

NewsBytes

Modeling Sex's (Evolutionary) Appeal

Sex is a costly undertaking. Finding partners takes time and energy. Sexual contact can transmit disease. And if reproductive success is measured by how many genes you pass on, females would be better off reproducing asexually. But sex must be beneficial in some way—besides being fun—since so many plants and animals do it without going extinct. A new computational model described in the March 2, 2006, issue of *Nature* confirms one existing theory about why sex is advantageous on the genetic level.


“This is very difficult to measure in real organisms,” says **Ricardo Azevedo, PhD**, assistant professor of biology and biochemistry at the University of Houston. “But

things that take years or decades in the lab take only hours in the computer.”

Evolutionary biologists have posited several reasons for the success of sexual reproduction. The mutational deterministic hypothesis suggests that sex helps remove harmful mutations from a population because offspring receive genes from two parents. But the benefits of mutation purging can only overcome the costs of sex if the rate of harmful mutations is high. Multiple mutations must also be more harmful than would be expected from their individual effects, a condition known as negative epistasis. Azevedo’s model suggests that the mutational deterministic hypothesis may be true.

Azevedo along with **Christina Burch, PhD**, assistant professor of biology at the University of North Carolina, Chapel

Hill, and three graduate students created a model that treats each “organism” as a network of interacting genes. The network is expressed as a matrix of numbers (positive, negative or zero) representing the effect of each gene on the activity of every other gene in the organism. Large populations of sexually and asexually reproducing cyber-organisms (networks) were created with different rates of spontaneous mutation. In the first part of the simulation, each organism’s genes interact. Organisms that produce stable patterns of gene expression produce offspring in the second part of the simulation; unstable networks don’t—natural selection at work. When the populations reached equilibrium in their sensitivity to mutations, the sexual populations had become more insensitive to mutations than asexual popula-



“If the conditions in the model are real, then when sex evolves it creates conditions that help sustain itself over time,” Ricardo Azevedo says.

tions and had also evolved negative epistasis. Compared to asexual creatures, they more effectively purge negative mutations from the gene pool.

“If the conditions in the model are real, then when sex evolves it creates conditions that help sustain itself over time,” Azevedo says.

“The prevalence of sex begs to be studied,” comments **Andreas Wagner, PhD**, an associate professor of biology at the University of New Mexico. “To the extent that an abstract model can tell you anything about the evolution of sex, [Azevedo and Burch] have made an important contribution.” But, he says, he’d like to see the work confirmed in living systems.

Azevedo agrees this paper is a first step. He is trying to make the model more applicable to multicellular organisms while his collaborator, Burch, conducts experiments with viruses in order to confirm the model’s results.

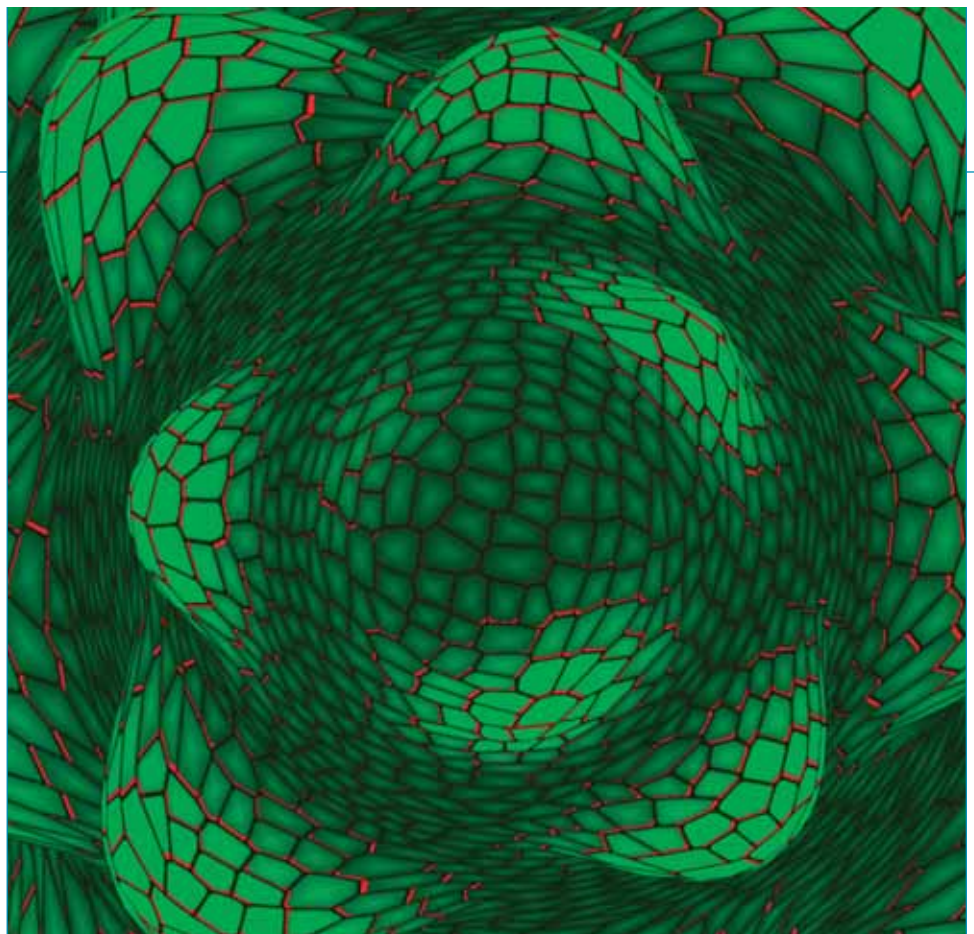
—Linley Erin Hall

Modeling Whorls of Leaves

The petals of every flower and the leaves sprouting from every plant stalk have characteristic arrangements, a phenomenon called phyllotaxis. For two centuries, botanists have puzzled over the force driving such regularity.

“If you want to understand how plants acquire their form, this is one of the very key questions,” says **Przemyslaw Prusinkiewicz, PhD**, professor in computer science at the University of Calgary in Alberta, Canada. He and his colleagues recently presented a new cellular-level computer model of the process. The work appeared in the January 31, 2006, issue of the *Proceedings of the National Academy of Sciences*.

Previous experimental work by Prusinkiewicz’s Swiss collaborators had shown that a plant hormone, auxin, plays a crucial role in phyllotaxis, as does a protein called PIN1, which regulates the transport of auxin. The team hypothesized that there was a feedback mechanism in which the distribution of



A computer simulation of the growing tip of a seedling of Arabidopsis thaliana, viewed from above. PIN1 proteins (red) facilitate transport of the plant hormone auxin (green), which in high concentrations promotes budding of leaves, seen here bulging out from the stalk. The feedback interaction of the protein and hormone produce the characteristic spiral pattern of leaves that form as the plant grows. Movies simulating the development of four different leaf arrangements can be seen in the paper’s supplemental material online at <http://www.pnas.org/cgi/content/full/0510457103/DC1#M1>.

auxin determined the location of the PIN1 proteins, the position of which, in turn, governed the flow of auxin.

They devised a computer model to test the theory quantitatively by simulating the properties of individual cells during the growth of a small flowering plant of the mustard family, *Arabidopsis*.

The model assumes that the tip of

basal tissues of the stem and flowed up to the growing tip, they were only able to get the leaf patterns observed in nature when they altered the model to have auxin produced locally at the tip.

They also found that by varying the parameters of the model, they could produce the leaf patterns found in other plants, which, Prusinkiewicz says, “rein-

By varying the parameters of the model, the researchers produced the leaf patterns found in various plants.

forces our belief that what we have shown is actually true, and it is not just true in *Arabidopsis*, but also in other plants.”

the stem develops at the same time that the pattern of leaves is forming, all of which relates to the pattern of cell division. The results of the simulation confirmed that the proposed interplay of auxin and PIN1 on the molecular level could produce the characteristic spiral leaf pattern found in *Arabidopsis*, but also yielded some surprises.

Though the researchers initially assumed auxin was produced in the

forces our belief that what we have shown is actually true, and it is not just true in *Arabidopsis*, but also in other plants.”

Prusinkiewicz characterizes their model as part of a broader inquiry into how genes and molecular level processes determine the macroscopic forms of organisms, which he calls “one of the most fascinating questions in developmental biology right now.”

—Louis Bergeron, MS

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.

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

“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**.

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.

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**

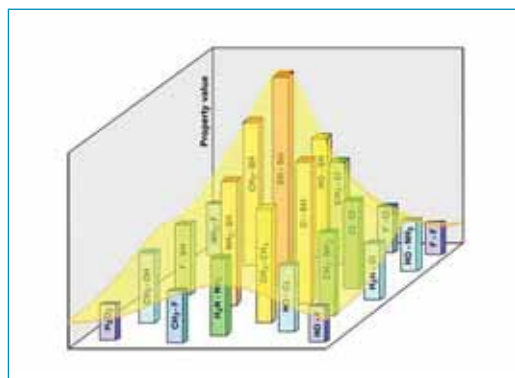
Whole Virus Simulation

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

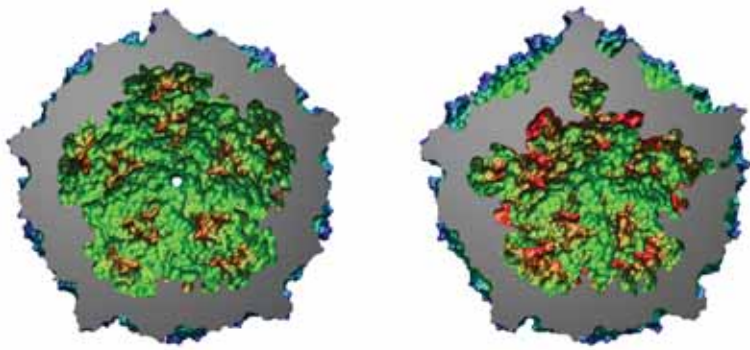
“It wasn’t clear before that one could do a simulation of such a large living system at an atomic level and learn something from it,” says **Klaus Schulten, PhD**, professor of physics at the University of Illinois at Urbana-Champaign. But when he and graduate students **Anton Arkhipov** and **Peter 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



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.



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

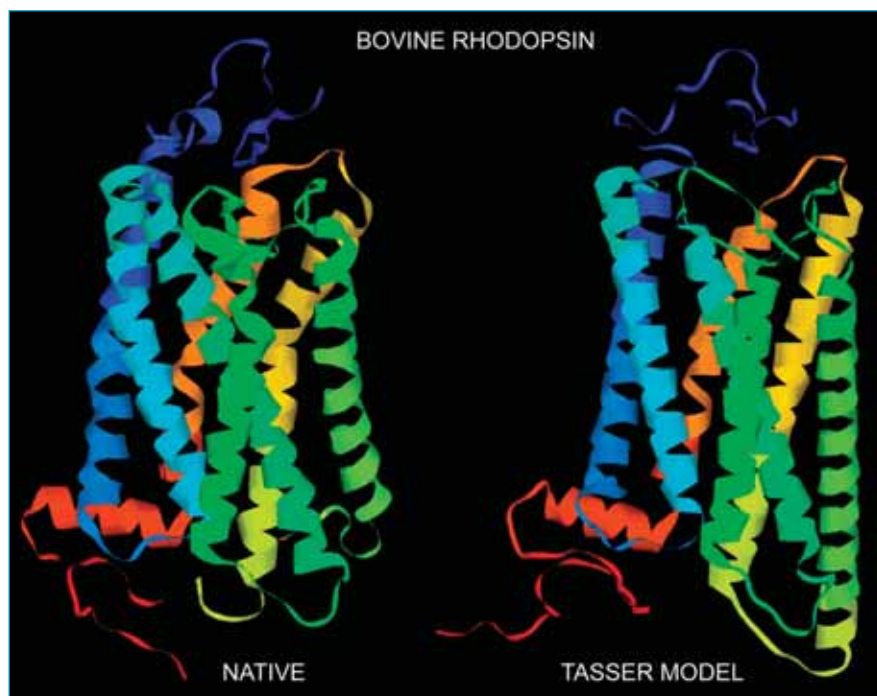
drugs, that is sometimes the precise problem: They seek a molecule that will snug itself into a nook whose shape is unknown, difficult to determine, and capable of changing as the fit is induced.

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.

820 models that are likely to be correct. The work was published in *PLoS Computational Biology* in February 2006.

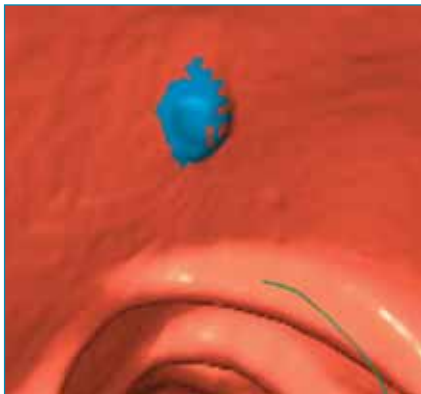
Because no one has determined the structure of these 900 proteins, an algorithm that can produce accurate predictive models should prove significant, comments **Harold Scheraga, PhD**, emeritus professor of chemistry and chemical biology at Cornell University.

Skolnick emphasizes that while he's confident most of the TASSER-generated models provide new insight into the GPCRs structures, he doesn't expect that many of the structures have been fully deciphered by this round of modeling. "What we're trying to do as best we can, is establish the plausibility of these [models] as hypothesis generators," he says, which should help guide drug development research away from dead ends and into productive avenues, where the tabs and slots of medication and receptor are most likely to mesh.

—**Louis Bergeron, MS**

Computation Competitions Take Off!

From all parts of the computational spectrum, researchers are duking it out: They are throwing their algorithms into the ring to see which one will out-perform all others on a particular task. Contests that feature algorithms for protein structure prediction, natural language processing, and computer-aided disease detection



Virtual colonoscopy image of a 0.8 cm polyp identified by CAD algorithm (shown in blue). Courtesy of R.M. Summers, MD, PhD, National Institutes of Health Clinical Center.

are giving researchers a jolt of adrenalin and moving these fields forward.

"When you have a field with a quantitative basis and competing approaches in which high performance is one of the main outcomes, it seems like a natural setting for having a competition," says **Ron Summers, MD, PhD**, senior investigator and staff radiologist in the department of radiology at NIH. "It's also beneficial to the field. The spirit of competition encourages hard work to solve difficult problems."

Protein-structure prediction has been competitive since 1994 when the CASP

"The spirit of competition encourages hard work to solve difficult problems," says Ron Summers.

(Critical Assessment of Techniques for Protein Structure Prediction) contest drew 34 groups to register. Since then, the biennial event has steadily grown in popularity: 263 groups are registered for the 2006 bout, including several that will rely only on *in silico* tools, without help from human instinct (See Human vs. Machine feature story in this issue).

This year, competitive natural language processing (NLP) gets a boost from one of the National Centers for Biomedical Computation. In conjunction with the fall meeting of the American Medical Informatics Association, i2b2 (Informatics for Integrating Biology and the Bedside) is extending an open invitation to anyone who wants to challenge their own NLP tools using real clinical records.

"Clinical data is not easily accessible to a lot of people who want to work on this type of data," says **Ozlem Uzuner, PhD**, assistant professor of information studies at the State University of New York at Albany. "I2b2 and its partners have put together these data and that's what makes this a unique opportunity."

The competition is two-pronged. Researchers compete to effectively remove patients' identifying information from clinical data. (Note: I2b2 has already removed the real infor-

mation and replaced it with fictional data to protect patient privacy). In addition, they will parse hospital discharge summaries to accurately extract information on patients' smoking status. The work will help set the stage for researchers to work with clinical data without violating patient privacy.

A computer-assisted polyp detection "bake-off" is also on the horizon. In a traditional bake-off, says Ron Summers, the cooks are given the ingredients and they compete to produce the best cake. In the CAD polyp bake-off, the American College of Radiology Imaging Network

(ACRIN) provides researchers with a data set consisting of CT colonoscopy scans from about 200 patients. The researchers then run their CAD systems using these data. About a dozen academic and commercial researchers have expressed interest in participating.

"Various researchers have been producing systems and claiming outstanding performance on very small data sets," says Summers. "It was competitive but not fair. It was like everyone deciding the terms of their own race." Since the ultimate goal is to help patients, results need to be standardized, Summers says. "We need to know which approaches are better so everyone can move toward that and improve their systems." Hence the CAD competition, which Summers hopes will be underway by November.

—**Katharine Miller** □

DETAILS

CASP:
<http://predictioncenter.gc.ucdavis.edu/>
Challenges in Natural Language Processing for Clinical Data (sponsored by i2b2 in conjunction with AMIA):
<http://www.i2b2.org/NLP/Main.php>
Virtual Colonoscopy CAD Bake-Off:
For more information, contact **Ron Summers: rms@nih.gov.**