

Texas Advanced Computing Center - Blueprint for the Affordable Genome

Blueprint for the Affordable Genome

Ranger helps University of Illinois researchers simulate nanopore gene sequencer

The first human genome took 13 years and \$3 billion to produce. Today, geneticists generate the same information in a matter of months for a fraction of the cost.

Transformational developments in the technology and methodology of DNA sequencing have made these advances possible. “Next-generation” sequencers deployed in 2005 — which create millions of duplicate segments of DNA and piece them together to determine their order — are making their mark on the life sciences. Meanwhile, teams around the world are working to develop new and improved “Generation 3” and “Generation 4” DNA sequencers that can ingest a strand of nucleotide bases and “read” its genetic code directly.

The medical community eagerly awaits the arrival of the \$1,000 personal genome promised by these *next-nextgeneration* sequencers. They predict the advent of a \$1000 genome will lead to breakthroughs in the study of disease and evolution, and the development of personalized medicine.

Aleksei Aksimentiev, a computational physicist at the University of Illinois Urbana-Champaign (UIUC), is working with experimentalists at UIUC, Notre Dame, and the University of Washington to develop a class of cutting-edge sequencers with a unique functionality.

The sequencers use an electric field to drive a strand of DNA through a small hole, or “nanopore,” either in silicon or a biological membrane. If this process can be controlled, the sequencer will be able to read each base pair, in order, by measuring the change in ionic current as the pair moves through the pore of the membrane.

A great concept — but does it work?

Not yet, according to Aksimentiev. Blockages and enhancements, noisy signals, and base pairs that pass too quickly through the pore, plague the experimental designs of these sequencers.

“Because there are so many factors, this intuitively simple physical picture doesn’t apply in practice,” Aksimentiev said. “The experiments cannot see what’s going on inside a nanopore. That’s why we do all-atom molecular dynamic simulations of their experimental systems — to explain what processes give rise to the signals they measure.”

Using the *Ranger* supercomputer at the Texas Advanced Computing Center (TACC), Aksimentiev produced atom-by-atom models of both experimental and untested nanopore designs, and set them in motion. The simulations revealed the microscopic conformation of DNA in a nanopore, leading to insights into how to calibrate the signals of the bases and optimize the design of the nanopore most effectively.

Aksimentiev discovered that the key requirement for sequencing DNA directly is positioning the DNA in the pore for a time interval that is long enough to read its sequence.

“If the DNA moves too fast, then one cannot read out the signal to distinguish the difference between the base pairs,” he said. “We have to find a way to trap the DNA.”

Aksimentiev and his colleagues believe they’ve found a way.

By stretching DNA with an electrical field, the strands fit into a pore smaller than their unstretched diameter. Turning off the field traps the DNA in the hole. Then, by pulsing the field, stretching and relaxing the DNA, the strand moves base-by-base through the pore. Since the diameter of the hole forces the DNA to tilt, the pairs may be disentangled in the signal detection. The team filed for a provisional patent in 2009 for their design of the first nanopore sequencer able to accommodate double-stranded DNA.

High-performance supercomputing systems are required for this kind of research because of the extreme scale and resolution of the simulations. Aksimentiev's classical molecular dynamics simulations model 200,000 atoms interacting over billions of time steps.

"The difference between A, C, G and T nucleotides is just a few atoms, literally between four and eight," said Aksimentiev. "So you have to have all-atom resolution, you have to get the physics right, and you have to simulate for a long time."

These many-atom, long-time-step simulations are a computationally expensive endeavor. Last year alone, the project required more than 7 million computing hours on *Ranger*, TACC's most powerful system. In 2010, the project will use more than eight million hours on *Ranger*, and 10 million hours on the Department of Energy's *Jaguar* system, for which Aksimentiev won a special [INCITE](#) award.

In addition to the scale of the problem, speed is a key factor driving the need for more powerful supercomputers. In order to be relevant, simulations need to be performed in a time frame that can keep pace with laboratory testing.

"Our collaborators cannot wait for us to run a simulation over three months," Aksimentiev said. "This is a very competitive field, so having *Ranger* was a tremendous help."

Aksimentiev produced atomic resolution simulations of his collaborators' proposed systems in under a week, keeping up with the progress of their experiments, and supplying key insights.

"Dr. Aksimentiev's simulations represent our eyes — we can't see without them," said [Greg Timp](#), professor of electrical engineering at the University of Notre Dame and Aksimentiev's collaborator. "His molecular dynamics simulations provided a microscopic view into the translocation process, showing with atomic precision the configuration of the molecule in the pore, as well as the corresponding current and voltage signals that develop as each base travels through the constriction."

The most recent simulations aim to improve the functioning of the double-stranded sequencer through variations in its pore geometry and the concentration of electrolytes.

"Aksimentiev's work is a brilliant example of how high performance computing is becoming a key tool of experimentalists from across a broad range of scientific disciplines," said Michael Gonzales, computational biology program director at TACC. "Advanced compute systems are no longer relegated to the analysis of experimental results. Rather, computing is in many ways driving the experimental design."

The nanopore system considered by Aksimentiev's group has no limit for the read length and doesn't require labels to read out a sequence; thus, the new device promises a drastic reduction in cost and a commensurate increase in speed. Furthermore, if it's made from a solid-state nanopore, Aksimentiev believes the sequencing device can be integrated easily with existing electronics, which would allow it to be scaled up to make a massively parallel device.

The development of such a sequencer would have important ramifications for medicine, biology, and human health.

"If we succeed, it will have a noticeable impact on the way we understand and treat human diseases," Aksimentiev exclaimed. "Everyone will be able to afford their DNA sequence."

April 19, 2010

The Texas Advanced Computing Center (TACC) at The University of Texas at Austin is one of the leading centers of computational excellence in the United States. The center's mission is to enable discoveries that advance science and society through the application of advanced computing technologies. To fulfill this mission, TACC identifies, evaluates, deploys, and supports powerful computing, visualization, and storage systems and software. TACC's staff experts help researchers and educators use these technologies effectively, and conduct research and development to make these technologies more powerful, more reliable, and easier to use. TACC staff also help encourage, educate, and train the next generation of researchers, empowering them to make discoveries that change the world.