computationally intensive."

FDTD has been around for a while, but it has been applied to biology only within the last few years. It's a numerical modeling technique that can be applied to interactions between electromagnetic waves and objects whose structural details are small compared to the wavelength of light. "Inside and in the vicinity of the target object, divide your space into a 3-D grid system and divide time into small steps," Lu says. "When the light hits the object, the electric and magnetic field distributions at each point in the grid space are calculated for each time step. Then put everything together to calculate the scattering pattern."

> "We're looking inside the cell without opening it." —**Jun Qing Lu**

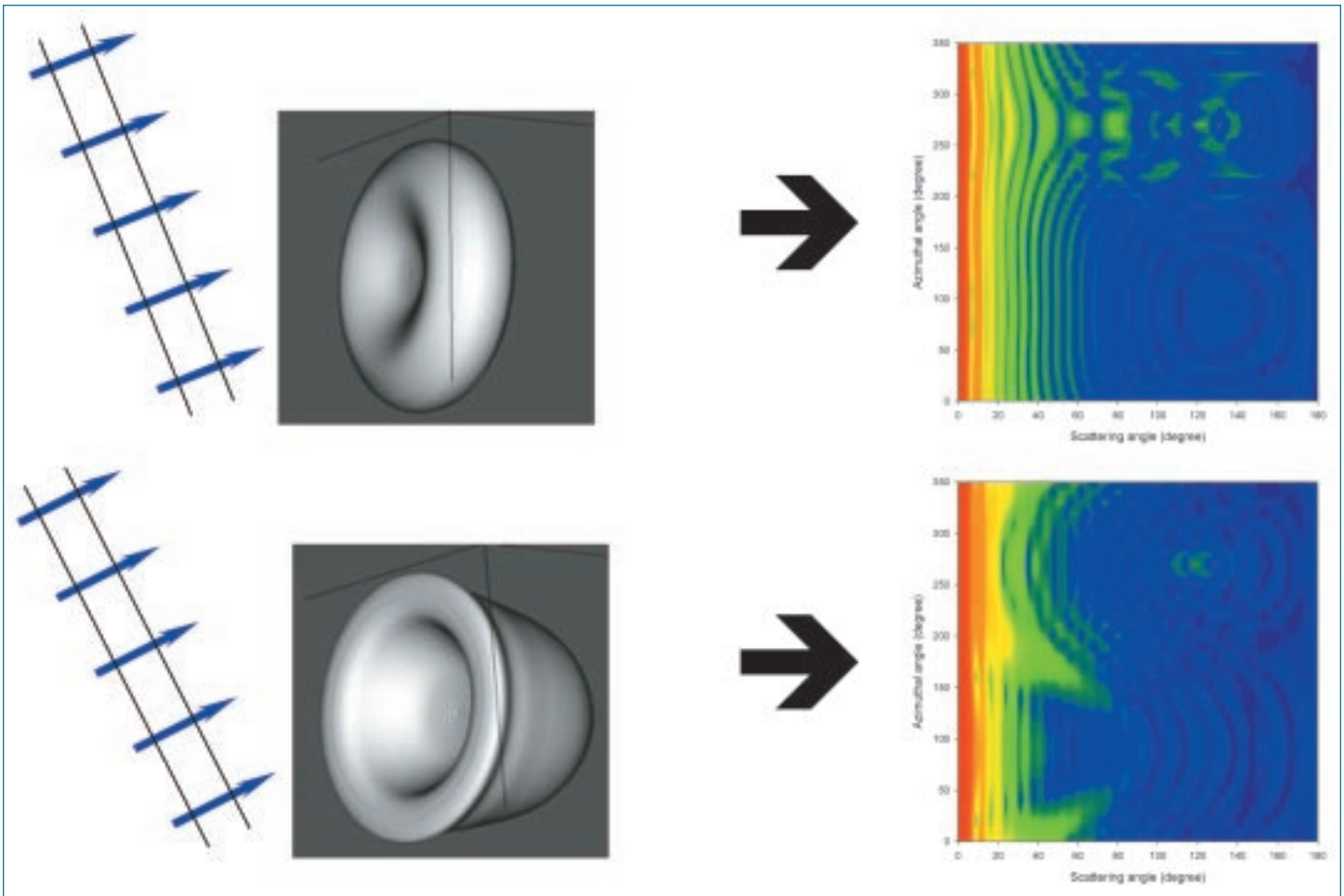
With about a million grid points, about two thousand time steps, and six finite difference equations for each grid point, it's clear why the process requires lots of computational power. If you also want to see how the light scattering changes with different cell types or the same cell in a different life stage, that requires even more power. "Parallel computing makes it faster," Lu says.

Lu and her colleagues work side by side doing computational modeling and experimental work. "I'm a theoretician," Lu says. "But I have scientists by my side doing experiments. So far, the models match reality pretty well."

Thus far, Lu's group has been studying light scattering by individual cells. Eventually, they will use the FDTD technique to do tissue studies—with hopes of distinguishing tumor from non-tumor. "People are showing lots of interest in this method," says Lu. "It's the right direction to pursue."

Integrative Cancer Biology Program is Born

The National Cancer Institute launched the Integrative Cancer Biology Program (ICBP) in October 2004, providing a total of $15 million to nine multidisciplinary centers. The goal: to use predictive cancer modeling to better understand how the disease



Light scatters differently from normal and deformed red blood cells.

develops and progresses.

"Only high-level computation can handle the explosion of information that we've seen in the last ten years as a result of genomics, proteomics and molecular imaging," says Daniel Gallahan, PhD, associate director of the Division of Cancer Biology at the NCI. "Cancer is such a complex problem that we really have to approach it with all the tools in our arsenal. By modeling how cancer develops from initiation to metastasis, we hope to predict and better understand the cancer process."

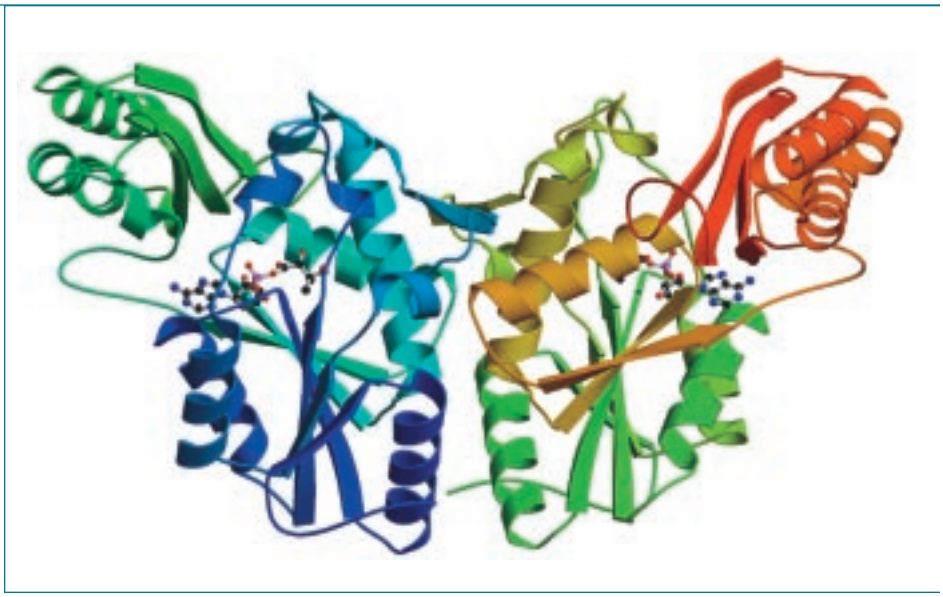
Until now, cancer researchers have used computation only in a fragmented way. "Hard-core modeling hasn't been addressed in the cancer community," Gallahan says. "There has been some modeling of cell migration, some statistical analysis of microarrays, and some modeling of risk factors and predictors, but nothing at the level that we're taking it to with the ICBP."

Making the leap to more complex computation means that the cancer biologists who head up each of the nine centers had to enlist experts from other fields. "All of these grant applications had to include computation on an equal footing with biology," Gallahan says.

Initially, the projects will be taking the steps necessary to integrate vast amounts of genomic, proteomic, imaging, and other data so that they are usable. Each center will then develop computational methods to make models that address a specific set of biological problems.

The nine centers cover the gamut of the cancer process—from initiation through signaling, DNA repair, tumor progression, invasion, angiogenesis and metastasis. One center, at Harvard, will be doing three-dimensional modeling of the tumor itself.

In principle, the ICBP should first lead to models at each step of the cancer process, but ultimately, Gallahan says, these should become modules that can be integrated. "Once these models are available in a modular way, we would then piece them together and look at how the cell transforms," he says. "By increasing our understanding of the cancer process, the models will help us identify and design better prevention and treatment strategies."



Model of an enzyme, PanC, which is involved in the last step of vitamin B5 biosynthesis in M. tuberculosi. PanC is essential for the growth of M. tuberculosis, and is therefore a potential drug target. Credit: Mycobacterium Tuberculosis Center

A Crescendo of Protein Structures

A ten-year, $600-million program known as the Protein Structure Initiative (PSI) has already, in its five-year pilot phase, greatly increased the speed at which protein structures can be determined, and added 1100 structures to the Protein Data Bank (PDB). Several thousand more may be added over the next five years. Completion of the project should lead to more rapid determination of protein function.

"The key is to make protein structures useful by getting them out there and in the hands of scientists all over," says John Norvell, director of the PSI at the National Institute of General Medical Sciences (NIGMS), which funds the project. "Lots of interesting science will come from this large collection. It will allow people to think in structural ways when designing experiments or hypotheses. It will permit better attack on protein-folding problems. And it will lead to better and quicker work on target drug designs."

A few thousand protein structures might not sound like a lot, given that the PDB—a federal repository for structural information about proteins—already contains about 30,000 structures. But the large majority of the banked structures are closely related to one another.

According to Jerry Li, MD, PhD, program director at the Center for Bioinformatics & Computational Biology at the NIGMS, "We really have only a few thousand structures that are relatively unique," says Li. "We

> The PSI is producing a catalog of structural information not only about a large number of proteins but about a larger variety of proteins than had previously been examined.