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."

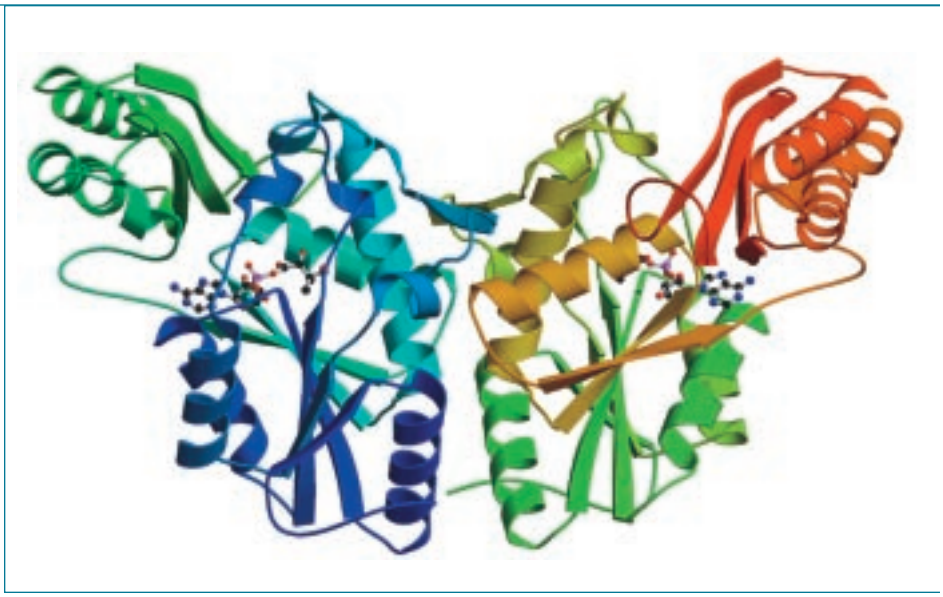
Until now, cancer researchers have used computation only in a fragmented way. "Hard-core modeling hasn't been addressed in the cancer community," Gallahan says. "There has been some modeling of cell migration, some statistical analysis of microarrays, and some modeling of risk factors and predictors, 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.

need a whole lot of structures that are not so homologous to each other."

That's why the PSI targets representatives of a wide range of protein families. As a result, 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.

For 50 years, scientists have been determining the structure of proteins in order to better understand their function, but the PSI marks a shift in how structural biology is done. "The PSI is discovery-driven rather than hypothesis-driven." Norvell says. "We're systematically sampling the universe of protein structures."

PSI's efforts have also reduced the cost of determining protein structures, from $420,000 per protein down to about $125,000. Norvell hopes to reduce the cost even further to under $100,000 or even as low as $50,000.

The program is now moving into its second phase, with plans to identify more protein structures in two ways—in the lab and *in silico*. Under one set of grants, production centers will be established to elucidate 4000 or more additional protein structures over the next five years. Meanwhile, another set of grants will focus on improving methods for computational modeling of protein structures. The shapes of protein family representatives (PSI's experimental targets) serve as rough templates for the other structures in the family, which will be determined using computer-based homology modeling.

"In the end," says Li, "the PSI will generate a few thousand experimental structures, but it will produce tens of thousands of modeled structures."

Spit Diagnostics

If spit could talk, it might tell us whether we're sick or healthy.

According to David Wong, DMD, DMSc—professor and associate dean of research at the School of Dentistry at the University of California, Los Angeles—the protein profile in our saliva might distinguish a person with oral cancer or breast cancer from one who has neither disease. That's why the National Institute of Dental and Craniofacial Research last fall funded grants to Wong's group and two others who will identify all of the proteins in human saliva.

Because spit can be collected noninvasively, Wong says, it is ideal for diagnostic testing. But there's a hitch: "Saliva contains all the information we know is in blood, but at much lower magnitudes," Wong says. "So you need different tools to measure it."

In recent years, those tools have been developed. Seven groups around the country, including at UCLA, have been building biosensors for saliva diagnostics.

Wong and his colleagues have already determined that the RNA transcriptome profile in the saliva of people with oral cancer is markedly different from that of healthy controls. So why bother with the proteome? "At the end of the day, we'll have genomic, proteomic, and transcriptomic information," Wong says. "The question will be