

Klaus Schulten Talks about the Evolution of Computational Biophysics

Professor Klaus Schulten will deliver the opening ISC'14 keynote address on computing in biomedicine and bioengineering. At this year's International Supercomputing Conference (ISC'14) on June 22-26, in Leipzig Germany, Professor Klaus Schulten will deliver the opening [keynote address](#) on computing in biomedicine and bioengineering. Schulten, a physicist by training, now devotes his time to computation biophysics. He has contributed to several key discoveries in this area, has garnered numerous awards and honors for his work, and is considered one of preeminent leaders in field. In a recent interview with ISC, he talks about the evolution of computational physics, its practical applications, and what the future holds.



Computational biophysics has been characterized as a "computational microscope" for studying living systems. What types of structures are researchers most focused on today?

Schulten: There has been quite a change over the last decade in the type of structures biologists are looking at, but much of the work is centered on proteins. Proteins are very small machines in cells — the human cell has as many as the United States has citizens. These many millions of proteins work together, and that's what makes a biological system "living."

In the 50s, computational researchers were able to look at small individual proteins, which led them to many Nobel Prizes. Today, they're looking at large assemblages of proteins. Simple ones like virus capsids are often constructed of thousands of proteins, but all of one kind. Others are involved in energy transformations like respiration or photosynthesis, which again have thousands of proteins but, in this case, many different kinds.

There are well-known structures like ribosomes, which read genetic information and turn it into the synthesis of proteins. And then there are sensory proteins, involved in things like vision and taste and olfaction. There are also proteins that are important for neurons in the brains, which facilitate the communication between the neurons. Others are involved in the genetic machinery, like chromatin, which acts as a cradle for DNA.

Whereas the individual proteins we used to study often had a just a thousand or a few thousand atoms, these protein assemblages involve many millions of atoms — up to a billion. These are much more representative for what we are now studying in living entities.

What advantages do these computer simulations have over real-world observations?

Schulten: The advantages are huge. You just have to be careful trusting the computer.

Physical measurements can only be taken under certain conditions. For example, the light microscope is really a versatile instrument. But it can only resolve things down to a certain size, and that size is limited by the wavelength of light — a wavelength much too big to see details like molecules in living cells. In an electron microscope, you have a much higher resolution, but you have to use a vacuum environment and dry conditions to examine what you're looking at. So, most of these experimental instruments are very limited.

That's where the computer comes in, to be a microscope where real microscopes don't work. Just as Boeing uses a computer to simulate airplanes, we simulate what we know is in the cell. Now, we're not quite as good as what Boeing can do with airplanes, but we're pretty good and getting better every year.

The computer simulations combine the data of experimental results, including light and electron microscopes, in addition to other kinds of research. The computer is also very good at applying the basic principles of science to living systems. Only the computer can put apples and oranges together and give you the structures that are consistent with all of these measurements. That's its strength.

Could you give some examples of recent discoveries made possible with computational biophysics? For example, have any specific results found practical applications in medical treatments, agriculture, energy production or other areas?

Schulten: Computational methods have made astonishing advances on how certain animals navigate, many of which possess a magnetic sense and are able to determine magnetic north and south. Some animals, like fish, have an incredible sense of touch, so can stay completely still in flowing water — and thus hidden from predators — by measuring the flow on the surface of their skin. These types of discoveries were uncovered by computer simulations.

The computer is also an extremely useful instrument for pharmacological research because the targets of most of our medical treatments are specific molecules in the body. That's what the computer can help with — describing the target molecules and suggesting pharmacological treatments against what the computer sees.

A relatively new discovery made possible by the latest generation of computers is the structure of the HIV virus. Now that this structure is known, great progress has been made with HIV treatment. And we can develop new treatments that outrace the virus as it adapts.

In my own research, I study the ribosome from a pharmacological perspective. The ribosome is a target of antibiotics, and we constantly have to find new ones because bacteria become resistant to the antibiotics we've developed.

We can also develop strategies using nanotechnology, which is a relatively new technology area being looked at for medical use. Nanotechnology uses devices so small, even light microscopes can't see them. They can be used to monitor cell activity, for example, to diagnose certain types of cancer. Here again, the computer can help develop these devices.

Another active area my research group is looking at is what we call "second-generation" biofuels. We already know it's possible to make gasoline from ethanol and valuable food sources like corn and sugarcane, but that takes food away from people. Now, instead, we want to make biofuel from industrial waste. This requires unique molecular transformations, which are being developed jointly between experimental and computational scientists.

So, we're seeing a revolutionary rise in the use of computers for bio-nanotechnology, medicine and the life sciences in general.

Software applications like NAMD and VMD have been critical to the work of biophysicists. Could you give a layman's description of how these applications work and why they're important?

Schulten: NAMD and VMD are scientific software that has been systematically funded by the National Institute of Health (NIH) for more than 25 years. Recently, the funding has been renewed for another five years. The result has been professional software that is used by about 300,000 researchers worldwide.

This software has the same user interface from a laptop to the most powerful supercomputers. These include petascale systems like *Blue Waters* at the University of Illinois at Urbana-Champaign, where I'm a professor, the *Titan* supercomputer at Oak Ridge National Lab, and the *Piz Daint* computer just installed in Switzerland.

NAMD performs the molecular simulations, while VMD sets up the simulations and helps analyze and visualize the results. Today, simulation outputs are so complex that you need this second step for interpretation. Most of the intelligence of the human brain is in the visual cortex, so VMD exploits that. This analysis and visualization step now occupies half of the compute time.

Physicists from CERN could tell a similar story. They developed the initial software for the Internet and are heroes of computing, too. But, when I go to meetings today, I realize the technologies we're using are at least as advanced as those of my physics colleagues. Life scientists are now big users of computers and drive a lot of the technology advances in this area.

Will this biophysics software be able to scale up and use the power of larger supercomputers as the hardware advances to the exascale level?

Schulten: To get to petascale computing, it took five years of software development. Nobody knew what they were doing, since the computers were not available yet. But today, we are able to scale up to the largest petascale systems, even the GPU-accelerated ones. And now we're doing the same thing for exascale.

For the first time, we are also power profiling our software algorithms. The problem we have is that we can only optimize what we can measure, and there are no good general-purpose tools available today for measuring how much power your software really uses. So, we developed detectors for power consumption with very high time resolutions — microseconds — for things like ARM chips and GPUs.

Now, we're beginning to power profile our algorithms and trying to do the same computing at one-tenth or one-hundredth the power consumption. I think, in about four years, these kinds of power-profiled algorithms will be the absolute key that permits the jump to exascale.

What will exascale supercomputers allow computational physicists to undertake that cannot be done with today's systems?

Schulten: When I was a young professor, to simulate a single protein required the most powerful and expensive computer available. In those days, it was a Cray. At the moment, we are in a phase in which simulations are getting better and better as more powerful computer becomes available.

We are now studying the macromolecules of thousands of proteins working together. That is a big step that was completely impossible to study before. Today, the work is being done on the biggest petascale computers. With the future exascale one, we will be able to do even more: chemically resolve the details of the cell.

The goal of modern life science is to characterize these biological systems from the atom to the cell. We are now somewhere in the middle. A human cell is around 10 micrometers long, and we can simulate it at a scale of about one-hundredth to one-thousandth of that. To reduce it by a factor of 10 — a factor of 1,000 by volume — we will need a computer 1,000 times as powerful.