The EYES Have It:

Biomechanical Models Explore Disorders of the Eye

PLUS:
Untangling
Integrative
Analysis

contents

ContentsWinter 2013

Cover Art: Created by Rachel Jones of Wink Design Studio using: Eye image (courtesy of National Eye Institute) and fluid dynamics image (for credit, see page 21 image caption).

GuestEditorial

BY ANDREW D. McCULLOCH, PhD

Multiscale Modeling in Biomedical Research

In an era of increasingly comprehensive molecular characterizations of living systems, computation has emerged as a key technology to facilitate integrative understanding of biological mechanisms. Computation can be integrative in biomedical science in several different ways. Perhaps the best recognized of these is the role of computing for *information* integration. This is a central goal of bioinformatics. A related but distinct type of integration is *functional* integration. Once bioinformatics has organized and annotated the molecular components of the biological system, we can build functional interaction networks whether genomic, transcriptional, signaling, metabolic or physiological. These reconstructed networks serve as a foundation for developing comprehensive functionally integrated systems models of the cell or living system. This kind of functional integration is at the core of systems biology. A third way that computation can integrate is *structurally*, across physical scales of biological organization, from molecule to organism. This type of computational biology is often now commonly referred to as multiscale modeling. Indeed the Inter-Agency Modeling and Analysis Group[1] comprised of officers from nine federal agencies (plus the Canadian agency, MITACS) has formed the Multiscale Modeling Consortium (MSM) of over 100 funded investigators around this shared interest. Whereas systems biology modeling is very frequently data-driven and data-limited, multiscale modeling is more often physics-driven and compute-limited.

Of course these different types of integration are neither independent nor mutually exclusive. On the contrary they are interdependent. There is no single computational approach that can integrate comprehensively from molecule to organism, from genotype to phenotype or across interacting physiological subsystems. But generalizable paradigms are emerging that are making the prospect of models that span increasingly broad spatial and temporal scales of biology—blue sky pipedreams just a short time ago—seem much more feasible, at least for certain classes of problem.

Much of this progress has been driven by improvements in technology such as higher resolution three-dimensional measurements of biological structures made possible by improved microscopy and structural biology approaches. A recent example is the three-dimensional model of the cardiac myocyte calcium release unit shown below that was used to model the calcium "spark" that occurs when a single calcium release unit on the sarcoplasmic reticulum membrane opens.[2] This model was made possible by improved electron
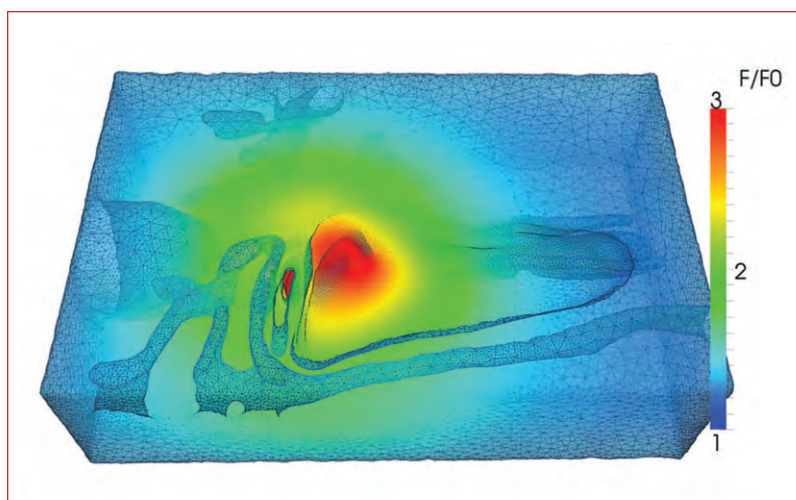
Multiscale model of a calcium release event (spark) from the sarcoplasmic reticulum of a cardiac myocyte. The model includes part of a t-tubule, sarcoplasmic reticulum and a mitochondrion. Color codes for predicted intensity of a fluorescent calcium probe. The dimension of the mesh is 1430×940×406 nm. Reprinted with permission from Hake, et al (2012).[2]

tomography techniques[3] and new methods for generating high-quality computational meshes.[4] New probes for localizing receptors and macromolecular complexes within these microanatomic domains will further improve these models.[5] Improved computational performance and novel algorithms are also allowing increasingly large-scale particle-based

1 See http://www.imagwiki.nibib.nih.gov.

2 Hake J, Edwards AG, Yu Z, Kekenes-Huskey PM, Michailova AP, McCammon JA, Holst MJ, Hoshijima M, McCulloch AD. Modelling cardiac calcium sparks in a three-dimensional reconstruction of a calcium release unit. *J Physiol* (Lond). 2012 Sep 15;590(Pt 18):4403–22.

3 Hayashi T, Martone ME, Yu Z, Thor A, Doi M, Holst MJ, Ellisman MH, Hoshijima M. Three-dimensional electron microscopy reveals new details of membrane systems for Ca2+ signaling in the heart. *J Cell Sci* 2009 Apr 1;122(Pt 7):1005–13.

4 Yu Z, Holst MJ, McCammon JA. High-fidelity geometric modeling for biomedical applications. Finite Elem Anal Des 2008;44, 715–723.

5 Shu X, Lev-Ram V, Deerinck TJ, Qi Y, Ramko EB, Davidson MW, Jin Y, Ellisman MH, Tsien RY. A genetically encoded tag for correlated light and electron microscopy of intact cells, tissues, and organisms. *PLoS Biol.* 2011 Apr;9(4):e1001041.

DETAILS

Andrew McCulloch, PhD, is professor of bioengineering and Jacobs School Distinguished Scholar at the University of California San Diego, La Jolla, CA.

models to be implemented. The particle-based Monte Carlo modeling of vesicular release from pre-synaptic neurons by Nadkarni and colleagues[6] is an exciting example. And improved algorithms together with higher resolution whole organ imaging are making possible large-scale organ models that can investigate the effects of fine spatial heterogeneities such as distributed scarring on clinical phenotypes such as cardiac arrhythmias[7].

Another way to push the spatio-temporal boundaries of multiscale models is through emerging new strategies for course-graining molecular models and for bridging molecular models to cellular scales. Fedosov and colleagues[8] recently demonstrated how coarse-grained techniques such as Dissipative Particle Dynamics can be used to model cell membrane dynamics and adhesive forces which are then used in multi-cellular models of red blood cell aggregation and in turn included in continuum models of whole blood non-Newtonian viscous properties. In 2009, Silva and Rudy[9] introduced another strategy for bridging from all-atom molecular dynamics and molecular electrostatics simulations to whole cell models by using a Markov model of a delayed rectifier potassium channel as the intermediate and then exploring the mechanisms by which disease-causing mutations can affect clinical electrophysiological phenotypes.

I have focused on new approaches to extending multiscale models to represent details at cellular mesoscales and on bridging molecular to cellular models. But there are other opportunities too. The span of important temporal scales in biomedicine is even larger than that of the spatial scales. There is tremendous potential to extend multiscale models from the time-scale of physiological responses to those longer time-scales of growth, remodeling, and the natural history of disease and aging. The challenges will ensure an exciting and robust future for integrative multiscale modeling in biomedicine. □

6 Nadkarni S, Bartol TM, Sejnowski TJ, Levine H. Modelling vesicular release at hippocampal synapses. *PLoS Comput Biol.* 2010;6(11):e1000983.

7 Winslow RL, Trayanova N, Geman D, Miller MI. Computational medicine: translating models to clinical care. *Sci Transl Med.* 2012 Oct 31;4(158):158rv11.

8 Fedosov DA, Pan W, Caswell B, Gompper G, Karniadakis GE. Predicting human blood viscosity *in silico. Proc Natl Acad Sci USA.* 2011 Jul 19;108(29):11772–7.

9 Silva J, Pan H, Wu D, Nekouzadeh A, Decker K, Cui J, *et al.* A multiscale model linking ion-channel molecular dynamics and electrostatics to the cardiac action potential. *Proc Natl Acad Sci USA* 2009;106(27):11102.

NIH Announcement

NIHAnnouncement

BY KATHARINE MILLER

Big Data Gets Big Support

In December, the National Institutes of Health (NIH) approved a new trans-NIH initiative called Big Data to Knowledge (BD2K). With this action, the NIH signals its intention to invest a significant amount of money in computational research above and beyond the ongoing investment by individual institutes, says **Mark Guyer**, **PhD**, deputy director of the National Human Genome Research Institute (NHGRI).

One significant piece of BD2K will be new "big data" centers of excellence for which RFAs (Requests for Applications) will be announced in FY 2013 and funding provided in FY 2014-15. While total dollar amounts and sizes of these center grants are yet to be firmed up, **Lawrence Tabak**, **DDS**, **PhD**, deputy director of NIH indicated in his December presentation that up to 15 investigator-initiated centers and 2 to 5 NIH-specified centers could be funded under the program. In addition, BD2K will fund significant new training opportunities in biocomputing.

While the NIH does not call BD2K a "replacement" for the National Centers for Biomedical Computing (NCBCs)—which are set to term out in 2014/2015—it is certainly a follow-on to that program, Guyer says. And it is likely that several current NCBC grant recipients will be among those seeking funds through the new initiative, says **Russ Altman**, **MD**, **PhD**, professor of genetics, bioengineering and medicine at Stanford University, and principal investigator for Simbios, the National Center for Physics-Based Simulation of Biological Structures.

The NIH will host a series of workshops over the course of the next several months to figure out in specific detail what the BD2K grant programs will be and how they will complement existing programs, Guyer says. These workshops will help NIH gather the full spectrum of community opinion in short order, he says. "We are planning on an accelerated basis because the money starts in FY14." □

DETAILS

The December 7, 2012 BD2K announcement is posted at http://acd.od.nih.gov/meetings.htm. NIH will create a BD2K web site in the near future and all announcements about the initiative will also be provided at the NIH guide to grants and contracts web page, http://grants.nih.gov/grants/guide/.

BY JOY P. KU, PhD, DIRECTOR OF SIMBIOS

Journey to the NIH: Insights and Inspirations from the 2012 NCBC Showcase

If he were a graduate student now, Francis Collins would be studying computational biology. That's what the Director of the National Institutes of Health (NIH) told a rapt audience at the November 2012 National Centers for Biomedical Computing (NCBC) Showcase. The field of computational biology is "raining opportunities," Collins said.

It was a message welcomed by the six distinguished postdoctoral fellows attending from Simbios, one of the six NCBCs. "Hearing that our field is at the forefront of research gave me a boost in continuing to pursue this line of study," says **Saikat Pal**, **PhD**, a Simbios postdoctoral fellow who works on multiscale modeling of the knee.

But the showcase inspired the postdocs in other ways as well. In addition to sharing their research in a unique setting (see details box), they stepped outside their everyday research focus to take a look at the entire field and get their first inside view of NIH grant-making.

"Giving postdocs the opportunity to attend this kind of event is an essential part of their training," says **Russ Altman**, **MD**, **PhD**, principal investigator of Simbios. "We are preparing the next generation of leaders in the field."

Eye-Opening Breadth

Even in an interdisciplinary field like computational biology, it's all too easy for researchers to develop blinders as they hone in on a specific research focus. With presentations about electronic medical records, medical imaging, genetics, and ontologies, the NCBC Showcase broadened that view for **Lee-Ping Wang**, **PhD**, a postdoctoral fellow researching molecular dynamic force fields. "It was definitely a real eye-opener to see all the different types of research," he says.

But it was the distribution of research areas represented—or not—that left the biggest impression on **Diwakar Shukla**, **PhD**, a postdoctoral fellow in bioengineering. Given that so many people work on molecular dynamics simulations—as he does—Shukla says he was surprised there was only one NCBC focused on it. **Jenelle Bray**, **PhD**, postdoctoral fellow in structural biology, made a similar observation. "None of the other NCBCs' talks were even close to what I do," she says. It's a situation that points to both current NIH priorities and the challenges of supporting research in such a diverse field.

Getting Behind the Scenes

For the postdocs, the NCBC Showcase was also a first behind-the-scenes view of what goes on in the minds of NIH decision makers. For example, **Gert Kiss**, **PhD**, a postdoctoral fellow in chemistry, was impressed that, in addition to research for research's sake, program officers care about the economics of NIH investments. "I was surprised to learn that they want at least some of their investments to eventually produce a profit or cost-savings and, in that way, help sustain the field."

And **Enrique Rojas**, **PhD**, a postdoctoral fellow studying bacterial growth, was struck by the collaborative relationship between the NIH and the NCBC principal investigators. "It's not as if you're trying to get their money and they're conservatively doling it out. Ideally, the relationship between researchers and the NIH would be one of cooperation and collaboration, and I did get that sense. That was encouraging."

As they return to their labs at Stanford, the postdocs bring with them these inspirations and insights. "It wasn't a normal science conference," Rojas says of the NCBC Showcase. And perhaps that's the point. □

DETAILS

Simbios postdoctoral fellows presented their latest research developments at the NCBC Showcase:

"Torsion Angle Normal Mode Analysis"
—Jenelle Bray

"Protein Engineering through Computational Design: From Proof of Principle to Real Life Applications"
—Gert Kiss

"Multiscale Modeling to Evaluate the Mechanisms Underlying Patellofemoral Pain Syndrome"
—Saikat Pal

"Chemical and Mechanical Regulation of Bacterial Cell Wall Expansion"
—Enrique Rojas

"Mapping the Conformational Landscape of G-Protein Coupled Receptors Using Novel Computational Paradigms"
—Diwakar Shukla

"Simple and Systematic Parameterization of a Polarizable Water Model"
—Lee-Ping Wang

Abstracts for these posters can be downloaded at http://meetings.nigms.nih.gov/assets/ncbc/NCBC_Showcase_posters.pdf.

Simbios (http://simbios.stanford.edu) is the National Center for Physics-Based Simulation of Biological Structures at Stanford.
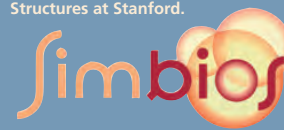
FOLLOW THE MONEY:
Big Grants in Biomedical Computing

By Kristin Sainani, PhD

Several biomedical computing projects received big money in the fall of 2012. If there's one clear winner, it's "Big Data": three of the grants focus on building new computational infrastructure and tools for dealing with massive biological datasets. A fourth grant focuses on building new tools for multiscale modeling.

Advancing Bioinformatics in Africa

Genomics research in Africa will receive a major boost thanks to a $10 million grant to establish a sustainable bioinformatics network on the continent. The project, H3ABioNet, is part of the H3Africa initiative—a joint venture of the National Institutes of Health (NIH) and the Wellcome Trust to promote large-scale genomics research in Africa.

"It's very important not only to generate data but also to have the infrastructure and know-how to make sense of them," says **Victor Jongeneel**, **PhD**, director of the High-Performance Biological Computing program at the University of Illinois. "The idea is to make sure that the data produced by H3Africa projects are analyzed in Africa and not shipped out for analysis to groups in Europe or the U.S. We have a moral imperative to make sure that the benefits of the research and the credit for the work are reaped locally." The H3ABioNet team includes researchers from the University of Illinois, Harvard, and the University of Cape Town.

H3ABioNet will help set up the computational infrastructure needed to analyze high-throughput sequence data, including establishing and providing access to high-performance computing centers, installing analysis software, and facilitating data storage. H3ABioNet will also train local scientists how to handle and analyze high-throughput data, including establishing internships to provide practical, hands-on experience. Centers in the H3ABioNet network will go through an accreditation process that will ensure their proficiency in analyzing human genomic data.

Though the focus of H3Africa is on human genomics, the H3ABioNet network could provide support for any kind of computationally intensive biological research in Africa. "I'm hoping that the capacity devel-

> "It's very important not only to generate data [in Africa] but also to have the infrastructure and know-how to make sense of them," says Victor Jongeneel.

opment that will be funded in this project will also have an impact on fields other than human genetics," Jongeneel says.

Making Sense of Metabolomics Data

Metabolomics (the systematic study of the end products of the cellular processes that allow cells to grow and reproduce) is coming of age—leaving scientists with a mountain of new "omics" data to decode. But a new data repository and coordination center at the University of California, San Diego, will help deal with the data deluge, thanks to a $6 million grant from the NIH.

"In the last few years, mass spectrometry technology has matured to a point where one can do reasonably robust medium-throughput to high-throughput metabolomics," says principal investigator **Shankar Subramaniam**, **PhD**, professor of bioengineering. "We've been funded to help figure out, 'what do we do with these data?'"

The new center will serve as the data hub for research cores financed through the NIH Common Funds Metabolomics Program and similar metabolomics research initiatives. The center will provide a national repository for metabolomics data and will provide publicly available, user-friendly tools for data access and analysis. Among the challenges, Subramaniam's team will develop strict standards for data and metadata. "Data without meta-data is almost al-
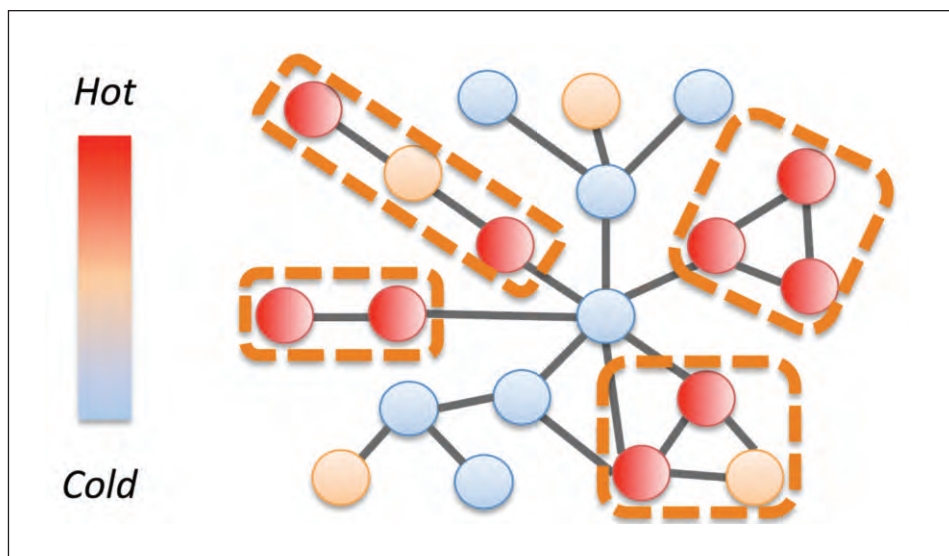
***Pattern Extraction.** HotNet is an example of a statistically robust algorithm for extracting information from big genomic datasets. "Hot" gene networks are more correlated with clinical phenotype. Courtesy of: Eli Upfal, Brown University.*

ways useless," Subramaniam says.

Subramaniam's team is developing robust statistical methods for "metabotyping" diseases—identifying metabolic patterns that correlate with disease states. They are also working on complex algorithms for reconstructing metabolomics networks and integrating metabolomics data with data from proteomics and transcriptomics. "This data integration is a very complex research task; there's no plug and play application that you can buy off the shelf," he says.

Rigorously Mining Genomic Data

From personal and cancer genomes to social networks to online buying habits, we have entered the era of Big Data. Computer scientists have developed efficient and successful machine learning algorithms to mine these data for patterns. But no one quite knows: how reliable are the results? So, a team from Brown University has received a $1.5-million "Big Data" grant from the National Science Foundation and the NIH to develop rigorous statistical tools that answer this question. Their work focuses on data from the Cancer Genome Atlas.

"Machine learning is very popular today. But we don't have statistical guarantees on the quality of our results," says **Eli Upfal**, **PhD**, professor of computer science and principal investigator on the project. "We want to build techniques that will still be efficient and practical but also would quantify the quality of the results." Fellow computer science professors **Ben Raphael**, **PhD**, and **Fabio Vandin**, **PhD**, will help lead the effort.

Genomic data are large, complex, and noisy on many levels, so they are a good prototype for testing Big Data tools, Upfal says. His team has already built a tool called HotNet that helps identify—with high statistical confidence—pathways of mutated proteins that are involved in cancer. Upfal's team also aims to develop statistical tools to answer the question: how big of a sample does one need to answer a particular question in genomics? Ultimately the tools will have applications in many domains beyond genomics, Upfal says.

Bridging Gaps in Multiscale Modeling

Multiscale models in neurobiology can help scientists understand the brain from the small molecule to the tissue level. But to realize the promise of these models, scientists first must bridge critical gaps between biologists and computer scientists, as well as between computer scientists working at different scales. A new $9.3-million NIH center at the University of Pittsburgh—the Biomedical Technology Research Center—aims to do just that.

"Our first goal is to bridge the gap between experiments and computations. A first step toward this goal is to reach out to those people already doing experiments and assist them in solving their neurobiological problems," says principal investigator **Ivet Bahar**, **PhD**, professor of computational and systems biology. The center, which is a joint collaboration between the University of Pittsburgh, Carnegie Mellon University, Pittsburgh Supercomputing Center, and the Salk Institute, will develop multiscale models for five different driving biomedical projects across several institutions. "The tools we build will be tailored to their needs," Bahar says. The projects focus on understanding brain signaling and may lead to new treatments for neurological and behavioral disorders.
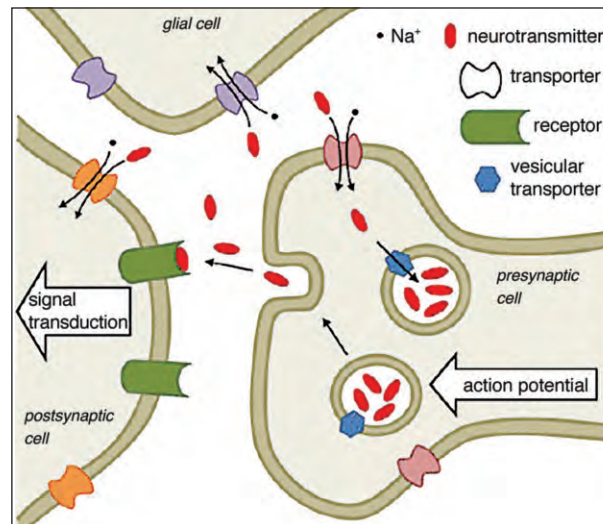
Modeling the Synapse. *The new Biomedical Technology Research Center at the University of Pittsburgh aims to combine computational models and experimental data to understand how messages are sent across synapses; and how that affects the brain. Courtesy of: Ivet Bahar, University of Pittsburgh.*

The center will also foster collaborations among modelers at different scales, such as the molecular, synapse, and brain tissue levels, she says. Currently, she says, "There is a serious disconnect. We would really like to overcome that," Bahar says. For example, her team will use molecular simulations to calculate parameters that can be fed into cell simulations created by other researchers. Those cell modelers will in turn provide feedback that can improve her team's molecular models. □

Cancer's Drivers. *Scientists applied HotNet to The Cancer Genome Atlas data for glioblastoma multiforme (a type of brain tumor). The algorithm extracted subnetworks of highly mutated genes. Courtesy of: Eli Upfal, Brown University.*
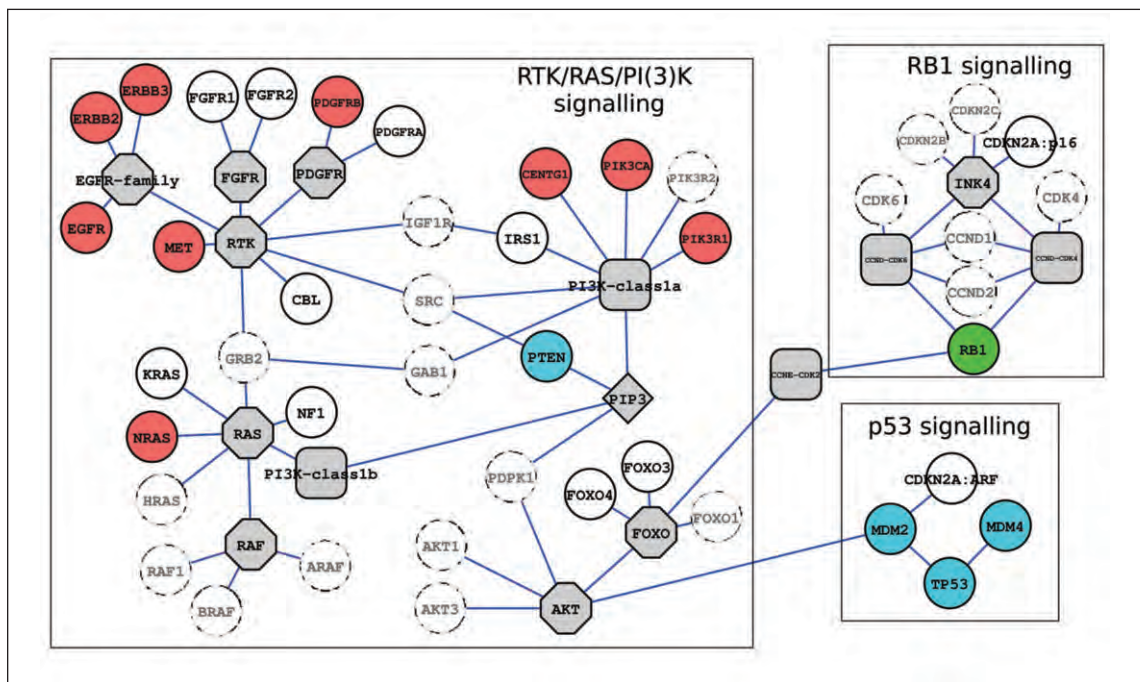
COMPUTING BETTER ENZYMES:
Optimizing Directed Evolution

By Kristin Sainani, PhD

Enzymes are among nature's crowning achievements: they accelerate chemical reactions, making life possible. People have co-opted natural enzymes for industrial use for thousands of years (think cheese-making). But it's only recently that scientists have been able to create made-to-order enzymes for applications ranging from detoxifying deadly nerve gas to converting waste into fuel.

amples of directed evolution of enzymes for industrial processes," says **Gert Kiss**, **PhD**, a postdoc at Stanford University, who did his doctoral work in the lab of **David Baker**, **PhD**, professor of biochemistry at the University of Washington and a pioneer in directed enzyme evolution.

Several advancements are fueling recent progress in the field. Chief among these is the coupling of directed evolution with com-

using a replication process that randomly introduces errors (a procedure called error-prone PCR); they then select or screen the resulting variants for increased catalytic activity. Enzymes with higher catalytic activity undergo a second round of mutagenesis and selection/screening; and the process is repeated (typically around 10 to 20 times). The number of possible variants is astronomical (20^{250} for an average-sized

> "You won't get around the actual experiments with these [computational] approaches, but by providing more and more rational input, the process becomes less random and thus more effective," Kiss says.

In a process called "directed evolution" scientists re-enact natural evolution in the laboratory: they iteratively mutate an enzyme and select for mutants with the desired feature. Within months, directed evolution can increase an enzyme's ability to catalyze a particular reaction by as much as 1000-fold—and sometimes even beyond. "In the past five years alone, there have been over 60 publications containing ex-

putation. Though simulating directed evolution exhaustively *in silico* remains beyond reach, computation can help narrow the search space for directed evolution; guide mutagenesis; and create *de novo* enzymes for catalytic activities that don't exist in nature.

Narrowing the Search Space

Directed evolution follows two steps: researchers mutate the starting enzyme

protein), so even with large libraries of mutants, one can cover only a tiny fraction of the search space. "You're shooting with a shot gun in a dark room and you're just hoping to hit something," Kiss says. When structural information is available, computational approaches can reduce the search space and improve the odds of a hit. Rather than randomly mutating the whole protein, scientists focus only on those amino acids that are likely to yield dividends, such as those in the active site. "You won't get around the actual experiments with these approaches," Kiss says, "but by providing more and more rational input, the process becomes less random and thus more effective."

For example, a 2012 paper in *Chemistry and Biology* described the directed evolution of human PON1, a protein that in its native form can weakly detoxify the nerve gas sarin. Researchers used the computational docking program AutoDock to explore the effects of different mutations on substrate binding—and identified eight key amino acids in the active site likely to affect sarin binding. These were subjected to repeated rounds of site-
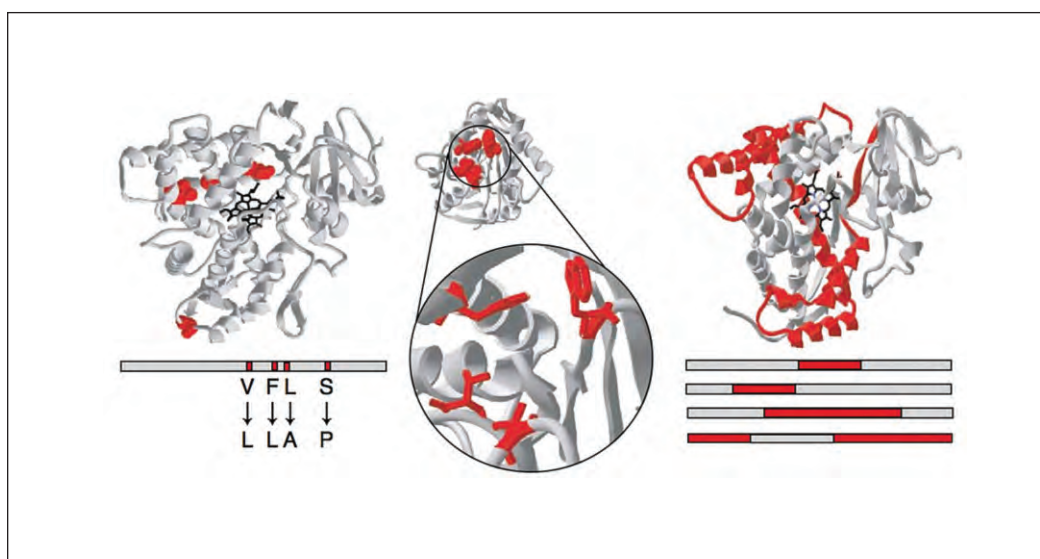
Triple Threat. *Scientists use three strategies for creating protein libraries in directed evolution: a) Random mutagenesis across the full sequence; b) Targeted mutagenesis focused on a specific site (often the active site); and c) Protein sequence recombination for the replacements of entire segments. Reprinted with permission from Bloom, JD,* **et al.,** *Evolving strategies for enzyme engineering.* **Curr Opin Struct Biol** *2005, 15, 447-452.*

specific mutagenesis. The resulting laboratory-evolved variants had up to 3400-fold increased activity (relative to wild type), enough to block the action of sarin on human target proteins for 24 hours.

To make the problem computationally tractable, modeling programs must rely on certain approximations. For example, many approaches assume a rigid protein backbone when in fact, proteins are "like spaghetti in a bowl—continuously vibrating and breathing—capable of adapting to their environment," Kiss says. Though they can't fully model backbone flexibility (as this requires massive computer resources), Kiss and colleagues have applied a program called RosettaBackrub that can incorporate backbone flexibility on a smaller scale.

Guiding Mutagenesis

Using Backrub, as well as several other computational strategies for guiding directed evolution, Kiss and colleagues increased the catalytic activity of the enzyme KE70 (a computer-designed enzyme, see below) by 400-fold. The work is described in a 2011 paper in the *Journal of Molecular Biology*. "In many cases, the mutations were suggested computationally," says first author **Olga Khersonsky**, **PhD**, a postdoc in David Baker's lab at the University of Washington. "Many other labs are also now using computation to guide their directed evolution processes."

Directed evolution can enhance protein function, but this is often at the cost of protein stability. "It really comes down to what you select for," Kiss says. "If you care about improving catalytic activity, you might end up losing thermal stability." Sequence analysis can help here: scientists may focus their mutagenesis only on "hotspots" of mutation and avoid mutations in more conserved regions. The fact that nature has disfavored changes in these areas suggests that they are destabilizing. In a 2010 paper in *ChemBioChem*, researchers mutated four residues in the active site of an esterase enzyme. But they only allowed substitutions with amino acids that commonly appear in these sites in other enzymes from the same family (as determined by the alignment program 3DM). Indeed, control experiments con-

firmed that the strategy significantly increased their hit rate.

Making Enzymes from Scratch

Directed evolution can only work if there is a starting enzyme that has at least a weak ability to catalyze the reaction of interest. When no natural enzymes exist,

*"We are generating enzymes for which there was no actual evolutionary pressure in nature," Kiss says.*

scientists can now create them from scratch on a computer. "We are generating enzymes for which there was no actual evolutionary pressure in nature," Kiss says. Though these designed enzymes display only weak catalytic activity ("we're very good at making bad enzymes!" Kiss says), they provide a starting point for directed evolution.

Baker's lab has provided some of the first examples of *de novo* enzyme creation. They first build an idealized active site: using quantam mechanics calculations, they determine which amino acid groups are needed—and in what orientation—to stabilize the transition state of the chemical reaction. Then, using the RosettaMatch program, they search through a database of over 86,000 crystal structures to geometrically fit this theoretical active site (also called a theozyme) into a groove or cavity on an existing protein. "The challenge is that it's a huge search space. Finding a way to efficiently search through this library of scaffolds to find a good geometry is very challenging," says **Daniela**

Grabs-Röthlisberger, **PhD**, a cofounder of Arzeda Corporation, which uses Baker's technology to make designer enzymes. Finally, they graft the theozyme onto the protein scaffold *in silico*.

Most of the proteins turn out to be duds: they fail to express, fold, or show the desired activity, or they aggregate in solution. But a few percent work, Kiss says. For example, Baker's lab created enzymes to catalyze the Kemp eliminase reaction (for which no natural enzymes exist). They came up with 57 designs in 17 different scaffolds, 8 of which showed Kemp eliminase activity. Three, including KE70, were further optimized by directed evolution to increase their activities up to 2000-fold. "This is really an uphill battle. But it's so cool to make progress and to eventually find a way through," Kiss says. The Kemp elimination is a model reaction that has no practical applications for industry or medicine. But, Kiss says, "we're now starting to go from proof of principle to a place where we're starting to apply these methods to real problems." □

Do-It-Yourself Enzymes. To make your own enzyme: (1) design an idealized active site (theozyme) using quantam mechanics; (2) use RosettaMatch to find a protein scaffold that's a geometric match for the theozyme; and (3) graft the theozyme into the protein scaffold in silico *using RosettaDesign. Courtesy of: Gert Kiss, Stanford University. Reprint from Kiss G, Olcum NC, Moretti R, Baker D, Houk KN, "Computational Enzyme Design,"* Angew Chem Intl Ed, In press.)
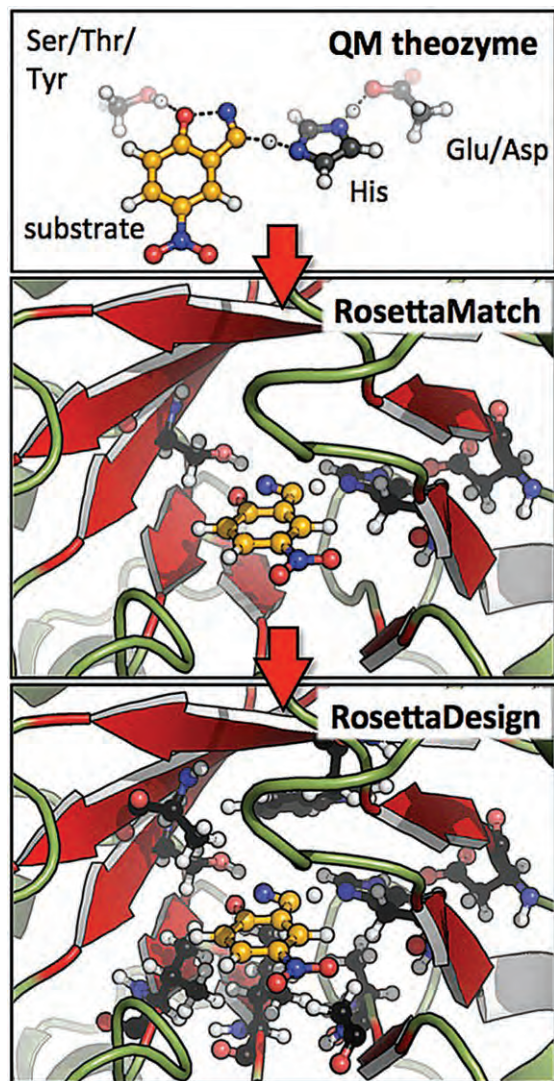
13 years ago

Markus Covert, PhD,
read a *New York Times*
article that changed his life.

Untangling
INTEGRATIVE

By Alexander Gelfand

The article quoted a prominent microbiologist who suggested that the ultimate test of one's understanding of a simple cell wouldn't be to synthesize an artificial version of the thing, but rather to build a computer model of it—a model that could predict all of the proteins expressed by the cell's genes, their behaviors and interactions. "I think about that article every day," says Covert, who was a graduate student at the time and is now an assistant professor of bioengineering at Stanford.

To be fair, he's done more than just think. In 2012, Covert himself appeared in the *Times*, garnering widespread attention for having created a computational model of the bacterium *Mycoplasma genitalium*. Covert's whole-cell model simulated all of the microorganism's molecular components and their interactions over the course of its life cycle; accounted for the function of every annotated gene product; and predicted a wide range of behaviors with a high degree of accuracy.

It was also a model of many parts. Twenty-eight, to be exact—28 individual submodels, each describing a different cellular function (ribosome assembly, cell division, DNA repair, etc.). Those submodels are defined by thousands of parameters and compute a comparable number of unknowns represented by 16 different categories of cell variable (chromosome, mass, geometry) that in turn represent different data types. "Chromosome," for example, might refer to the degree of chromosomal replication, or the location of every single protein on the chromosome; "mass" might refer to the mass of DNA or of proteins; "geometry" might refer to cell radius or shape. Over the preceding decade, Covert explains, he and his colleagues had come to the conclusion that no single computational approach would suffice to model a whole cell; instead, the task would require "a lot of different approaches"—approaches that would somehow need to be integrated into a unified whole.

That integrative ethos is becoming increasingly common. This is true whether the problem under investigation requires combining disparate data types, such as the ones flowing from next-generation high-throughput sequencing technologies; or simulating systems that contain many different moving parts, each one amenable to different mathematical treatment. And the trend will only intensify as integrative modeling and analysis becomes the *modus operandi* of biomedical research in general. As **Bernhard Palsson**, **PhD**, Covert's former doctoral advisor at the University of California, San Diego, says, "It's clear that over

> ## "It's clear that over the next 10 years, this kind of activity will take center stage in the life sciences," Palsson says.

the next 10 years, this kind of activity will take center stage in the life sciences." And it will likely come in many forms.

Integrating Data about Gene Regulation

Mark Gerstein, **PhD**, and his colleagues at Yale University have been leading contributors to the ENCODE project, which aims to delineate all of the functional elements in the human genome. Using techniques that they originally developed in model organisms such as worms and mice, Gerstein and his team recently employed several different types of ENCODE data to build an integrative model of transcription that can predict gene expression based on the presence of particular regulatory elements. Among other things, ENCODE has established that as much as 18 percent of the human genome, most of which was once considered to be "junk" DNA, helps regulate the 2 to 3 percent that actually codes for proteins.

The team began by building individual models that correlated expression with different kinds of regulators—most notably, the transcription factors and histone modifications that are found at transcription start sites directly upstream from sets of genes, and which exert considerable influence over whether or not those genes are transcribed and therefore expressed. Transcription factors are proteins that activate or repress the flow of genetic information from DNA to messenger RNA. Histones—the spools around which DNA winds within the chromosome—are modified in various ways that also affect gene regulation.

The models used machine-learning methods to look at the values of thousands of these regulators in small regions around the transcription start sites; multiplied them by coefficients in order to weight their relative significance; and added them all together to create accurate predictors of gene expression. "That's the stuff of integration right there," says Gerstein, who is professor of biomedical informatics, molecular biophysics and biochemistry and computer science. By comparing the relative impact of the various regulators, Gerstein was able to determine which transcription factors
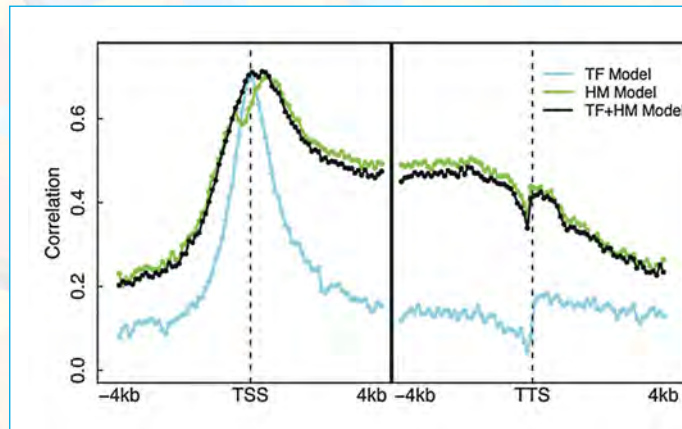
This graph compares the prediction accuracy of three models: one based on transcription factors (TF Model), one based on histone modifications (HM Model), and one incorporating both (TF+HM Model). Surprisingly, no accuracy is gained by combining the TF and HM models. Reprinted by permission of Oxford University Press from Cheng C, Gerstein M, Modeling the relative relationship of transcription factor binding and histone modifications to gene expression levels in mouse embryonic stem cells, Nucleic Acids Res 40(2):553-68 (2012).

MODELING

and histone marks were most important to prediction. As reported in a recent paper in *Nucleic Acids Research*, the team found distinct differences in predictive strength based on location, with transcription factors achieving their highest predictive power in a small region of DNA centered around the transcription start sites, and histone modifications demonstrating high predictive power across a wide region around the genes. As a final step, Gerstein and his colleagues built a model that included both histone modifications and transcription factors, but discovered that integrating the two did not improve accuracy. "They're actually somewhat redundant; you can't do better by combining them," says Gerstein—a surprising result that may help illuminate the basic biology of transcriptional regulation.

Interestingly, Gerstein doesn't consider the integrative aspect of the undertaking to have been especially challenging. "In a sense, the integration is carried out in the actual mathematical machinery as it's put together," he says, referring to the automated manner in which the machine-learning algorithms go about sorting and multiplying, adding and predicting. Instead, most of the heavy lifting comes earlier: before the data on the various regulators can be fed into the models, they must first be normalized and placed in the same coordinate system, put in the correct format and properly scaled. "There's a huge amount of upstream work [required] to be able to do

"There's a huge amount of upstream work [required] to be able to do this integration," Gerstein says.

this integration," Gerstein says, adding that the project is "a nice case study" of "the overall process of putting all this information together and making predictions."

Integrating a Whole Cell Model

The idea that the "integrative" part of an ambitious integrative analysis project should turn out to be fairly straightforward might seem surprising. But it's hardly uncommon.

For example, yoking together 28 indi-
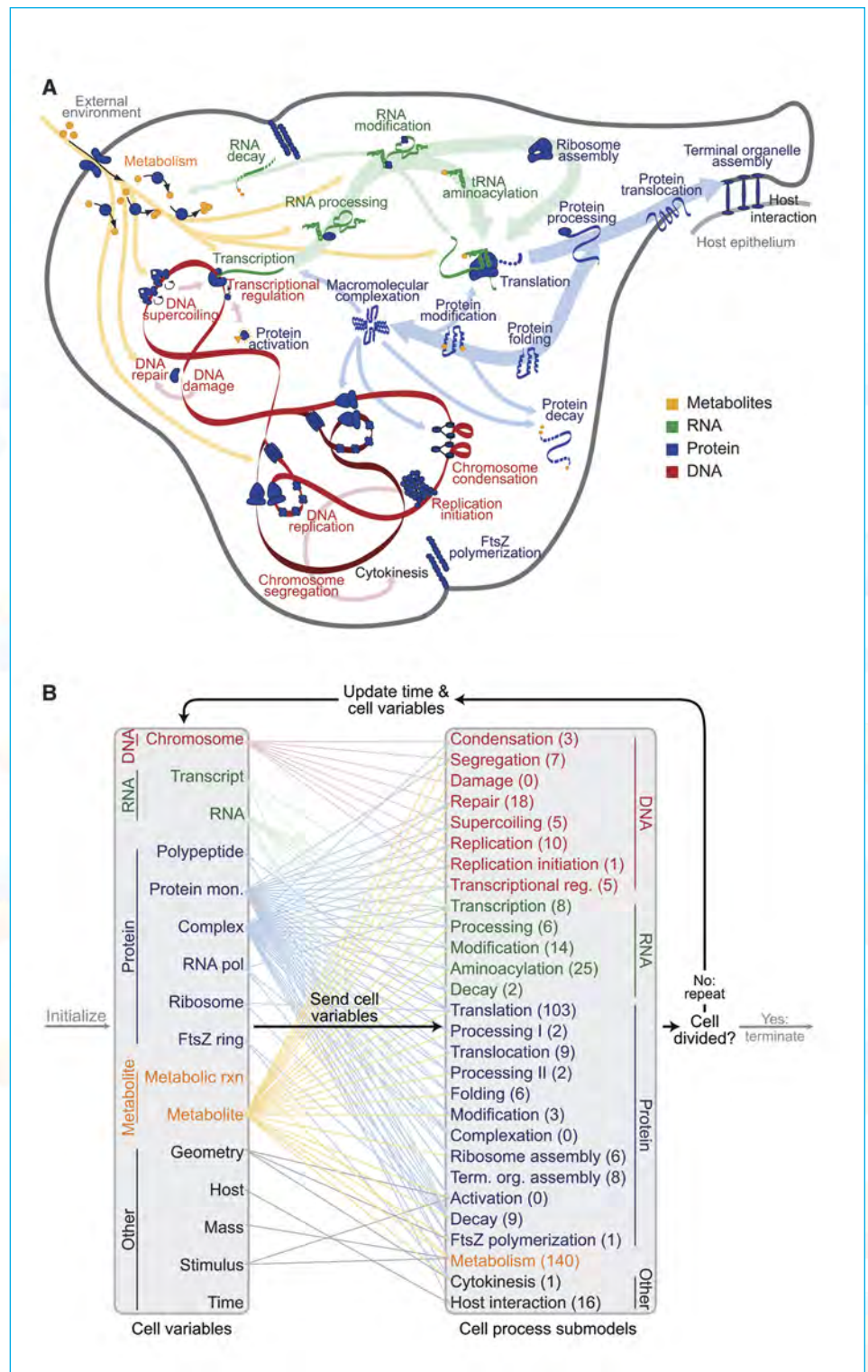
Whole-Cell Model Integrates 28 Submodels of Diverse Cellular Processes. (A) The 28 submodels in Covert's whole-cell model are represented by colored words in the context of the flask-like shape of an M. genitalium *cell. Submodels are connected through common metabolites, RNA, protein, and chromosome, which are depicted as orange, green, blue, and red arrows, respectively. (B) The model integrates cellular function submodels (right-hand column) through 16 cell variables (left-hand column). For each one-second timestep (dark black arrows), the submodels retrieve the current values of the cellular variables, perform their computations, and update the values of the cellular variables accordingly. This process is repeated thousands of times during the course of each simulation, ending only when the cell divides. (Colored lines between the variables and submodels indicate the cell variables predicted by each submodel, while the number of genes associated with each submodel is indicated in parentheses.) Reprinted with permission from Karr JR, et al., A Whole-Cell Computational Model Predicts Phenotype from Genotype, Cell 150:2:389-401 (2012).*

vidual submodels representing different biological functions into a single, integrated über-model of *M. genitalium* might appear to be a Herculean task—especially when many of those functions operate at different time scales, and are computed using mathematical approaches ranging from Boolean logic to stochastic methods. Yet while Covert and his colleagues did indeed describe integration as a "key challenge" in the *Cell* paper announcing their results, it wasn't the only one. And in the end, it was amenable to a reasonably simple solution. At least, for the most part.

"We decided that we could assume that at a short timescale, [the submodels] were independent," Covert says, adding that in this case, "short" meant less than a second. There were exceptions to this rule, most notably in the case of energy, which was in such high demand amongst all the submodels that Covert and his team had to develop a special means of allocating it before anything else could be set in motion. Once that had been worked out, however, Covert, *et al.*, could simulate the whole-cell model by proceeding in one-second timesteps, using the same method employed to integrate ordinary differential equations. For each timestep, they collected the latest data computed for every variable and fed it into the 28 different submodels. Each submodel would then return fresh data, which served as the inputs for

> ### "Integration," Covert says, "happens at the level of data."

the next time step. "Integration," Covert says, "happens at the level of data."

So decoupled, each submodel could even be run serially; though in practice, multiple whole-cell simulations run concurrently on a 128-node computer cluster. Thus far, the team has plowed through thousands of simulations, including hundreds of wild-type cells and hundreds more in which *M. genitalium*'s 525 genes have been disrupted one by one.

Debugging an integrated simulation of this kind can be hairy, and Covert gives credit to a former Google engineer who helped the team develop automated testing procedures for their tens of thousands of lines of MATLAB code. Still, echoing Gerstein, Covert says that a good deal of the toughest

work took place long before anything was integrated. And much of that work involved selecting the most appropriate mathematics for each of the 28 cellular functions, a task that took many years to complete.

Those choices were driven by how well understood each function was, and how much quantitative data was available for it. The most detailed submodels, like the ones for RNA and protein degradation, use stochastic processes to allow for variability. The sparsest rely on Boolean operations. Others still employ flux balance analysis, which analyzes the flow of metabolites through a metabolic network without specifying their actual concentrations. "We really tried to let the process itself, and our understanding of it—together with the data that had been generated with regard to it—be our guide,"

Covert says. All of the code is available online, and Covert looks forward to the day when someone writes a competing submodel and then runs the whole-cell model with both versions to see which works best.

Integrating a Multiscale Genome-scale Metabolic Network

Flux balance analysis lies at the heart of many cellular models and plays an important role in multiscale modeling efforts as well. Typically used to investigate metabolism, the method begins with the reconstruction of a genome-scale metabolic network that describes all of the metabolic reactions that are likely to occur in a given cell based on its DNA, and can then be used to model its various metabolic pathways. Palsson, who helped pioneer the approach, refers to such networks as "supply chain models"—albeit ones that map the relationships between all of the metabolites and enzymes that carry out the biochemical reactions necessary to sustain life.

Recently, Palsson and his colleagues combined a metabolic model for the bacterium *Thermotoga maritima* with a model of macromolecular expression that describes the synthesis of every single one of the organism's proteins. (They created the same kind of integrated model for *E. coli*, as well.) The expression model, which is based on a

network that represents the biochemical reactions that drive transcription and translation, simulates the machinery that a cell uses to build its gene products, and therefore accounts for many things that a standard metabolic model ignores. By integrating the two different kinds of models, Palsson and his team vastly expanded the range of cellular phenomena they could compute and predict. "You just wouldn't believe what we are calculating with this model now," he says before going on to list regulons (collections of genes all governed by the same regulators); metabolic engineering designs; and a variety of cellular functions. In the future, Palsson would like to add genetic regulation to the modeling mix, using the kind of data Gerstein has been exploring with the ENCODE project.

> ## We really tried to let the process itself, and our understanding of it—together with the data that had been generated with regard to it—be our guide," Covert says.

Plenty of challenges remain. Palsson points out that modeling the kinetics and thermodynamics of the many biochemical reactions that take place within a cell is computationally difficult, and will require algorithmic advances. "We understand a lot of individual events," he says, "but putting them all together in a coherent whole is tough."

Integrating Multiscale Models of Tissues

It's equally difficult to simulate the behavior of a population of cells distributed in three-dimensional space, such as one might find in a bacterial infection or a major organ.

That was precisely the problem tackled in a study recently published in *PLoS Computational Biology*.

Ron Weiss, PhD, and his colleagues at the Massachusetts Institute of Technology developed a novel combination of computational methods to design and analyze an artificial tissue homeostasis system—one that uses a synthetic gene network to cause stem cells to grow and differentiate into a stable population of insulin-producing beta-cells of the sort found in the pancreas. (Such a system could be used to help treat Type I diabetes, in which beta-cells are destroyed as the result of autoimmune defects.)

The network is comprised of several dis-

crete modules assembled from standard genetic circuitry components: toggle switches and oscillators to control population growth; sender-receiver systems to permit intercellular communication. The question, says Weiss, who is a dual associate professor of biological engineering and of electrical engineering and computer science, was whether he and his colleagues would be able to predict the behavior of the entire system once all the modules were connected. "What happens when you take these known modules and try to integrate them into a much more complex system?" he asks.

Weiss and his team designed several different iterations of their system, each one more sophisticated than the last. And they simulated those systems using three different mathematical models that progressively accommodated more and more complexity: one that used ordinary differential equation simulations, and two that used stochastic differential equation simulations to allow for noise and spatial effects. The spatial effects are important because cells that are distributed across space are exposed to different environmental conditions and can't communicate instantaneously with each other. The noise effects matter because two cells that contain the same genetic circuits can still produce different amounts of a particular protein due to unpredictable fluctuations in gene expression. To his surprise, Weiss found that having some noise in the system was actually helpful. "Normally, in synthetic biology, you think of noise and heterogeneity as being bad things—things that tend to destabilize the system, things that you want to get rid of," he says. "In our system, the addition of noise actually stabilizes the system and makes it more robust."

Once again, the low-level integration of the modules—*i.e.*, the act of joining them together to form a larger system—was not the most challenging aspect of the project; for the most part, Weiss explains, it involved defining the interfaces between the various modules and then "gluing one module onto another." The tricky part was figuring out which bits mattered most to overall system performance—especially since the modules affected one another in unexpected and often non-linear ways.

Weiss and his colleagues first examined how certain module behaviors combined to produce optimal system performance, and then applied Bayesian network inference, which graphically represents the probability that different variables may be related to one another, in order to identify the individual behaviors that had the greatest impact. Weiss feels that such methods represent broadly applicable techniques for aiding integration, just as the team's decision to proceed from simpler systems and models to more detailed and accurate ones offers a general approach towards system design and understanding that could be useful to others.

Integration from Cell to Organism

If Weiss proved that integrated modeling and analysis could bridge the gap be-
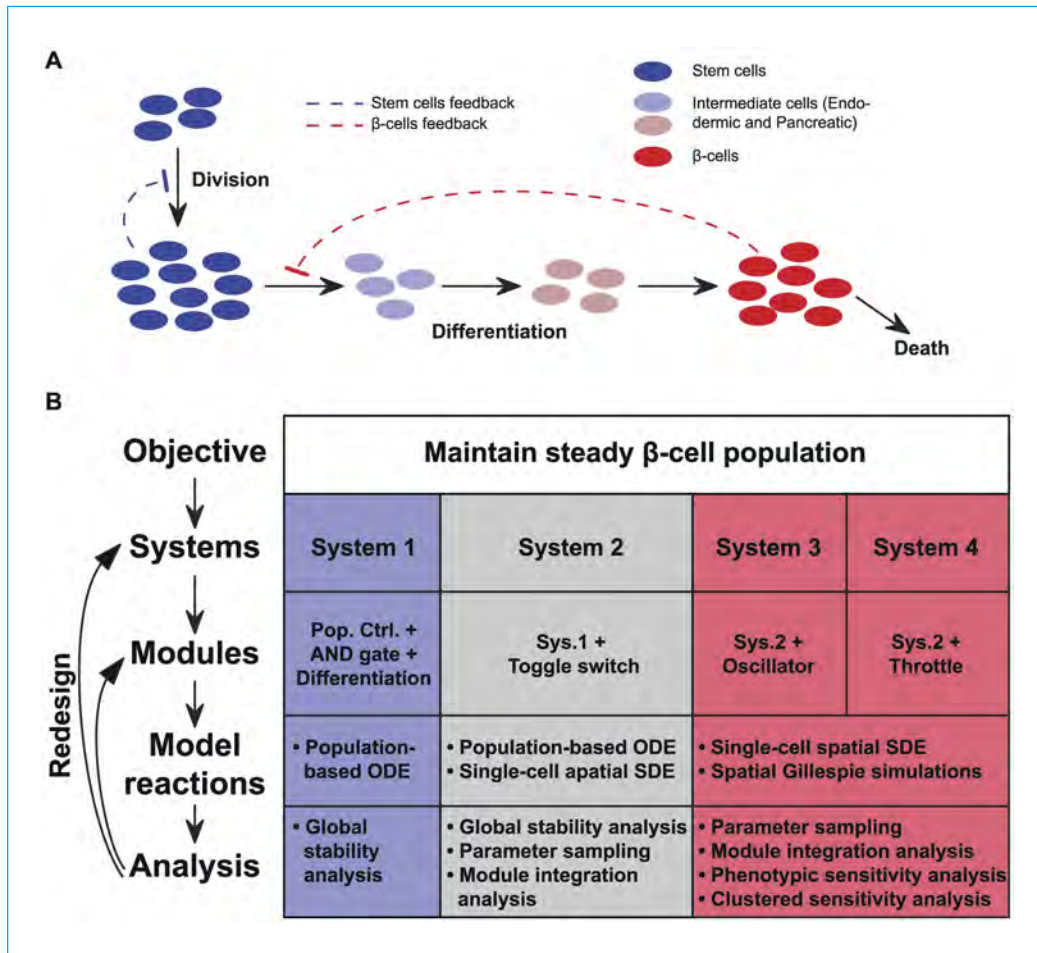
Tissue Homeostasis Model Design. (A) Stem cells (blue) divide and then differentiate through several phases to become beta (ß)-cells (red), a process that is governed by negative feedback from the ß-cells. (B) Weiss and his colleagues created an integrative model of tissue homeostasis by designing a modular system and subjecting it to repeated analysis and redesign. The system went through four iterations, each one more sophisticated than the last. Reprinted from Miller M, et al., (2012) Modular Design of Artificial Tissue Homeostasis: Robust Control through Synthetic Cellular Heterogeneity. PLoS Comput Biol *8(7): e1002579. doi:10.1371/journal.pcbi.1002579.*

"Normally, in synthetic biology, you think of noise and heterogeneity as being bad things— things that tend to destabilize the system, things that you want to get rid of," Weiss says. "In our system, the addition of noise actually stabilizes the system and makes it more robust."

tween genetic circuitry and heterogeneous populations of cells, **Lars Kuepfer, PhD**, showed they could do the same for cells and entire organisms.

Working as part of the Virtual Liver Network, a national initiative funded by the German Federal Ministry for Education and Research, Kuepfer and his colleagues in the computational systems biology group at Bayer Technology Services in Leverkusen,

whole-body level, thereby revealing how cellular and extracellular mechanisms influence one another. In a series of case studies, they examined the cellular basis of acetaminophen poisoning; probed the workings of allopurinol, a drug used to treat gout; and looked at how variations in individual physiology (such as liver size) can interact with metabolic disorders (such as an impaired ability to eliminate ammonia)

his next steps, and would like to integrate patient-specific metabolome data into the whole-body model. The key challenge, he says, is not the integration *per se*, which involves using the data generated by the PBPK model to constrain the metabolic one, or pumping the output from the metabolic model into the whole-body simulation. Rather, it is in knowing enough to do both kinds of modeling in the first place.

Kuepfer, who did his graduate work in metabolic modeling and now works for a company that uses pharmacokinetic and pharmacodynamic modeling to evaluate drug candidates, has the tools and experience to work both sides of the street. As things stand today, however, most specialists in genome-scale metabolic network reconstruction probably wouldn't share his familiarity with PBPK modeling—though Kuepfer does expect a growing number to extend their focus beyond the cellular scale in the future. Palsson, meanwhile, points out that metabolic models, though highly scalable and easy to compute, can also be hard to understand. "How to use them and apply them requires a certain skill set that isn't commonly available," he says.

Integrating Metabolic Models with Gene Expression. (a) illustrates how genome-scale metabolic models are constructed, with specific metabolic paths inferred from the presence of particular genes. (b) illustrates how integrated models of metabolism and macromolecular expression (aka ME-Models) like the one in Palsson's paper link various biological sciences and relate gene products to genetic perturbations and gene functions in the context of cellular physiology. Reprinted with permission from Lerman JA, et al., In silico *method for modelling metabolism and gene product expression at genome scale,* Nat Commun *doi: 10.1038/ncomms1928 (2012).*

Germany, yoked a genome-scale metabolic model of the kind used by Covert and Palsson to a physiologically-based pharmacokinetic (PBPK) model of the sort used to simulate the availability of drugs in tissues throughout the body. More precisely, they integrated a genome-scale network reconstruction of a human hepatocyte into the liver tissue of a PBPK model representing an adult human being. The resulting multiscale model enables the calculation of thousands of cellular reactions within a whole-body framework containing 200 or so ordinary differential equations and several hundred parameters, ranging from anthropometric details like age and height, to physicochemical ones like the solubility and molecular weight of the compounds under investigation.

Kuepfer and his team were then able to introduce changes at the whole-body level—administering a therapeutic agent or a drug overdose, for example, or generating an abnormally high level of some naturally occurring compound—and track the effects at the cellular level, and even feed the ensuing metabolic perturbations back to the

to create otherwise inexplicable levels of biomarkers in the blood.

The Integrative Skill Set

Palsson, who has used metabolic network reconstructions to model the interactions between multiple tissue types, says that multiscale, cell-to-whole-body models like Kuepfer's are going to make "astonishing progress over the next decade or so," in

part because so many diseases, from cancer to psychiatric disorders, have a metabolic component. Kuepfer is already considering

And that might be the common take-away from all of these studies. Figuring out how to integrate the various models and data types involved is one thing; but it is not the only thing, nor is it necessarily the hardest thing. Often, the thorniest issues involve overall design, or conceptual clarity, or individual expertise. Integrative modeling and analysis may hold the keys to many complex computational and biological problems. But they will only lead to meaningful results if researchers give

careful thought to the individual components they are attempting to combine—and the problems they are trying to solve. □

…metabolic models, though highly scalable and easy to compute, can also be hard to understand. "How to use them and apply them requires a certain skill set that isn't commonly available," Palsson says.

Vitreous body
(filled with vitreous humor)

Medial rectus muscle

Retina

Sclera

Retinal pigment epithelium

Ciliary muscles

Choroid

Lens

Pupil

Optic nerve head

Iris

Optic nerve

Central retinal artery and vein

Cornea

Anterior chamber
(filled with aqueous humor)

Lamina cribrosa

Courtesy of:
National Eye Institute

Lateral rectus muscle

The EYES Have It:

By Katharine Miller

Biomechanical Models Explore Disorders of the Eye

Squint,
and you can almost make out that bird

soaring over the horizon. But determining whether it's a hawk or a raven will be nearly impossible for someone with myopia, also known as nearsightedness. In this common condition, light focuses on a spot in front of, rather than on, the retina. Eyeglasses can correct the defect, as can refractive surgery in which a lens-shaped portion of the cornea—the outer layer of the eye in front of the pupil—is removed in a precise way, pushing the focus back to the retina.

But when **Anna Pandolfi**, **PhD**, asked her doctor if her myopia could be treated with refractive surgery (commonly known by such trademark names as LASIK or LASEK), he said her eyes were "too thin" to endure the surgery.

Pandolfi, associate professor of structural mechanics in the structural engineering department at the Politecnico di Milano in Italy, wanted to know more: How did her eye abnormality affect the surgical outcome? After some initial research, she realized that researchers didn't really know the answer: They couldn't conduct experiments on humans without great risk to the patient; and there were no adequate computer models of the eye.

Pandolfi also realized that she could help. As a civil engineer with a longstanding interest in computational mechanics and the study of materials, she could create a computational model of the

cornea that might help explain how refractive surgery would impact problematic eyes such as hers.

Pandolfi is not alone in realizing the need for computational modeling of the eye's biomechanics. The risks of experimentation are too great. "You can't start hacking around with peoples' eyes to see what works," says **Harvey Burd**, **D. Phil.**, university lecturer in engineering science at the University of Oxford.

Moreover, given the eye's multiple interconnected parts, each of which is comprised of layers of cells, connective tissue, and fluid that must all function properly to give us sight (our most valued sense), it's perhaps no surprise that there are myriad biomechanical ways the eye can fail. The array of diseases and disorders that afflict the eye include glaucoma, cataracts, macular degeneration, and retinal detachment, to name just a few.

Despite the eye's complexity, says **Phil Luthert**, **PhD**, professor of pathology at University College, London, it lends itself well to computational

> "There is something fantastically tractable about the eye," Luthert says. "[It] ought to be a really neat place to try to get an integrated model in the next ten years."

approaches. "There is something fantastically tractable about the eye," he says. The discrete steerable globe has a blood supply; allows light in one end; and sends signals through the optic nerve at the other end. "The eye ought to be a really

neat place to try to get an integrated model in the next ten years."

Already, biomechanical models of practically every part of the eye—from the muscles that control eye movement to the cornea, lens, vitreous humour, sclera, lamina cribrosa and retina—are contributing to a better understanding of both the normal and the diseased eye.

What's in the Blink of an Eye?
Biomechanics of eye muscles

Eyes can move extremely fast. Indeed, the jerky eye movements called saccades are the fastest movements produced by the human body. Eyes also engage in smooth pursuit as they follow a moving object. And they converge and diverge as well—moving toward or away from each other to maintain binocular vision. The muscles and tissue that surround the eye control all of these movements.

Biomechanical models of the eye muscles can give insights into diseases in which the muscles fail. For example, a condition called strabismus can result when the eyes are not properly aligned; and gaze palsy prevents the eyes from moving in the same direction. Surgeries for these problems often go wrong because they involve guesswork about which muscles to shorten or lengthen.

The model SEE++ can help guide surgeons to improve outcomes. It consists of three parts: a geometric model of the muscles; a muscle force model; and a kinematic model that brings the geometry and forces together to define stable eye positions.

Using SEE++, physicians can enter a patient's response to diagnostic testing (such as the classic "head tilt test" in which tilting the head results in telltale eye movements) and try to work backward to understand the muscle forces that cause the patient's particular pathology. "It's not always perfect," says **Thomas Kaltofen**, a researcher at RISC Software GmbH, a limited liability company in Austria that developed the program. "There's often more than one way to achieve a particular pattern." But in conjunction with a physician's expertise, the program can hone in on the causes of a particular pathology and then simulate how to correct it surgically. SEE++ is currently used for

surgical planning, mostly in Austria.

"Often, we're ruling things out as well as ruling them in. You get closer and closer to the pattern measured from the real patient and at some point assume that you might be simulating the actual case," he says. "It works quite well when you have enough data."

Kaltofen and his colleagues recently added the skull to the visualization, and hope to soon integrate it into the model as a constraint on muscle movement. They also plan to integrate a more flexible model of the signal that comes from the brain to the muscle; and add the capacity to simulate new, innovative surgeries.
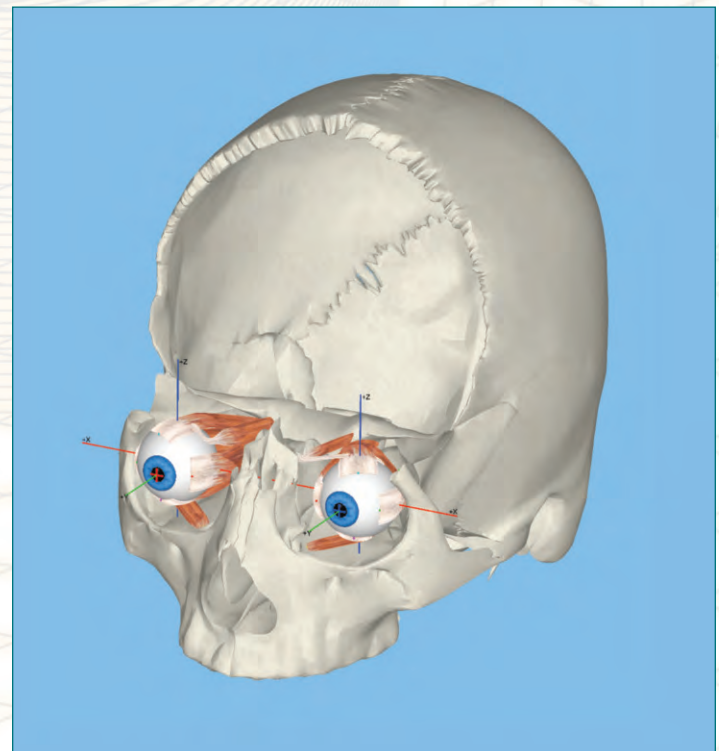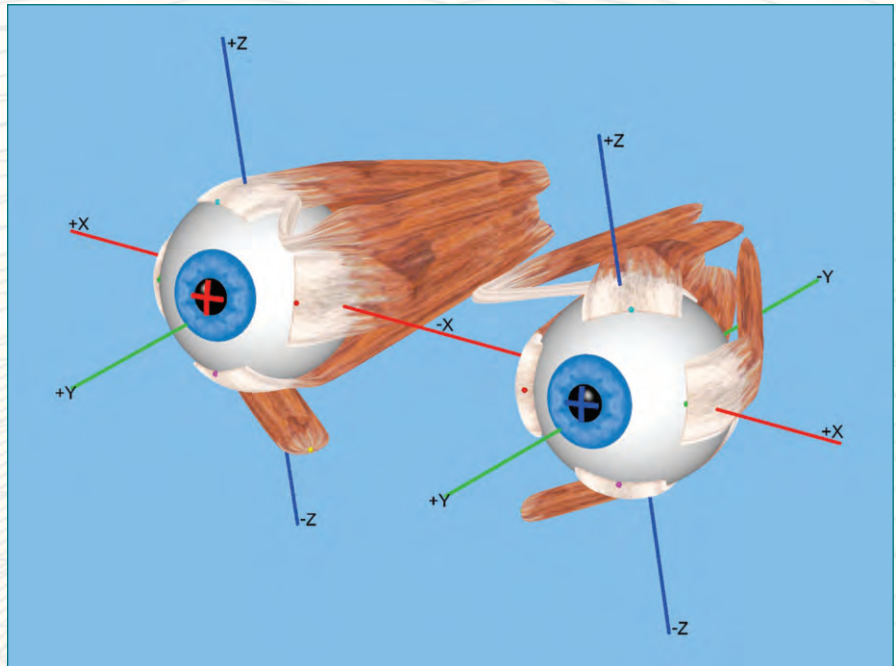
A Model of Extraocular Muscles. The software program SEE++ creates individualized models of the eye muscles that can be used to both diagnose muscle problems and plan surgical treatments. Courtesy of: Thomas Kaltofen, www.see-kid.at.

Despite its usefulness, SEE++ simulates only static movements (where the eye ends up) rather than dynamic movements (the trajectory and accelerations it used to get there). A new model created by **Dinesh K. Pai**, **PhD**, professor of computer science at the University of British Columbia, takes this extra step. "Static movements assume you can ignore the inertial term—the effect of mass or accelerations—and that motions are balanced at all times," he says. "But eye movements are some of the fastest movements we make; and the dynamics are significant."

> "[E]ye movements are some of the fastest movements we make; and the dynamics are significant," Pai says.

result, rates of success at eye surgery are relatively poor," Pai says. "If we can model these, we can use the model for surgical planning."

SEE++ has also added a model of the pulleys, Kaltofen says. Pulleys detour rather than going straight; and they move when the eye moves. So SEE++ includes pulleys in the geometric model as well as in the force and kinematic models. "The whole model changes by introducing this behavior, though it's a slight change," Kaltofen says. Why the pulleys move with the eye is not fully clear. They may help simplify the brain's job of coordinating eye movements, Kaltofen says, "but that's only one theory." Research is ongoing.

The Window into the Eye:
Models explain cornea's transparency

The cornea is a curved lens that bends light toward a focal point inside the eye, providing about two-thirds of the eye's optical power. Structurally,
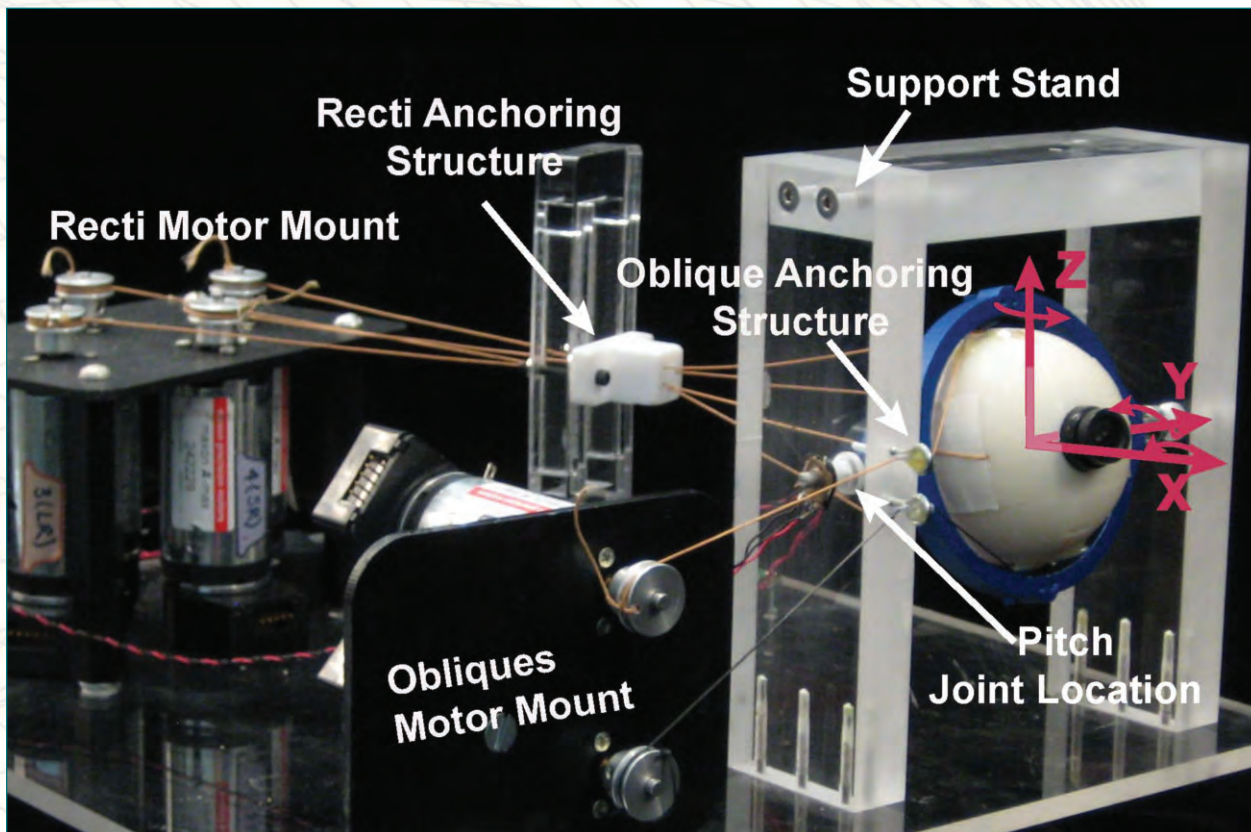
A Robotic Eye. Pai's computational model of extra-ocular eye movements was the basis for this tendon-driven robotic eye, which can even engage in the extremely rapid eye movements known as saccades. Courtesy of: Dinesh Pai, Mahkameh Lakzadeh and Tony Hodgson.

Using their model, Pai and his colleagues have learned how to build robotic eyes that can do both saccades and smooth pursuit. Their model includes "pulleys," the connective tissues that lie between the muscles on the sides and top of the eye. These tissues play a greater role than once believed. "Many surgeries cut away at these functionally important connective tissues. As a

"the cornea is a perfect pressure vessel," says **Peter Pinsky**, **PhD**, professor of mechanical engineering at Stanford University. "It's floppy and flexible when isolated, but assumes a highly precise and stable shape when internally pressurized like a balloon." It is also—rather remarkably—transparent, allowing about 90 percent of incident light to enter the eye.

Computational biomechanics can help provide a theoretical understanding of both corneal shape and transparency, issues that matter for a variety of clinical problems, Pinsky says. For example, corneal shape is an important issue for refractive

surgery—the procedure Pandolfi was denied due to the thinness of her cornea. Also, in certain diseases such as keratoconus, corneal shape is severely changed as a result of ultrastructural rearrangements within the tissue. And other disease processes can compromise corneal transparency, leading to blindness.

Pinsky spent most of his career working on the theory side of finite element method (FEM)—a numerical technique that divides a complex object into simple elements that can be solved in relation to each other, often in a triangular mesh. But about 10 years ago, he began applying FEM to model the cornea. He's specifically interested in how the cornea maintains transparency as well as why it assumes the shape it does.

Prevailing theory holds that a cornea's transparency stems from its nanoscale organization. Scientists know that within the stroma—the thickest layer of the cornea—lies an intricate system of ribbon-like fibers called lamellae. "These flattened strands lie one on top of another in up to 500 layers," Pinsky says. The lamellae are also like large cables—cut through them and you will see the finer level of smaller cables made up of thousands of individual fibrils that are beautifully organized in an almost perfect regular lattice. But researchers don't really know precisely how the fibrils are maintained in this pseudo-lattice arrangement as required for transparency.

Proteoglycan molecules—proteins with "GAG" (glycosaminoglycan) chains that branch out in a star-like structure—are known to play a critical role. Theory has it that these GAGs bridge from one collagen fibril to another to form an elastic network. But this doesn't match up with imaging or fit what's known about GAG chemistry, which suggests the fibrils and GAGs should be mutually repulsive, Pinsky says.

To study the phenomenon theoretically, Pinsky and his colleagues created a numerical model of a portion of the cornea. Working with a representative volume, they characterized the system in an idealized way—endowing the fibrils and proteoglycans with nicely organized properties and characterizing the system's electrostatic and elastic energies. "We then try to find the model parameters that make the energy of the system stationary," he says.

The result: The fibril lattice's response to external forces is better explained by osmotic stress perturbations resulting from electrostatic interactions than by the GAG elastic bridging theory. And the model explains a curious quality of corneas: when isolated and placed in a salt bath, they swell enormously. No other theory successfully predicts this behavior of the cornea at various levels of hydration. The model is being applied to

explain the fundamental mechanism for Fuch's dystrophy, a condition in which the ion pumping mechanisms at the posterior surface of the cornea are compromised, resulting in swelling and a loss of corneal transparency.

The second thrust of Pinsky's cornea work involves understanding how refractive surgery impacts the shape and health of the cornea. When removing part of the cornea or adding implants, Pinsky says, it is important to know how the cuts affect the tissue mechanics. Imaging is starting to provide a good 3-D understanding of how the lamellae are organized in the stromal layer of the cornea. And Pinsky's lab has been doing experiments to mechanically test super-thin slices of the cornea. His group has also put all that data together to produce a predictive FEM model and a theory of the tissue's 3-D structure.

"This is of interest to the laser companies," Pinsky says. "LASIK, which is really an ingenious procedure, is nevertheless a pretty significant attack on the cornea." It involves cutting a flap, lifting it, and then vaporizing some of the stromal tissue in a precise way. The flap is then carefully repositioned over the cornea. But, says Pinsky, "the flap remains mechanically defunct because wound healing does not fully integrate it with the underlying stroma." And the removal of tissue changes the state of stress in the remain-
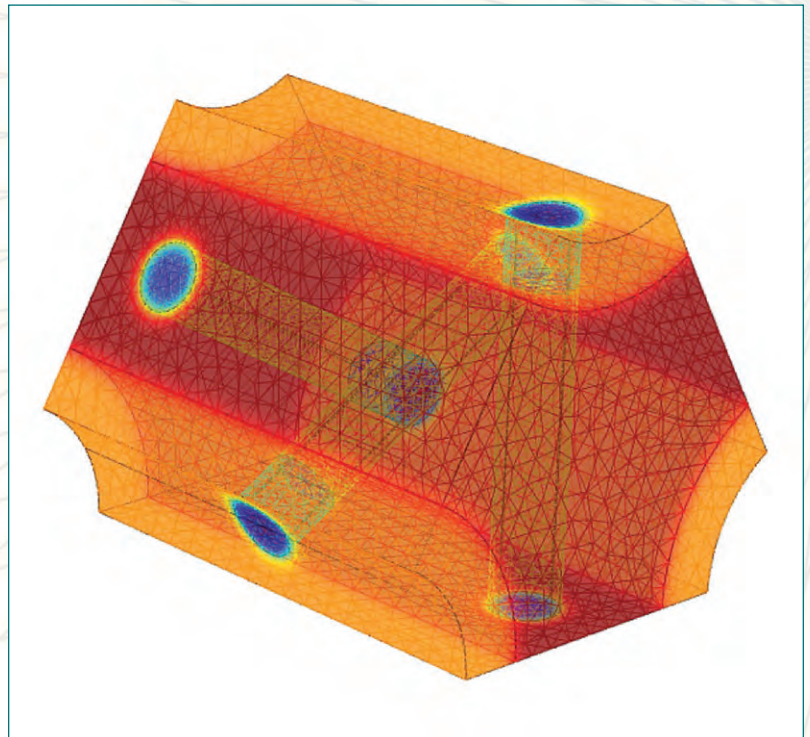
Corneal Tissue "Unit Cell." *To model the nanoscale structure of the cornea, Pinsky and his colleagues worked with a unit cell (representative volume) of tissue, as illustrated here, showing collagen fibrils (cylindrical zones with blue ends) and GAGs arranged in a next-nearest neighbor connectivity (in the red zone—not explicitly modeled). The team used numerical methods to analyze the unit cell and obtain the free energies of the system. The color contours illustrate the electrostatic potential. Courtesy of: Peter Pinsky.*

ing tissue. Using his models, Pinsky seeks to understand how the cornea responds to various surgical procedures with the goal of improving the predictability of outcomes. He also hopes the models will provide insight into how the lamellae rearrange to produce an altered corneal shape in keratoconus.

Pandolfi is also interested in using FEM to understand how refractive surgery affects the eye, with a particular interest in problematic eyes like hers. Unlike Pinsky, she models the corneal material as a whole, rather than at a molecular level. "Once you have a numerical model, you can run a simulation to see if the eye can undergo a surgical intervention or not," she says.

To create her model, Pandolfi started with the geometry—building the exact shape of the cornea using photographs and measurements. "This was easy in that you have all the information you need," Pandolfi says. Describing how the material behaves is trickier, she says, because the mechanical properties cannot be directly measured in living tissue. She only had data on how dead pig or human corneas behave in response to displacement. So, she used inverse analysis—feeding the model with the geometry of these tests and observed displacements—to discover the mechanical parameters.

Using this approach, Pandolfi showed that, after refractive surgery, the eye's refractive power (visual acuity) is more sensitive to changes of intraocular pressure (IOP). "For an engineer, this is clear," Pandolfi says. "The cornea is thinner so it is more sensitive to a change of IOP." Pandolfi has now validated the results. In addition, she showed that after removing 10 percent of the cornea in a simulated refractive surgery, the remaining tissue is subjected to 15 to 20 percent more stress. It's a noteworthy result given that some surgeries reduce corneal thickness by as much as 50 percent. Imagine thinning the cornea from the thickness of a soccer ball to that of a helium balloon: it becomes more stretchable in response to internal pressure, and this in turn affects the cornea's ability to focus light in the right spot.

To date, Pandolfi's model has relied on average values for the cornea's geometry and material properties. Now she's moving toward understanding the differences between different individual's eyes. "We cannot speak about an ideal eye," she says. "We have to speak about the range of variability of the parameters that define the real eye."

> "Once you have a numerical model, you can run a simulation to see if the eye can undergo a surgical intervention or not," Pandolfi says.

One of her students is using geometric data from 20 pre- and post-operative eyes and simulating how the surgeries likely affect IOP and the stress state inside the cornea.
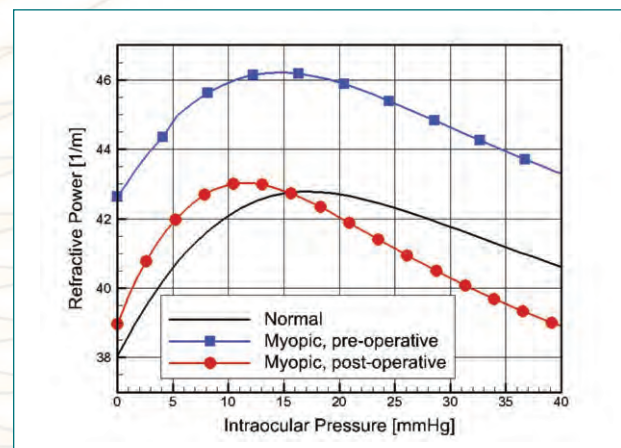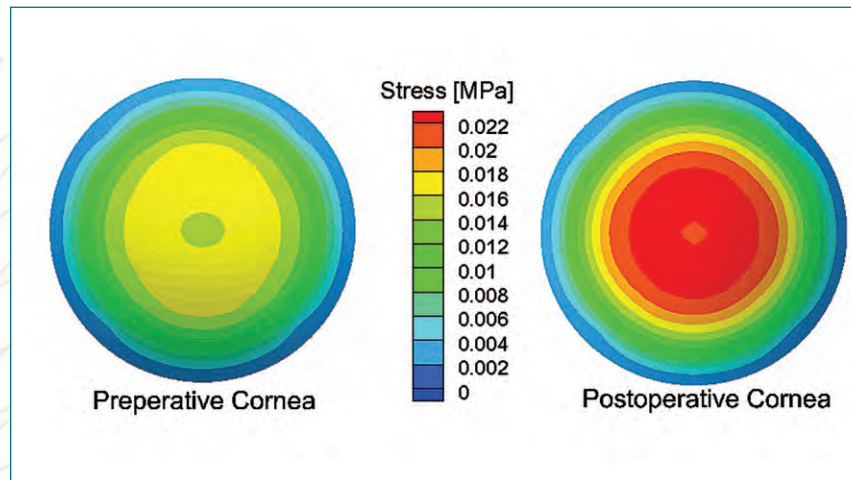
A Weakened Cornea. Pandolfi's finite element models of the cornea after refractive surgery revealed that the thinner cornea is under greater stress (top) and loses visual acuity more rapidly than normal corneas do if IOP goes up (graph, lower).

Clear-eyed:
Models of the lens explore cataract prevention and treatment

Behind the cornea lies the aqueous humor; and behind that floats the lens, a transparent unit that is responsible for about one-third of the eye's optical power. Shaped like a lentil bean—round, and convex on both sides—the lens has an outer capsule made of stiff collagen elements in a flexible matrix. Inside, lens fibers stretch from the front to the back in onion-like layers.

With age, the lens can become opaque, a condition called age-related nuclear cataract that is treated by removing the natural lens and replacing it with an artificial one. Also with age, the lens loses its ability to adjust (or "accommodate") its focus from objects in the distance to objects close by. Called presbyopia, this condition occurs in almost all adults after about age 50.

Using computer modeling, researchers are trying to get a better understanding of how the lens develops cataracts, how it responds to cataract surgery, and how it loses its ability to accommodate. "It's a neat system for an integrative model because at the cellular level, the location of physiological components determines the local tissue's optical function," says **Paul Donaldson, PhD**, professor of optometry and vision science at the University of Auckland, New Zealand.

Nuclear cataracts develop as a result of protein cross-linking at the center of the lens. "It's like molecules falling out of solution and scattering light," Donaldson says. The cause is uncertain, but Donaldson believes that protein cross-linking might be driven by a failure of the lens to maintain its normal physiological environment—a process in which internal micro-circulation likely plays a key role. As we age, Donaldson proposes, the micro-circulation system runs down, losing its ability to deliver sufficient nutrients to the lens center and initiating the biochemical changes that lead to protein cross-linking and compromised lens transparency.

The existence of a lens micro-circulation system was first proposed by Donaldson's collaborator, **Richard Mathias, PhD**, who is now professor of physiology and biophysics at the State University of New York, Stony Brook. He took electrical measurements, analyzed the circuits discovered, and modeled the resulting ion and water fluxes to demonstrate that micro-circulation occurs within the lens. Over time, Mathias' model evolved to cover many physiological components of the lens, says **Ehsan Vaghefi, PhD**, a bioengineering researcher in the department of optometry and vision science, also at Auckland. But Mathias solved his model analytically and in one dimension, which tended to become exponentially time-consuming and complex. Building on this work, Donaldson and Vaghefi encapsulated Mathias's analytical model into a 3-D finite element framework that includes parameters such as the spatial location of trans-

porters and ion channels/pumps throughout the lens. They then used brute force (iterating through a series of approximate solutions) to solve numerical equations that describe fluid circulation at a microscopic level throughout the lens. After many iterations, the results converged on a solution.

The advantage of this approach, Vaghefi says, is that the model can make predictions about what happens when the environment and/or physiology of the lens is perturbed. For example, the team has now shown that, in response to stimulation with an external probe, the model lens behaves the same way as a real lens.

Vaghefi and Donaldson both hope their model will be used as a tool to predict how age-dependent changes in lens physiology affect the progression of lens cataract—and ultimately improve treatment. "If we could up-regulate the circulation system of the lens," Donaldson says, "we could deliver antioxidants to delay onset of cataract." Delaying cataracts by just 5 to 10 years would actually cut their incidence in half, he notes, since many people will die of natural causes before they even get them.

Computer models of the lens could also help researchers design better cataract surgical procedures and lens replacements. Currently, cloudy cataract lenses are surgically replaced with new lenses that have a fixed optical power. "It's a plastic lens that behaves like a spectacle lens but inside your eye," Burd says. "A lot of people are trying to head toward a lens that can behave more like a natural lens." But that, of course, requires a good understanding of the natural lens physiology.

To that end, Burd created a finite-element, multiscale model of the lens capsule—the outermost layer of the lens. Like the cornea, the lens capsule has a complex microscopic structure that affects the material's behavior. So Burd's millimeter-scale finite element model of the lens capsule includes microscale structural information about the behavior of collagen fibers embedded in an elastic matrix. The model could be used to evaluate how the implantation of new lenses stresses the lens capsule.

Burd has also modeled how the lens loses the ability to accommodate as we age. He and his colleagues have proposed an empirical model that represents the changes in the elasticity of the lens over time. Combining that with Burd's lens capsule model, they showed that 80 percent of the age-related decline in lens accommodation is caused by increased stiffness in the lens fibers—not changes to the capsule itself. This suggests that inserting an appropriately flexible artificial lens within the cap-

Net Microcirculation in the Lens. *This 3-D vector map shows the net current densities calculated by Vaghefi and Donaldson's model of microcirculation in the human lens when the model is stimulated with a simulated vibrating probe at the pole. According to the model, net flow moves inward at the poles and outward around the equatorial plane. Reprinted from Vaghefi E, et al., Development of a 3-D finite element model of lens microcirculation, BioMedical Engineering OnLine 2012, 11:69 doi:10.1186/1475-925X-11-69.*
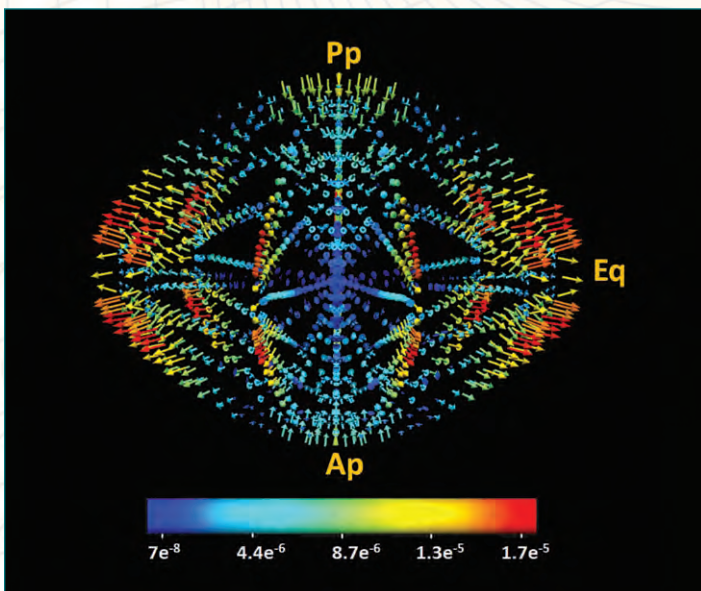
sule might help the lens retain its ability to accommodate. But, Burd says, researchers would have to know how such a surgery would alter the material properties of the lens capsule—another topic he hopes to address using his models.

Not a Dry Eye in the House:
### Fluid mechanics in the vitreous humour

Moving deeper into the eye, the next major structure is the vitreous humor—a gelatinous blob that fills the area between the lens and the retina. As the eye moves, so too does the blob, exerting mechanical forces on the surrounding tissue—primarily the retina. With age, the blob can become liquefied in parts and can also shrink and detach from the retina, allowing the liquefied portion to fill in the gap between the vitreous gel and the retina. This causes flashes of light and floaters in the visual field as the eye moves. More importantly, however, portions of the vitreous that are still attached to the retina can pull on or tear the retina. This, in some cases, leads to retinal detachment, a condition in which the photoreceptor-rich retina pulls away from the nourishing tissue behind the eye, resulting in blindness.

Interestingly, retinal detachment following vitreous detachment occurs more often in nearsighted eyes. **Rodolfo Repetto, PhD**, lecturer in hydraulics at Universita Degli Studi di Genova, and his colleagues hypothesized that mechanics plays a role. "Possibly myopic eyes' oblong shape affects friction at the interface between the liquefied vitreous humor and the retina during eye movement," Repetto says. His group's computational models of fluid flow in the vitreous during saccades—rapid flickering motions of the eye—confirmed this hypothesis. "Even with a homogeneous fluid, the stresses within the vitreous and on the retina are significantly higher [in myopic eyes]," he says.

Repetto and his colleagues are also modeling the effect of surgical treatments for retinal detachment. Typically, surgeons attach a silicon band that deforms the eye shape into a slight hourglass shape, pulling the retina back in touch with the tissue behind it. "It works, but there was no understanding about what happens in terms

of the motion of the fluid in the vitreous chamber," Repetto says. His group showed that the motion and stresses in the humour are significantly altered. "Reattachment depends on the stresses generated from the inside on the retina," he says. "It's a first step toward understanding why the process works."

Another surgery replaces the vitreous with an oil that pushes the retina back in contact with the essential underlying tissues. But sometimes the oil breaks down into an emulsion that is not
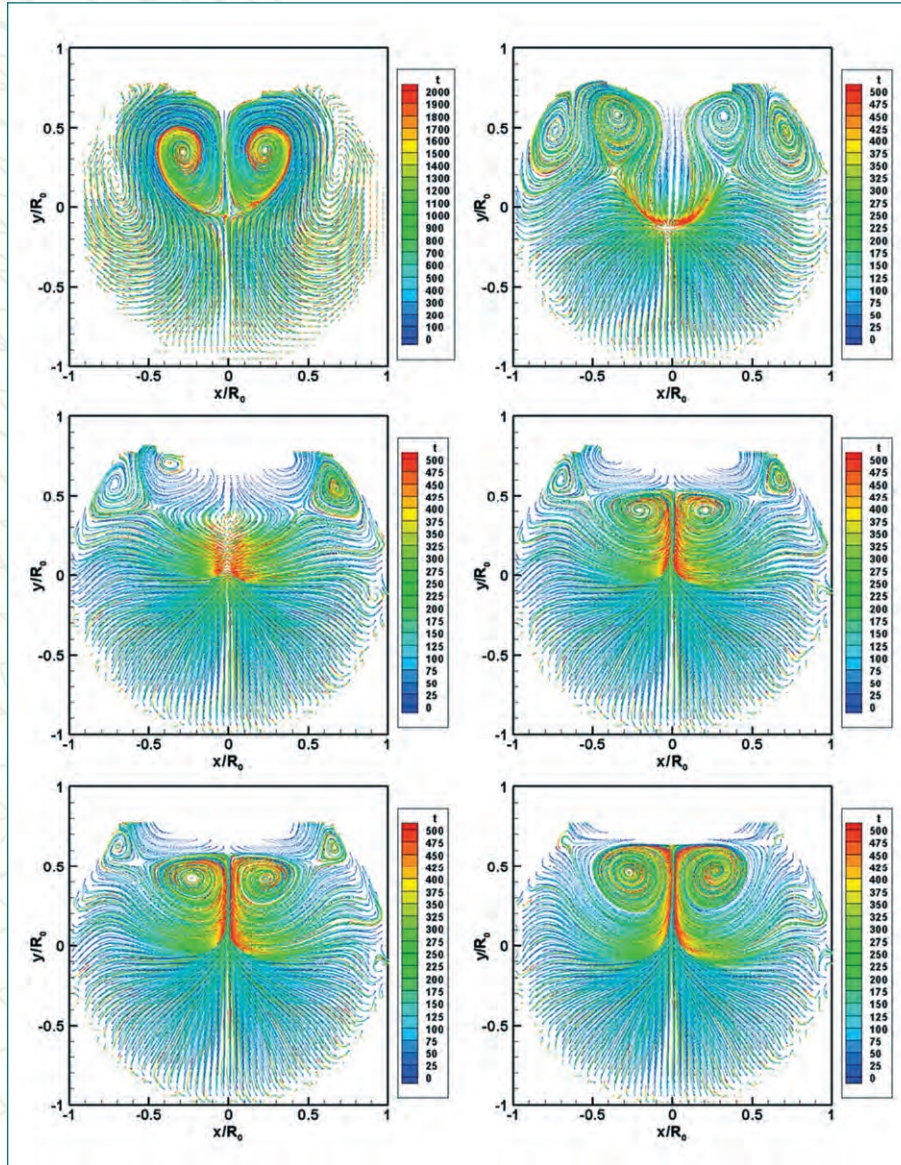
Mixing Processes in the Vitreous. *For some diseases, drugs are delivered to the retina by injection into the vitreous. Models of fluid flow in the vitreous can help researchers understand how much of the drug will actually reach the targeted cells before it is excreted. Here, Repetto and his colleagues modeled particle movement in the vitreous in an analytical model and showed that the mixing properties and fluid structures depend on fluid properties, such as viscosity, and eye motion (particularly, speed). Recently, Repetto's team computationally reconstructed 2-D images of the speed of these flows to get a fully 3-D image of fluid motion in the vitreous humour. Reprinted from Stocchino A, et al., Mixing processes in the vitreous chamber induced by eye rotations,* **Phys Med Biol** *55 (2010) 453–467, with permission from IOP Publishing.*

transparent. Repetto's group is modeling the oils to better understand why this happens. "It depends again on mechanics," he says.

With an Eye to Glaucoma:
Modeling the effects of intraocular pressure

At the back of the eye, more than a million nerve fibers (axons from ganglion nerve cells in the retina) extend through an opening known as the scleral canal within a zone called the optic nerve head (the eye's "blind spot," which has no photoreceptors). The lamina cribrosa (LC), a porous structure that resembles a loose foam, fills in the area around the nerves. In glaucoma, a disease whose most common symptom is elevated intraocular pressure (IOP), "the LC gets squished down, moves backwards, and becomes more like scar tissue," says says **Ross Ethier**, **PhD**, senior research investigator with the department of bioengineering at Imperial College, London. Ultimately, these changes somehow damage and kill off the nerves passing through the scleral canal, causing blindness. But there's little known about why one person might get glaucoma while another does not.

Ethier hypothesized that individual differences in the thickness of the sclera—the white of the eye—might mediate the effects of pressure on the optic nerve head. In humans, the thickness of the sclera is quite variable, so he and his team used 11 post-mortem eyes (7 normal and 4 with glaucoma) to build 11 finite element models of the scleral shell, each containing a model of an identical and idealized optic nerve head. In the models, scleral thickness turned out to be a significant factor in glaucoma risk: Differences in scleral thickness, particularly in the region next to the optic nerve head, produced significant variation

in strains across the LC.

Rafael Grytz, PhD, is also interested in the effect of elevated eye pressure on the optic nerve head. He recently left the Devers Eye Institute Research Labs in Portland, Oregon, to become assistant professor of ophthalmology at the University of Alabama, Birmingham, where he will
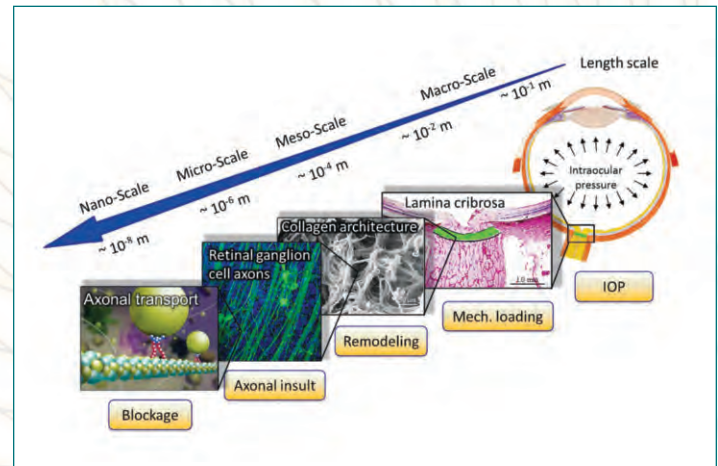
Multiscale Eye Models. *Grytz integrates eye models at a variety of scales. Courtesy of: Rafael Grytz.*

continue working on multiscale models to explore the biomechanics of the eye. "I look at glaucoma and other eye diseases and one thing that strikes me—and it's the theme of my work—is that the biomechanical mechanisms are occurring and interacting across very different length scales," he says.

Pressure creates a loading condition at the macroscale that translates down the scale of the collagen fibrils, the main load-bearing constituents of the LC. "If you load them beyond certain pressures, remodeling occurs," he says. That, in turn, might have an impact on the nanoscale—interrupting axonal transport.

So Grytz's models start with a biomechanical model of how the individual collagen fibril responds to loaded forces. From this, he derives material properties of eye tissues at the microscale, and then simulates growth and remodeling of the entire eye with particular attention to how the LC thickens in the early stages of glaucoma. This involves creating a generic finite element model of the optic nerve head and modeling the living tissues of the scleral canal as an adapting mixture of components—collagen tissues, axons, etc. He found, first, that the mixture would create an LC all by itself under or-

Lamina Cribrosa. *This photograph of a plastic rendering of lamina cribrosa was produced on a StrataSys 3-D printer. Courtesy of: J. Crawford Downs, PhD, Christopher A. Girkin, MD, and Kenneth R. Sloan, PhD, University of Alabama at Birmingham.*

"I look at glaucoma and other eye diseases and one thing that strikes me—and it's the theme of my work—is that the biomechanical mechanisms are occurring and interacting across very different length scales," Grytz says.

dinary IOP. And that at elevated IOP, the LC thickened by recruiting collagen from the tissue behind the LC. Eventually, the simulated tissue reached a steady state where it no longer thickened. This was the first study to show that biomechanics could drive the growth and remodeling of the LC in a way that matched observations of early glaucoma in monkey eyes.

Grytz acknowledges that a major challenge remains: determining the mechanism that leads to the insult to the axons. Thus far, he has simulated the human eye coupled with the micro-architecture of the LC and axons passing through LC's porous structure. Though still in the early stages, he says the simulation is fully coupled and shows that when he elevates IOP, the axons are highly shear deformed around the edges of the LC. "But there's so much we still don't understand, and I believe computer simulations will be a great benefit to explore this problem," he says.

Looking Deep into the Eye:
Retina modeling and simulation

The retina is perhaps the most complex component of the eye. In addition to 100 million photoreceptors and more than a million nerve cells that send visual signals to the brain, the retina is a highly layered anatomical and physiological structure. Existing models of the retina tend to simplify it into a slab of cells, says Vaghefi. "That's not accurate."

Luthert, who calls himself a "would-be" modeler of the retina, agrees. His "bugbear" with existing metabolic models of the retina is their use of systems biology without a spatial domain. "Biology occurs in a cell with a finite size and constraints about what can move where," he says. "The cell in turn sits in a tissue with blood vessels that are constrained in space." As a result, he says, multiscale modeling is needed to capture the metabolic and spatial elements within one integrated model.

So, a few years ago, Luthert joined with others to propose a multimillion-dollar grand scheme for computational modeling of the retina. The plan included several building blocks: a model of the cells that come together to create the outer retina; a model of blood flow in the retina; and the incorporation of imaging data to refine the models and then make them patient-specific.

Though the European Commission's Research and Innovation Department didn't fund the grant, Luthert remains completely committed to the plan. "It's absolutely the way forward," he says. "The major treatment for diabetic retinopathy involves injections into the eye with an intrinsic risk," he says. "The lack of clarity about when or how often to inject lends itself nicely to a computational approach."

He acknowledges the plan was perhaps a bit ambitious. "So what we're doing now is chopping it up into bits and running with those," he says. For example, one of his graduate students is modeling blood flow in the choroid—the remarkably rich vascular bed at the back of the retina. And Vaghefi and his colleagues are using various imaging methods to try to build models that are more anatomically and physiologically correct. "It's a move from simple equations put down in the '80s to make the models more precise."

Meanwhile, **Abbas Shirinifard**, **PhD**, a computational modeling scientist at St. Jude Children's Research Hospital in Memphis, Tennessee, has been building a multiscale model of the outer retina that covers multiple spatial and time scales. He simulates the cascade of events that lead to a common form of age-related macular degeneration—a primary cause of vision loss in older Americans—known as "wet" macular degeneration.

Scientists know that the wet form of macular degeneration begins with blood vessels invading the retina from a blood-rich zone (the choriocapillaris) that normally nourishes the retina rather than destroying it. Researchers have hypothesized that this process is caused either by an increase in a protein called VEGF (vascular endothelial growth factor) in response to inflammation or injury; or by breaks in the Bruch's membrane—a physical barrier between the capillary layer and the outer cell layer of the retina (the retinal pigment epithelium, or RPE). To explore these possibilities, Shirinifard modeled the cells of the choriocapillaris, the Bruch's membrane, and the RPE—and included more than 30 parameters related to transport of oxygen, diffusion, and cell behavior.

In the simulations, his team found that increasing VEGF did cause capillaries to invade the retina, but in the wrong way. Normally, in humans, blood vessels invade the space between the RPE and the Bruch's membrane first, and then sit there for a while before invading the retina. But in the simulation, there was no pause: "The cells jump in to invade the retina without any invasion of that sub-RPE space," Shirinifard says. "So that was shocking and surprising." Simulating holes in the Bruch's membrane produced the same result. "Both hypotheses failed—or my model was wrong," he says.

In an attempt to improve his model using experimental data, Shirinifard ordered some postmortem human eyes. When they arrived, some had been crushed in shipping. The damaged eyes had some patches of detached RPE and some patches that were still attached to the Bruch's membrane. Shirinifard then had an "aha!" moment: Cellular adhesion could be a key factor in macular degeneration.

In his simulation, Shirinifard began perturbing various adhesion properties (there are multiple types of adhesion—with and without mechanical coupling, for example) between different components at the back of the eye. The result: His simulations produced several different patterns of capillary invasion that are actually seen in people.

Validating the simulation is quite complicated,
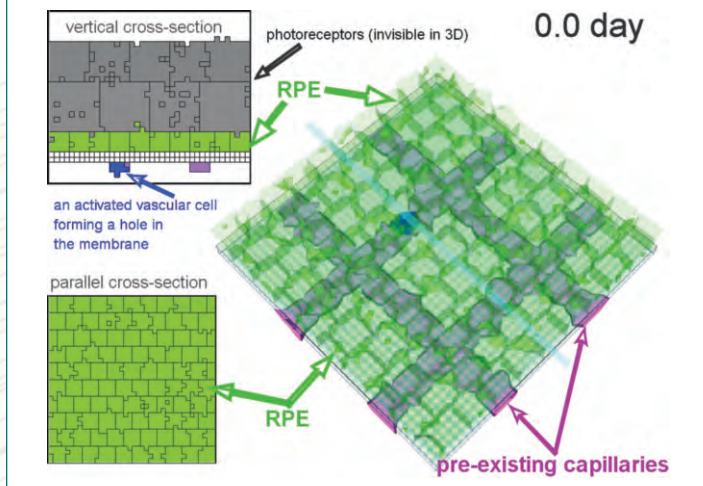
Shirinifard notes. In a clinical setting, there is typically just a snapshot of a moment in time, whereas the simulation shows a snapshot and also the evolution of a 3-D structure. "There's not much data to compare it with," he says.

Currently he is turning his attention to the inner retina, where blood vessel changes occur in diabetic retinopathy. "There's more data and access to lots of parameters for building these kinds of models," he says. "It's a unique opportunity."

All Eyes:
A vision for a physiome model of the eye

Peter Hunter, **PhD**, of the University of Auckland, has long been interested in creating a physiome model of the eye—a complete computational system that would capture all of the eye's complexity from the molecular and cellular levels up to the tissue and organ levels.

Luthert agrees there could be great value in an integrated model. "To some extent it's okay to have just the pieces of the eye modeled," he says. "But if you take one step back, it's clear that things are very interconnected." For example, although an integrated model of the outer retina is "utterly compelling" and could help us understand macular degeneration, it would benefit from integration with models of IOP and fluid flow through the vitreous. And, he says, "Front-of-the-eye glaucoma stuff generates input for the back-of-the-eye glaucoma stuff." Thus, models of how pressure became elevated in the first place—which involves aqueous production and retention in the front of the eye (models not discussed here)—could inform models of how the optic nerve head responds to increased IOP. "It's these truly complex disorders that need a whole eye approach," he says.

"It's not a bad idea," Pinsky says. "The components of the eye have to interact in very intimate ways. Many of us are experts in one area or another. The idea of a more comprehensive understanding is important to me. It would be a sensible scientific goal." □
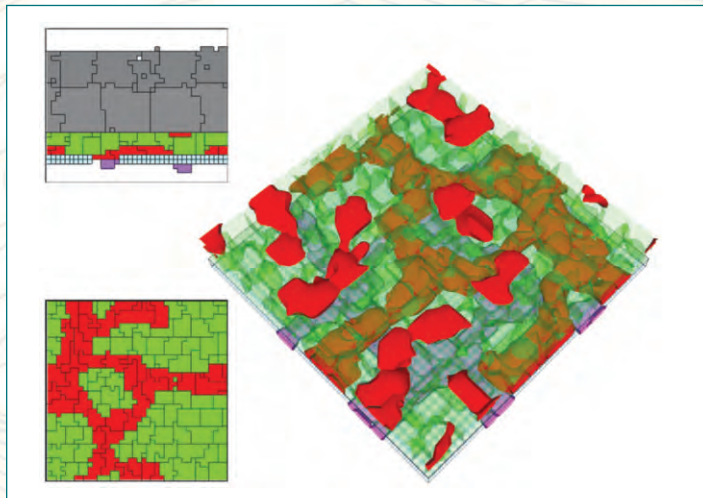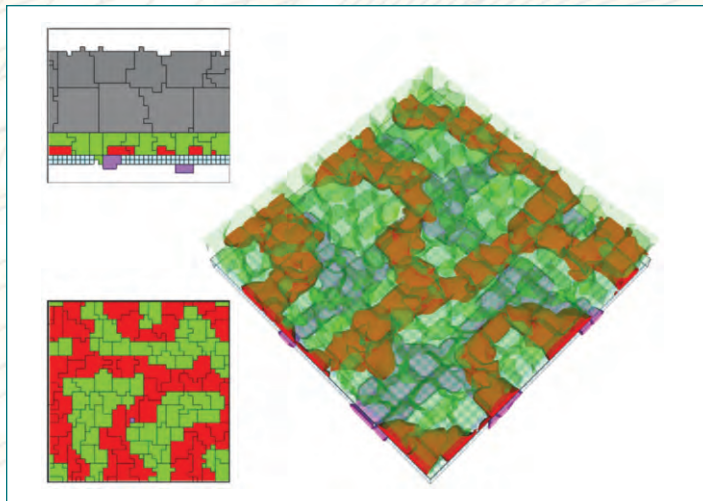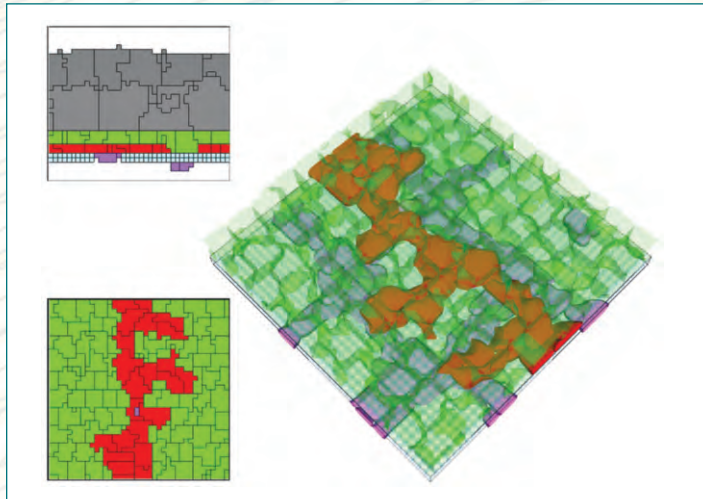
Simulating Capillaries Invading the Retina. Shirinifard ran many thousands of simulations of capillaries invading the RPE to see how quickly certain changes to the parameters lead to various types of vascular invasion of the choroid in wet macular degeneration. These four snapshots show one such simulation over the course of a year (months 0, 6, 9, and 12) in 3-D as well as in a 2-D vertical slice (upper left) and a 2-D horizontal slice at the level of the sub-RPE space. Vascular cells are shown in red; RPE in green; and the Bruch's membrane in light blue. Month zero starts with an activated vascular cell (purple) forming a hole in the Bruch's membrane. At 6 months, blood vessels have spread into part of the sub-RPE space. At 9 months, the sub-RPE space is well infiltrated by blood vessels. And at 12 months, these vessels are now invading sub-retinal space—the zone between the RPE and the photoreceptors. Courtesy of: Abbas Shirinifard, based on data presented in Shirinifard A, et al., Adhesion Failures Determine the Pattern of Choroidal Neovascularization in the Eye: A Computer Simulation Study. PLoS Comput Biol 8(5): e1002440. doi:10.1371/journal.pcbi.1002440 (2012).

BY KATHARINE MILLER

Visualizing Markov State Models Using MSMExplorer

An unfolded protein can move through thousands of intermediate structures (conformations) before finding its properly folded state. One approach to understanding this process involves simulating a protein's behavior with molecular dynamics, and then analyzing the resulting trajectories statistically to build a Markov state model or MSM, which describes a series of conformational states and the probabilities of transitions between them. Interpreting these MSMs can lead to new insights. Now, a new application is available to address a key bottleneck that has slowed the process of interpretation: the lack of a simple, automated way to visualize MSMs.

"With a picture, you gain a rapid way to qualitatively analyze MSMs," says **Bryce Cronkite-Ratcliff**, an undergraduate in computer science at Stanford University. He worked with **Vijay Pande**, **PhD**, professor of chemistry, structural biology and computer science at Stanford, to create MSMExplorer, an application for visualizing MSM data. "It can guide further quantitative analysis in new directions," Cronkite-Ratcliff says. And as a bonus, the application can generate graphics for publications and posters, a task that most MSM researchers currently do by hand.

To create an MSM of a protein folding simulation, a program like MSMBuilder—also out of Pande's lab and also available at simtk.org—runs simulations to watch how often a protein goes from one state to another, and clusters intermediate structures based on kinetic proximity—*i.e.*, how energetically easy it is to change from one structure to another. The resulting data matrices must then undergo extensive quantitative analysis to be understood. Visualizing the data as a network can guide researchers to ask relevant quantitative questions. "MSMExplorer is not a replacement for quantitative analysis, but it lets you see overall patterns that are otherwise hard to tease out of the numbers," Cronkite-Ratcliff says.

A few years ago, for example, **Greg Bowman**, **PhD**, then a graduate student in Pande's lab, crunched the numbers for a lot of MSMs and discovered the presence of kinetic hubs. These kinetic hubs represent a new and more physically accurate way to conceptualize how proteins fold, where the native (folded) state plays the role of a central hub (gathering point) kinetically, not just thermodynamically. The work was very labor intensive. But if MSMExplorer had existed at that time, Cronkite-Ratcliff says, the discovery would have been much more obvious. Because the visualization displays various protein-folding states and their connections in sizes that reflect their likelihood of occurring, the hubs pop out. "It's hard to look at a matrix or do calculations on a matrix and detect whether there's a kinetic hub," he says. "But it's pretty easy to open a picture and see whether there's a giant circle in the middle with spokes going out from it."

Certainly other software programs exist that can build visualizations of networks—including Cytoscape and Gephi, to name just two. But MSMExplorer knows how to collect the MSM files it needs and automatically turn them into a useful visualization "right off the bat," Cronkite-Ratcliff says. It also does a few things those other programs can't. For example, Cronkite-Ratcliff says, MSMExplorer includes some custom tools such as transition path theory (TPT) algorithms that
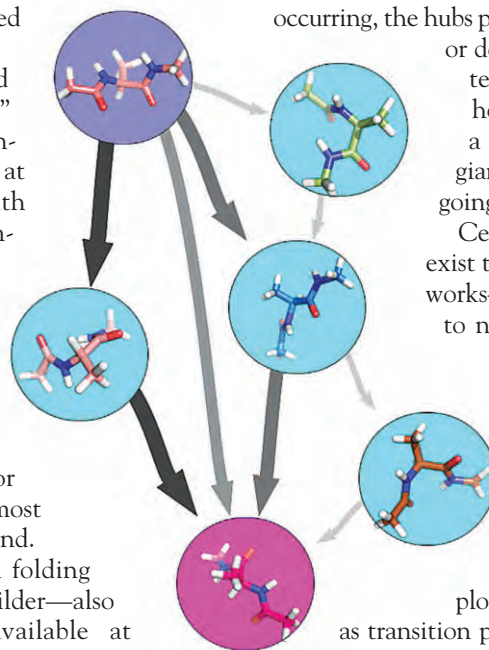
MSMExplorer uses transition path theory (TPT) to depict the likelihood of a protein folding via various intermediates. Courtesy of: Bryce Cronkite-Ratcliff.

extract and display the likely paths a protein will take to get from node A to node B, as well as how much activity occurs along each path. The nodes can also be displayed as images of the folded states they represent, a feature that can help users produce publication-worthy graphics. "You can make them look less like a hairball," Cronkite-Ratcliff says.

According to Pande, the MSM technique is already used in 20 to 30 labs around the world and many others are interested in adopting it. "A way to visualize MSMs will empower many researchers now and many more in years to come," he says. □

A Simbios website providing open access to high quality biocomputational tools, accurate models and the people behind them

Stanford University
318 Campus Drive
Clark Center Room S231
Stanford, CA 94305-5444

seeing science

SeeingScience

BY KATHARINE MILLER

autoPACK Visualization Challenge

Three-D animators have long sought algorithms that can pack odd-shaped things into tight spaces. Now, **Graham Johnson, PhD**, a QB3 Faculty Fellow in bioengineering at the University of California, San Francisco together with Arthur Olson's Molecular Graphics Lab at The Scripps Research Institute has created an algorithm called *autoPACK* that can pack anything into anything—including stuffing molecules into cells to visualize how they interact in space. And, with support from the CG Society of Digital Artists and Autodesk, he is challenging both professional animators and the scientific community to use the algorithm as part of the autoPACK Visualization Challenge (http://autopack.cgsociety.org).

The Challenge provides participants with the necessary ingredients to produce visualizations of HIV in blood serum using open-source models that are as biologically accurate as possible and constantly being updated. The goal: to convey hu-

manity's complex relationship with HIV in either a short film (under two minutes) or a JPG image.

Johnson hopes the competition will not only build a large community of users and developers who become hooked on the program, but also attract both artists and professional, Hollywood-caliber 3-D animators from outside biology to help build on the open-source project. "They will most want the generic packing algorithm, but our core distribution will always come with the biological applications as part of the GUI," he says. "And working with DNA and proteins as part of the

These renderings show an early version of the autoPACK Visualization Challenge model of HIV with the blood serum and lipid bilayer turned off. Moving the camera around to show different views we see (clockwise from top): HIV matrix capsid proteins with one spike protein sticking out; a nucleocapsid of HIV; half of the spherical virus with five visible spike proteins; HIV RNA with an RNA binding protein that is critical to HIV's lifecycle. Images courtesy of Thomas Brown and Graham Johnson.

competition will provide a biological hook into industry that we hope will continue long after the contest is finished." □