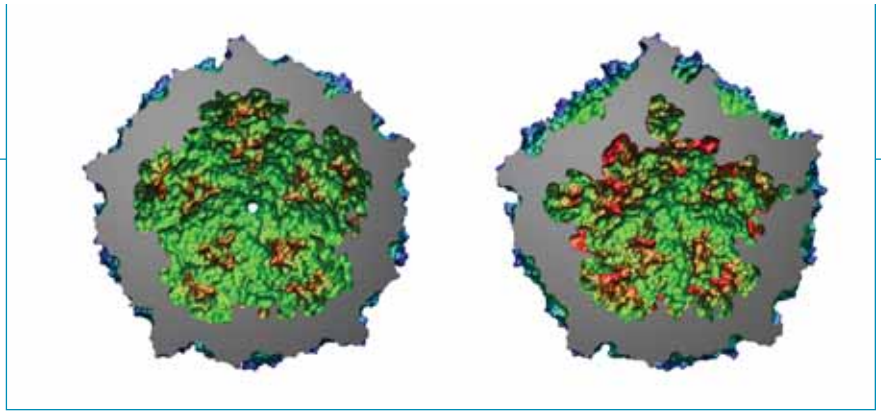


The collapse of the STMV capsid when simulated without the RNA core. The initial structure for this simulation (a) was the intact STMV capsid immersed in a drop of salty water (not shown). After only 5 nanoseconds of simulation, a prominent implosion of the capsid is observed (b). For both (a) and (b), 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

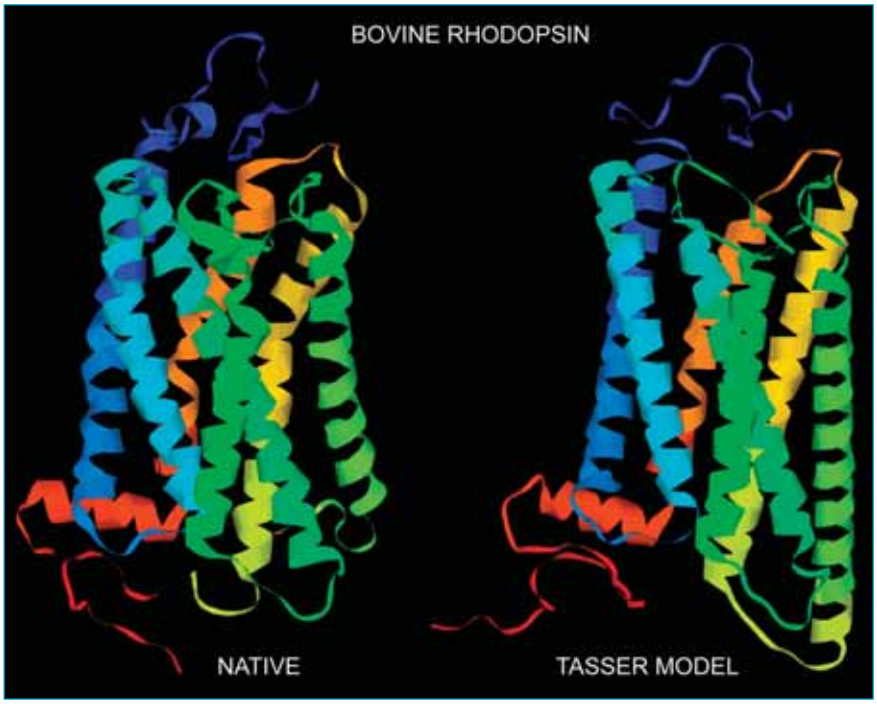
Predicting the Structure of Important Drug Receptors

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.

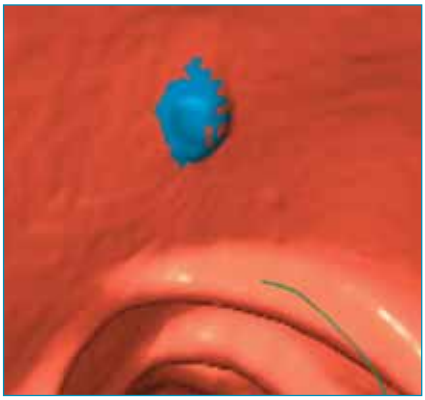
Because no one has determined the structure of these 900 proteins, an algorithm that can produce accurate predictive models should prove significant, comments **Harold Scheraga, PhD**, emeritus professor of chemistry and chemical biology at Cornell University.

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



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.

"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 (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).

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

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 information 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.