, forming, in effect, a halo. These nuclei are difficult to study because their lives are both short (often lasting only milliseconds) and fragile. Halo nuclei appear at the limits of nuclear existence, very near a place called the drip line. This is the perilous territory

where the number of protons and the number of neutrons are plotted against each other and one too many of either means the nucleus will not hold together. Halo nuclei also come with a large number of degrees of freedom—independent configurations required to explain how a system is built.

Hagen, Hjorth-Jensen, and Papenbrock set out to study fluorine-17, a "mirror nucleus" of oxygen-17. Each of these isotopes has an atomic number of 17, but with their protons and neutrons in flipped numbers (fluorine-17 has 9 protons and 8 neutrons, while oxygen-17 has 8 protons and 9 neutrons). Fluorine-17, in particular, has a "halo" formed by an excited proton orbiting far away from the oxygen-16 core that plays an important role in nucleosynthesis, the stellar processes that generate the elements that surround us.

The ORNL-Oslo-UTK team developed and implemented sophisticated theoretical and computational methods to solve the nuclear many-body problem— it is difficult to pin down precise calculations of a system with more than two interacting bodies—for the 17 interacting particles in this isotope. The team used first-principle *(ab initio* in Latin) interactions derived from quantum chromodynamics, which describes the strong interactions between elementary particles, to build the nuclear Hamiltonian, the operator that describes the energy of a system in terms of its momentum and positional coordinates. They used the coupled-cluster method — a numerical technique that solves such quantum many-body problems — and ORNL's Jaguar supercomputer to successfully complete the *ab initio* calculations of the proton halo state in fluorine-17. The researchers used nearly 100,000 processor hours on Jaguar to carry out the calculations, and showed a computed binding energy (what holds the nucleus together) that closely reflects experimental data.

"It's rewarding to see the time spent developing and scaling our application produce exciting results," Hagen explains. "Without the capabilities provided by Jaguar, not only would our time-to-solution be impacted, but even reaching a solution would be difficult. We depend on Jaguar because our computations are expensive in memory and computing time."

Code development and science-directed calculations were made possible through the Department of Energy Office of Science and Office of Advanced Scientific Computing Research (ASCR) programs. Computing resources for these calculations were provided through the Innovative and Novel Computational Impact on Theory and Experiment (INCITE) program and contribute to the Universal Nuclear Energy Density Functional (UNEDF) SciDAC (Scientific Discovery through Advanced Computing) program aimed at achieving a comprehensive and unified description of nuclei and their reactions.

The more tools scientists have to calculate the properties of nuclei—how long they live, what holds them together, and how they decay—the more clearly they can investigate the limits of nuclear existence, understand phenomenological models of the nucleus, and predict nuclear properties in applied fields like nuclear medicine or stockpile stewardship.

*Source: University of Tennessee at Knoxville with additional reporting from the Oak Ridge Leadership Computing Facility.*

**ORNL Staff, Jaguar Recognized at ISC'10**

*Staffers recognized for work with InfiniBand software while Jaguar again claims top spot on Top500 List*

On June 3, Richard Graham and Stephen Poole of the ORNL's Computer Science and Mathematics Division and Oak Ridge Leadership Computing Facility jointly accepted an HPC Advisory Council award with Mellanox Technologies at the 2010 International Supercomputing Conference (ISC'10) in Hamburg, Germany. The award was presented for their collaborative work on software technology for Mellanox's InfiniBand network, a switched fabric communications link between processor nodes and input/output nodes in high-performance computers.

"We've been working with Mellanox on the technical side for about two years to help them understand our application requirements to define a new network product," said Graham, Applications and Performance Tools Group leader at ORNL. Graham and Poole assisted Mellanox with software requirements that increase offload capabilities of the InfiniBand network, increasing application performance through faster, more balanced communication between processing units in high-performance systems.

"This award recognizes the fact that what we've done over the last couple of years has a major impact on the high-performance computing industry," said Graham.

The award was one of four HPC Advisory Council awards presented at the closing ceremony of the 25th annual ISC. Graham accepted the award in the category of "Innovative HPC Software Technology Announced Between June 2009 and April 2010" with Poole, chief scientist with the ORNL Computer Science and Mathematics Division. Mellanox representatives accepting the award were Michael Kagan, co-founder and chief technology officer, and Gilad Shainer, senior technical marketing manager. The HPC Advisory Council works to advance the state of high-performance computing through education and support for end users and application developers.

Also announced at ISC'10, the Oak Ridge Leadership Computing Facility's Cray XT5 "Jaguar" was again named fastest supercomputer in the world with a performance of 1,759 petaflops, or over 1.7 thousand trillion calculations per second. This designation comes from the Top500 list—a biannually published list of the fastest supercomputers in the world.