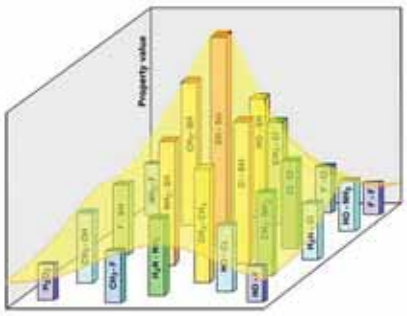
Finding the Best Molecule for the Job

Every pharmaceutical company wants to find the next blockbuster drug. Yet finding molecules with a complete set of desired properties is tricky because of the astronomical number of medium-sized organic molecules. Now researchers at Duke University have developed a novel way to design virtual molecules from scratch. The work was published in the February 17, 2006, online issue of the *Journal of the American Chemical Society*.

"The biggest challenge in chemistry is being able to design molecules for particular purposes," says **Weitao Yang, PhD**, a professor of chemistry at Duke University. "You can only do experiments on real molecules, but virtual techniques let you use non-real molecules to explore the molecular space."

Yang along with colleague **David Beratan, PhD**, professor of chemistry, and post-doctoral fellows **Mingling Wang, PhD**, and **Xiangqian Hu, PhD**, developed an innovative approach. Rather than calculate properties of an enormous number of possible individual molecules, their framework approximates the properties over a continuous landscape in which the individual molecules lie. The model relies on knowledge of how atoms can be joined based on the energy relationships between nuclei and electrons in atoms. This narrows down the possible combinations and smoothes out discrete characteristics, such as atomic number, and thus provides a continuous surface for optimization.



Yang's model allows researchers to find the best molecule for a desired property. In this graph, the bar heights represent the amount of a property that each candidate molecule possesses. The model finds the best molecule by evaluating different combinations of molecular groups along the smooth surface over the bars.

For their proof of concept, the researchers focused on the properties that determine the ability of an atom's electron cloud to be distorted by external electric fields. So, for example, if six groups of atoms could be located at each of two different sites, the model puts the different groups of atoms in the same spot simultaneously and then determines how well the different combinations fit. This repeats at a predetermined number of sites. Joining the best molecular groups or combinations—like snapping together Legos—yields a complete molecule with the best properties.

> "You can only do experiments on real molecules, but virtual techniques let you use non-real molecules to explore the molecular space," says Weitao Yang.

This approach quickly yields the molecular potential, but it doesn't necessarily map back to a molecule that can be made. For example, the best group at a particular site might be a combination of 13 percent of one molecule and 87 percent of another. This is impossible, of course, since only one molecule can occupy a single location, so the preferred molecule would be used.

"I think it's very elegant how Beratan and Yang approached the problem," says **Ursula Rothlisberger, PhD**, an associate professor of computer-aided inorganic chemistry at the Swiss Federal Institute of Technology in Lausanne, "But as a first step, it still has many limitations." For example, it can only create simple molecules, as Yang would agree. He and his colleagues are now refining it to handle more complex systems such as designing optical materials for electronic devices. They plan to extend their work to drug design as well.

"We want to uncover many new materials that researchers didn't know about before," Yang says. "This method explores the design space much more efficiently."

—Linley Erin Hall

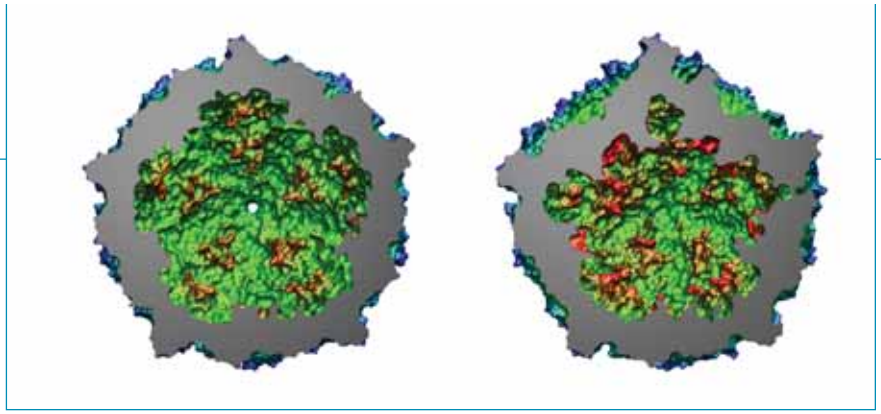
Whole Virus Simulation

Giving new meaning to the phrase computer virus, researchers have created a computer simulation of an entire biological virus comprising approximately one million atoms.

"It wasn't clear before that one could do a simulation of such a large living system at an atomic level and learn something from it," says **Klaus Schulten, PhD**, professor of physics at the University of Illinois at Urbana-Champaign. But when he and graduate students **Anton Arkhipov** and **Peter Freddolino** successfully simulated the satellite tobacco mosaic virus (STMV), they revealed some surprising features of the particle in the process. The work was published in the March 2006 issue of *Structure*, as a collaboration with virologists from University of California, Irvine.

Viruses must do two things: infect cells and transport their genetic material inside a stable container known as a capsid. In the case of the STMV, the capsid consists of 60 identical proteins produced by the virus's genome. Crystallographers who had imaged the small virus believed all 60 pieces were arranged in complete icosahedral symmetry. The computer simulation, however, showed this to be an incomplete picture of the virus.

Schulten and his colleagues started with the crystallography image of STMV and then allowed the atoms to move according to their physical properties. For just over 10 nanoseconds (broken into 10 million time steps), "we let the laws of physics take over," says Schulten. The



The collapse of the STMV capsid when simulated without the RNA core. The initial structure for this simulation (a) was the intact STMV capsid immersed in a drop of salty water (not shown). After only 5 nanoseconds of simulation, a prominent implosion of the capsid is observed (b). For both (a) and (b), a cut through the center of the capsid is shown. Courtesy of Klaus Schulten, Anton Arkhipov, and Peter Freddolino, University of Illinois at Urbana-Champaign.

result: Although the capsid remained generally spherical, some of the symmetry was lost. "The virus developed a belt around an equator of the sphere, and that belt engaged in a back and forth motion," Schulten says.

More important, simulation revealed that, unlike many other viruses, the STMV capsid is unstable without its RNA contents and depends on the RNA to assemble. "It seems that for this virus, the genomic material first aggregates into a sphere, and then recruits the 60 proteins to be a shell around itself," Schulten says. "This is opposite to what one expected."

Schulten and his colleagues hope that viral simulations of this type will help researchers understand how viral capsids shift from stable to unstable when they are infecting a cell. It's possible that one might be able to interfere in an infection at the point when the capsid breaks apart, he suggests. "We want to use information gained from simulations to protect people from viral infections."

In future projects, Schulten and his colleagues plan to simulate the poliovirus and other viral particles that are 4 to 10 times larger than STMV. Their success with STMV suggests that large scale simulations provide valuable, new information. "Had we done a partial simulation, we wouldn't have learned as much," he says.

—Katharine Miller

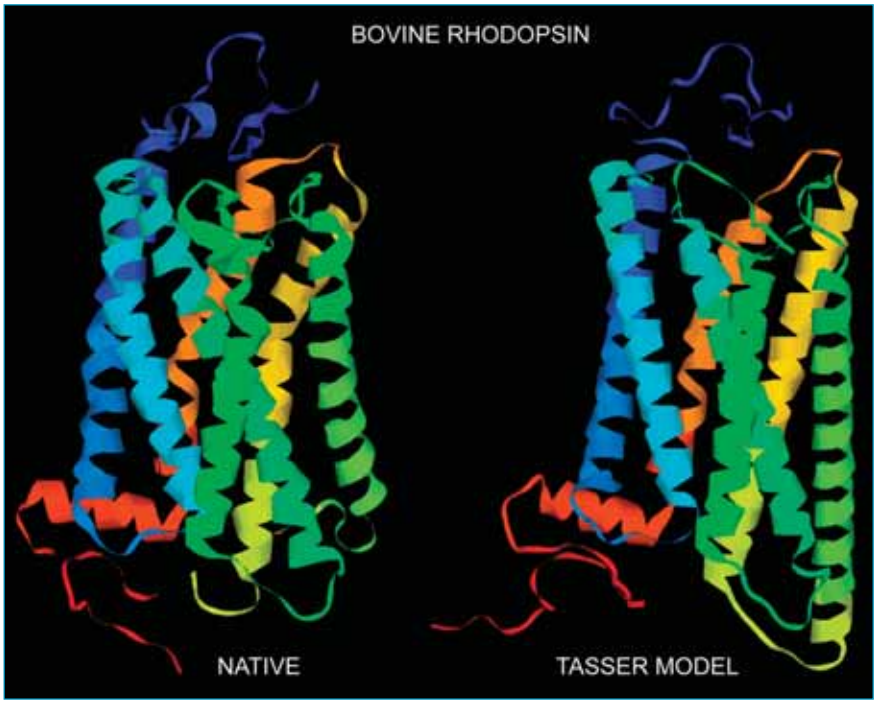
Predicting the Structure of Important Drug Receptors

If you want to find a Tab 'A' that will fit into a Slot 'B', you'll waste a lot of time if you don't know the shape of the slot. For scientists trying to design new drugs, that is sometimes the precise problem: They seek a molecule that will snug itself into a nook whose shape is unknown, difficult to determine, and capable of changing as the fit is induced.

Now, a new computational tool promises to help rescue researchers from the task of fitting square pegs into undefined holes. It models the structures of the largest family of cell surface receptor proteins in the human body: G protein-coupled receptors (GPCRs). These receptors are encoded by about five percent of human genes and are the targets of about 45 percent of all modern medications. The 3D structures of most GPCRs are unknown because the molecules are extremely difficult to work with. Like all proteins residing in cell membranes, they tend to fall apart when plucked from the membrane for analysis in a laboratory. Traditional approaches such as NMR and X-ray crystallography have only yielded a single GPCR 3D structure.

To sidestep the difficulties of the experimental approach, **Jeffrey Skolnick, PhD**, director of the Center for the Study of Systems Biology at the Georgia Institute of Technology in Atlanta, and his research team developed a structure prediction algorithm called TASSER. It takes whatever fragmentary information is known about a protein's structure—or can be reasonably inferred from knowledge about related proteins—and feeds it into a structure assembly algorithm that combines the data in different ways, searching for the most energetically stable configuration.

"By looking closely at structures that are similar, you should be able to enhance drug discovery by not only designing towards what you want, but away from everything else," says Skolnick, who estimates that of the 907 GPCRs in the human genome, TASSER has produced



Bovine Rhodopsin is a GPCR whose structure is known from experimental work. Here, that known structure compares favorably with that predicted by TASSER.