

Structural Biology Looks at the Ties That Bind

Experts in biochemistry, genetics, physical chemistry, and computational modeling are working together to understand the mechanistic basis for disease.

NOT so many years ago, no one knew how cancer and many other diseases occurred. Over time, scientists learned that smoking can cause lung cancer, overexposure to sunlight can cause skin cancer, eating too much of certain types of foods may lead to heart disease, and so on. But even when they knew what caused disease, they still did not know how the change took effect in the body.

It has only been in the last 10 years that researchers at Lawrence Livermore and elsewhere have discerned that subtle, permanent alterations to DNA cause changes in proteins and other biological molecules, sometimes leading to cancer and other diseases. In fact, the very act of living—of eating and breathing—can expose DNA to harmful agents that result in damage to genes and ultimately to proteins.

Humans produce as many as 100,000 different protein molecules, each of which is a long, folded chain of amino acids. Proteins activate essential chemical reactions, carry messages between cells, fight infections, control the growth and differentiation of cells, regulate the activity of genes, and provide structural and mechanical support. They also provide the motion required in cell division, muscle contraction, and cell propulsion, and they generate and transmit nerve impulses.

The link between proteins and DNA is strong: the amino-acid sequence of

each protein is specified by a unique DNA base sequence in the coding region of a single gene. Mutations in the DNA sequence may be caused by small molecules, called chemical mutagens, that appear everywhere in our environment and bind to the DNA bases. Changes resulting from mutations in the DNA base sequence of a gene can produce proteins that function abnormally and result in disease.

Scientists have known that changes in genes resulted in the production of proteins that did not function properly. But they had to know the specific structure of these proteins before they could make the technical advances needed to detect human disease and cancer successfully and design new drugs and treatment therapies. While amino-acid sequences of more than 20,000 proteins have been deposited in data banks that are available to medical researchers, complete three-dimensional structures have been identified for less than 5 percent of them.

The Need for a Closer Look

The impetus for Lawrence Livermore's Biology and Biotechnology Research Program (BBRP) Directorate to establish a structural biology capability was its work on the human genome, especially DNA repair processes and DNA damage.

Proteins known as DNA repair enzymes constantly scan the genome for

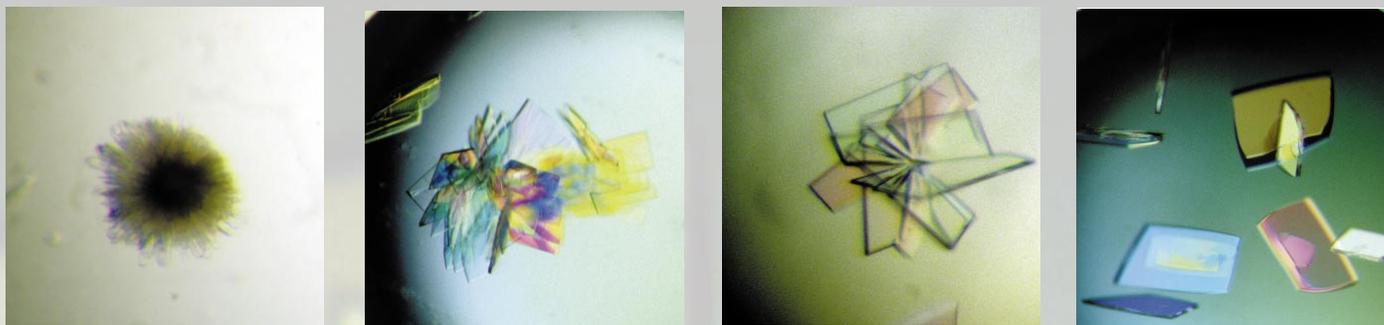


Figure 1. This series of images shows the gradual improvement in crystal quality as crystals of a benign portion of the tetanus toxin are grown under varying crystallization conditions. In the photograph on the opposite page, Mark Knapp and Sabine Ringhoffer collect data from a crystal like those on the far right above for use in Livermore's structural biology research efforts.

damage, remove the defective region of the molecule, and resynthesize the missing segments of DNA. But sometimes, the repair process stops working, or damage may be too great for the repair process to overcome. Unrepaired DNA damage eventually produces mutations that may trigger the growth of malignant tumors. Livermore scientists under Larry Thompson had been researching this repair process for 20 years. At the same time, Andy Wyrobek, Jim Felton, and others were studying DNA damage itself, in sperm and from food mutagens.¹ They had learned, for instance, that eating certain foods may cause mutations in DNA, changes that could later give rise to cancers.

Biochemist Rod Balhorn, who has spearheaded much of the structural biology work at Livermore, says, "After almost 15 years of research, both groups knew that they needed more information." They required a better look at the proteins responsible for DNA repair to figure out precisely how they recognize, bind to, and replace damaged segments of the DNA molecule.

Thus, in the mid-1990s, with funding from the Department of Energy's Laboratory Directed Research and Development Program, BBRP began developing a structural biology capability. They brought in experts in x-ray crystallography and nuclear magnetic resonance spectroscopy, which are the only methods for obtaining high-

resolution, three-dimensional information about individual molecules. Bernhard Rupp set up an x-ray crystallography laboratory, while Monique Cosman established a laboratory for nuclear magnetic resonance spectroscopy. Their teams began providing experimental data on protein structures, some of which are used by another new group under Mike Colvin that performs molecular modeling. Yet another new group led by Krzysztof Fidelis specializes in predicting the structure of proteins from information about the amino-acid sequences that are encoded in DNA.

Today, under the leadership of Jim Felton, these groups support a number of projects at BBRP. Some of them are a continuation of previous work, including identifying how chemical mutagens damage and perturb the structure and function of DNA as well as characterizing the structure of proteins that recognize and repair DNA damage. A newer project with the Gladstone Institute of San Francisco is identifying how mutations in proteins involved in lipid (fat) metabolism and plaque formations in the brain relate to cardiovascular and neurodegenerative diseases, especially Alzheimer's disease. The results of these and other studies are helping scientists understand why particular individuals are susceptible to cancer and certain diseases and how DNA repair proteins interact with and repair damaged DNA.

Another project is part of the Laboratory's work to reduce the threat of biological weapons. Scientists in BBRP are working to obtain high-resolution structure and function information for tetanus and botulinum toxins, which belong to the same family of bacterial toxins. Structural information is playing an important role in the development of antidotes, detection systems, and other countermeasures for minimizing the threat of exposure to biological warfare agents.

Examining in 3-D

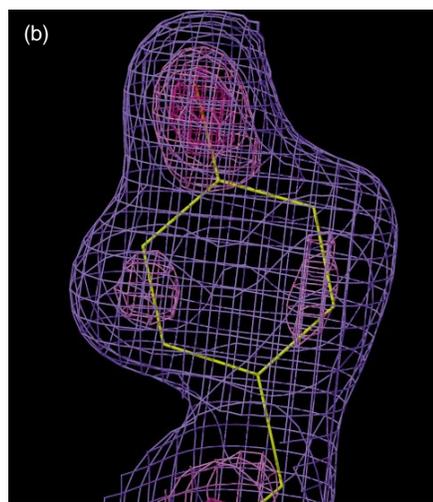
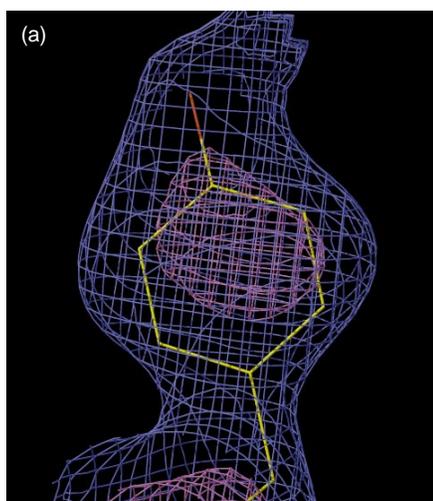
X-ray crystallography and nuclear magnetic resonance spectroscopy operate in very different ways, but both can determine the locations of the individual atoms that make up a biomolecule.

X-ray crystallography exploits the fact that x rays are scattered by the electron cloud around each atom in a crystallized molecule. Based on the diffraction pattern obtained from the assembly of molecules or atoms in the crystal, the electron density of the crystal's individual components can be reconstructed, resulting in a very accurate model of the crystallized protein's molecular structure.

Rock crystals or crystals of salt or sugar are hard objects because of their regular atomic structure. Protein molecules have irregular, folded shapes and produce fragile, soft crystals that resemble tiny jelly cubes (Figure 1).

Their fragility makes them sensitive to environmental variations and to radiation, including x rays. Flash-cooling to almost the temperature of liquid nitrogen (-196°C) eliminates their sensitivity to radiation.

Protein x-ray crystallography of large molecules has been around for 50 years, but advances are being made all the time to achieve higher and higher resolutions (Figure 2). Because the highest resolution data come from the highest power x-ray sources, Rupp and his team have used such DOE



accelerators as the Advanced Light Source at Lawrence Berkeley National Laboratory to achieve the highest possible resolutions. Work is also under way at Livermore to develop advanced computational methods for processing the data collected by x-ray diffraction.

Nuclear magnetic resonance (NMR) spectroscopy involves the interaction of the magnetic “moment” of each atom’s nucleus with an external magnetic field. When a molecule is placed in a magnetic field, the field will align the spins of the nuclei either parallel or antiparallel to the field, with each spin having a discrete energy level. Transitions can be induced between high- and low-energy states by the application of a radio-frequency perturbation, and a resonance signal for each spin can be detected. Because the chemical environment significantly modifies the properties of a nucleus, the position of an NMR signal can provide information about the structure and dynamics of a molecule.

Series of radio-frequency pulses and delays are designed to manipulate the nuclear spins and their interactions with neighboring spins. In this way, NMR spectra are generated containing

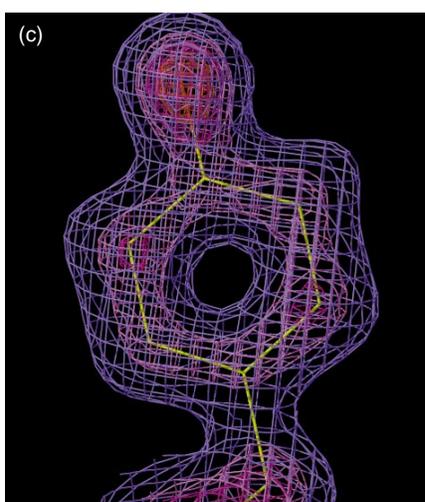


Figure 2. Images of the electron density of a molecule at three resolutions—(a) 3 angstroms, (b) 2 angstroms, and (c) 1.1 angstroms. The higher the resolution, the more accurate the model of the molecule.

information about the distance and angles between nuclei that are separated in space and/or through chemical bonds (Figure 3).

The two methods complement one another, providing different kinds of information to researchers. X-ray crystallography works with solid materials and results in very fine detail of molecules that are frozen in time. NMR spectroscopy, on the other hand, uses molecules in solution, which means that they are in motion. Spectral data is averaged to give information on the movement of atoms in the molecules in relation to one another.

Both x-ray crystallography and NMR require considerable time to reduce experimental data to usable structural information. After successful growth of a crystal, x-ray diffraction patterns can often be obtained in less than a week, but the actual definition of molecular structure from these data may require several years of effort. Similarly, the NMR spectra needed to identify the structure of a small protein can be obtained in a few weeks, but many months may be required to analyze and assign the data before the structure can be calculated.

Predicting Structure

Because of the time requirements for determining protein structure with x-ray crystallography and NMR, computational modeling and simulation methods have been used for many years to augment experimental efforts. Because these techniques are so computationally intensive, they have benefited enormously from the recent dramatic increase in computer performance—in particular, the development of massively parallel computers—and concomitant software developments. Using these computer advances, scientists can today model much larger molecular systems than before.

Mike Colvin’s computational biochemistry effort makes use of two primary modeling methods: quantum

chemistry (based on fundamental quantum mechanics) and empirically based molecular dynamics models.

Modeling with quantum chemistry allows the calculation of extremely accurate chemical structures and reaction energies. Until recently, this method was limited to small, simple molecules, but compounds with up to several dozen atoms can now be studied on inexpensive personal computers, while massively parallel computers are used for compounds with up to several hundred atoms. Molecular dynamics simulations involve much larger molecules, typically with tens of thousands of atoms. The two methods work together to constantly refine the model. Quantum chemical calculations are used to generate force fields and atomic charges for molecular dynamics simulations, which in turn are used to determine local structural constraints that are used in accurate quantum chemical-energy calculations.

Together, molecular dynamics and quantum chemistry are being used by Colvin's group to study a number of biological problems, including the mechanisms of enzymes that repair damaged DNA as well as drugs that bind to the DNA of cancer cells. The molecular dynamics simulations are used to determine the large-scale changes in the DNA helix due to damage or drug binding. Then, quantum chemical simulations are applied to smaller segments of the modified DNA to give more accurate energies and structural properties (Figure 4).

Krzysztof Fidelis and his colleagues are taking an entirely different tack to predict protein structure. Their approach works with whole proteins, which can involve tens of thousands of atoms. Furthermore, the method uses the sequence of amino acids and its environment in the protein as a starting point.

Two predictive techniques—comparative modeling and fold

recognition—operate on the proven assumption that similar amino-acid sequences will produce similar protein structures. With these methods, predicting the structure of an unknown protein would include a visit via computer to data banks containing information on known protein structures.

A third technique that Livermore has not yet used starts closer to ground zero: it combines sequence data with known physical and chemical properties of individual amino acids to predict the structure of the complete protein. Says Fidelis, "If scientists can predict even small structures with this method, it means they really know something about protein structure."

In 1994, Lawrence Livermore, together with researchers at the University of Maryland and Sandia National Laboratories, established an international organization for the prediction of protein structures. Today, Livermore is home to the Protein Structure Prediction Center, which acts as a clearing house for an ongoing



Figure 3. Monique Cosman at work with her team in Livermore's nuclear magnetic resonance (NMR) laboratory. Team members are Steve Chan (foreground) and (background, left to right) Kin Yan, Kevin Thornton, and Viswanathan Krishnan. (Inset) A solution-state, three-dimensional structure of a fatty acid binding protein as determined using NMR. The thickness of the lines provides information about the motions of the atoms in the molecule.

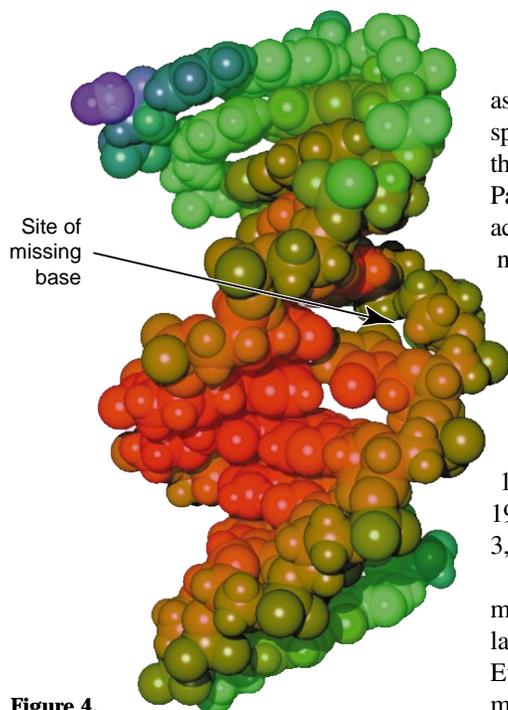


Figure 4. Model of a DNA double helix with a base missing in the middle. This type of damage occurs by natural processes thousands of times each day in every cell in the body and must be repaired to maintain good health. Each atom in this simulation is colored according to the amount it moved during a molecular dynamics computer simulation (red moved least; blue moved most). The damage seems to affect the DNA flexibility and is thought to play a role in the repair of DNA.

assessment of prediction methods and sponsors a biennial conference to discuss the most successful methods. Participating researchers receive amino-acid sequence information for a set of new structures that have been determined either by x-ray crystallography or NMR spectroscopy but not yet released to public data banks. Later, predictions are compared to laboratory results, often with excellent results (Figure 5). This is clearly a growing effort. In 1994, there were 130 predictions; in 1996, 980 predictions; and in 1998, 3,800 predictions.

The strong dynamic between these modeling and predictive efforts and laboratory experimentation is evident. Even with the largest computers, modeling cannot stand entirely on its own. It needs validation from experimental results in an ongoing, iterative process that constantly refines modeling results and methods.

New Inhibitors for Toxins

A recent structural biology success story at Livermore involves the tetanus toxin, a member of a family of toxins that could be used by an aggressor or terrorists as biological warfare agents. BBRP's goal is to learn how to develop

inhibitors for these toxins in case one of these bacteria is used in a biological attack.

Inhibitors are protective drugs that stop or slow the biological action of a toxin or other damaging molecule. Think of the protease inhibitors that patients with HIV receive. Inhibitors are weaker and easier to develop than antidotes, which reverse a toxin's damage after the fact. Inhibitors might be used if an exposure is anticipated, and they require constant dosing.

Tetanus is a paralytic disease caused by a neurotoxin produced by the anaerobic bacterium *Clostridium tetanii*. It is just one of a whole family of clostridial neurotoxins that are believed to have a similar cell invasion mechanism. All the deadly botulinum toxins belong to this family.

The tetanus toxin targets the membranes of the central and peripheral nervous systems to block the release of neurotransmitters, causing the nerve cells to fire constantly. The result is muscle rigidity—thus, the common name for tetanus, “lockjaw.” An effective inhibitor for the tetanus toxin must stop the toxin from binding to cells in the nervous system.

Tetanus and other clostridium family toxins have two parts: the light chain, which contains the enzymatic portion of the toxin and is responsible for its toxic effects, and the heavy chain, which binds to the neuron and aids delivery of the light chain to the interior of the neuron. The heavy chain has two parts or domains. The binding domain binds to gangliosides, which are sugar-based recognition molecules on the nerve cell membrane. The translocation domain makes a pore in the cell through which the toxin may pass.

Considerable research at several institutions has established the propensity of the binding domain to bind to gangliosides. But what had not been determined was which part of it bound to the ganglioside.

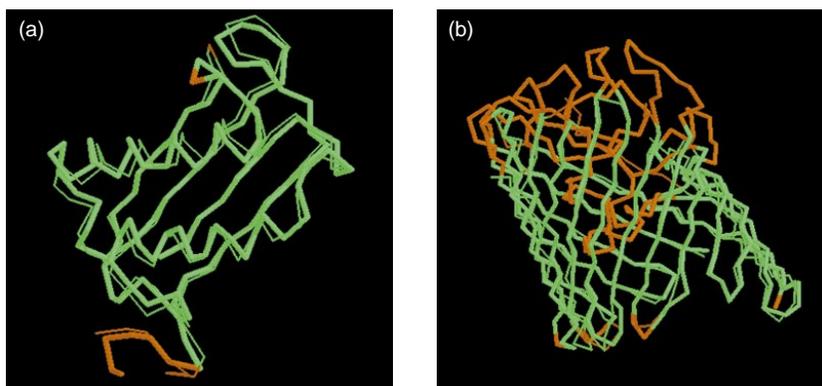


Figure 5. Comparison of the modeling prediction with the actual structure of two proteins, (a) human D-dopachrome tautomerase and (b) *C. acidovorans* OMP32. The thick lines represent the structure as determined by x-ray crystallography, while the thin lines correspond to atoms in the prediction. Regions colored green are correctly predicted to 3.5 angstroms.

Knowing the precise site of binding and what the site looks like is important. For an inhibitor to be effective, it must bind at the same site, which means that its molecular structure must fit there as neatly as does the toxin's binding domain. If binding by the toxin can be blocked, penetration of the cell will be stopped.

A major accomplishment in 1998 was the high-resolution structure determination of the binding domain of the tetanus toxin by Livermore's x-ray crystallography group. (The three parts of the toxin can function separately, so it is possible to do this research without working with dangerous, intact toxins.)

With the high-resolution protein structure in hand, researchers on Colvin's computational biochemistry team collaborated with scientists at Sandia National Laboratories to



Figure 6. High-resolution (1.56 angstroms) structure of the binding domain of tetanus toxin. This structure shows that this portion of the protein has two separate parts. One binds to sugars called gangliosides present on the surface of motor neurons. The function of the other is unknown.

computationally select compounds that might fit in the same binding site. They were able to quickly identify 30 compounds predicted to bind to the tetanus toxin protein from a database of 250,000 compounds (Figure 6).

Moving from those 30 possible compounds to an approved inhibitor drug will involve a long process that will likely take years. Rod Balhorn is currently testing the 30 compounds using mass spectrometry to see if they bind to the tetanus binding molecule. While the testing is incomplete, he and his colleagues have already discovered seven new molecules that will bind to the toxin. These compounds will be bound to the toxin, and the site of binding will be determined by x-ray diffraction or NMR spectroscopy. Armed with these data, a pharmaceutical company can then develop an inhibitor drug that is specific for this toxin.

The invasion mechanism of toxins might someday be put to another use entirely. The light chain, which carries the toxin, could be reengineered to remove the toxin portion of the molecule and add a drug. The formerly deadly protein could thus become a life-saving, drug-delivery vehicle. The drug might be designed to target specific cells, for example, cancer cells with anticancer drugs.

Experts Finding Solutions

With its strength in physical sciences and international recognition for work in genomics and DNA repair, Lawrence Livermore was ideally suited to develop capabilities in structural biology. Experts in biochemistry, genetics, physical chemistry, and computational modeling are working together to understand the mechanistic basis for disease. Molecular medicine is a new and rapidly evolving field and one in which Lawrence Livermore is beginning to play an important role.

—Katie Walter

Key Words: clostridium toxins, computational biochemistry, DNA repair, nuclear magnetic resonance spectroscopy, protein structure prediction, tetanus, x-ray crystallography.

References

1. "Food Mutagens, Part I," *Science & Technology Review* (July 1995), pp. 4–25, "Food Mutagens, Part II," *Science & Technology Review* (September 1995), pp. 5–23, and "The Genetic Contribution of Sperm: Healthy Baby or Not?" *Science & Technology Review* (November/December 1995), pp. 6–19. These articles are available on the Web at www.llnl.gov/str/str.html.

*For further information contact
Rod Balhorn (925) 422-6284
(balhorn2@llnl.gov)*

About the Scientist



ROD BALHORN is a senior biomedical scientist in Lawrence Livermore's Biology and Biotechnology Research Program (BBRP) Directorate. He joined the Laboratory in 1974 after receiving a B.S. in chemistry and a Ph.D. in biochemistry from the University of Iowa, where he was also a postdoctoral fellow. He was instrumental in initiating a structural biology capability at Livermore as a logical outgrowth of BBRP's human genome research into DNA damage and repair processes. He is the coauthor of numerous publications reporting advances in protein biochemistry, chromatin organization, and structural biology research and is currently working on the application of structural biology findings to the search for inhibitors of the deadly botulinum toxin.