Appendices

Appendix A
Technology Fundamentals

Static vs Dynamic Structures

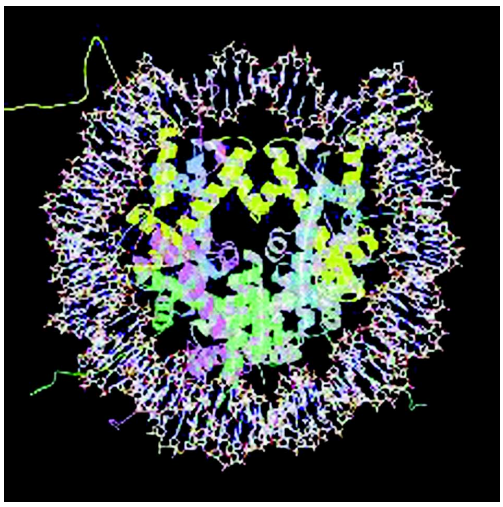
Over the years, the laborious and painstaking process of protein crystallography has resulted in an increasingly large number of protein and nucleic acid structures. Crystallography is essential for studying the 3-D structure of these biomolecules and for revealing some of the mechanisms of their biological activity. However, the structures solved by crystallography are static: they are a snapshot image of the molecule's motion and chemical activity.

In contrast with time-averaged still images from protein crystallography, other analytical tools can provide dynamic structural information, something that could be considered a motion picture of a molecule's behavior. These tools include a variety of spectroscopic techniques such as fluorescence studies, electron paramagnetic resonance, and nuclear magnetic resonance; and scattering techniques such as quasi-elastic and dynamic light scattering, and X-ray and neutron small-angle scattering. These techniques usually cannot provide the same level of molecular detail as that of crystallographic studies, but the measurable changes in structure associated with molecular activity can yield essential insights into how a molecular machine does its job. As a result of increasingly sensitive detection limits, observing molecules directly within a cell is now becoming possible using fluorescent labels, for example. Being able to observe proteins in vivo, without perturbing their natural environment, offers the ultimate means of understanding the many processes that occur within living cells.

Molecular Models from Crystallography

The most common experimental method for obtaining a detailed picture of a protein or protein complex is to interpret the diffraction of X rays from many identical molecules in an ordered array commonly referred to as a crystal. The method thus is called single-crystal protein X-ray crystallography. This experimental technique provides information on the positions of individual atoms within a biological complex. Having a detailed structure or "form" of the macromolecule will aid in beginning to understand its function.

Determining a protein's structure by X-ray crystallography consists of growing high-quality crystals of the purified biomolecule, measuring the directions and intensities of X-ray beams diffracted from the crystals, and using computers to transform the X-ray measurements. This method produces an image of the crystal's contents in much the same manner as a microscope's objective lens. Computers and crystallographer take the place of the microscope's objective lens because no lens can focus the highly divergent beams diffracted from the crystals. Finally, the image must be interpreted, which involves computer graphics to display the electron density of atoms in the molecule and the construction of a consistent molecular model.



Molecular structure of the nucleosome core complex

Recent developments in synchrotron radiation sources have revolutionized protein crystallography, opening the door to high-throughput protein-structure determination. These intense tunable X-ray sources have allowed the development of the Multi-wavelength Anomalous Dispersion (MAD) technique for solving the phasing problem. MAD enables the collection of data in mere minutes compared with the many hours required when conventional X-ray sources were used. To make effective use of synchrotron sources for protein studies, however, new approaches are needed for efficient, high-throughput production of protein crystals. In addition, improvements in detectors, data interpretation, and graphics display all will enhance the quality of the molecular model that is the end product of the macromolecular crystallography process.

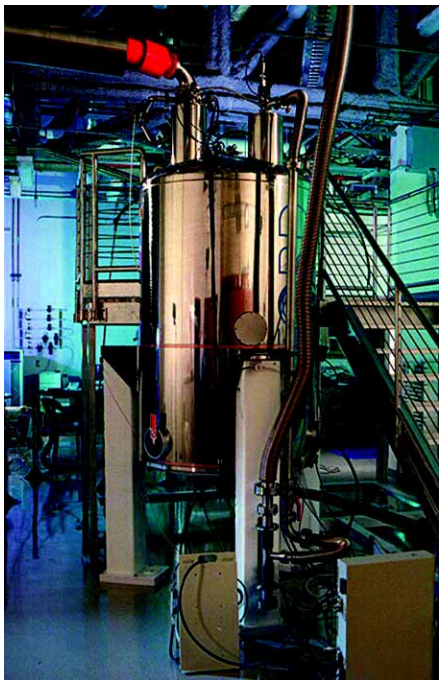
Neutron crystallography is a valuable technique to use when details of an enzyme mechanism or binding site for drug development are needed. Although not a high-throughput procedure, neutron crystallography can visualize the hydrogen atoms (about half the atoms in a protein structure). It can also visualize many of the more-mobile water molecules that cannot be seen, even with ultrahigh-resolution synchrotron X-ray radiation.

The accompanying figure shows the molecular structure of the nucleosome core complex, the chromosome's basic building block. This fundamental repeating unit is made of a complex of eight separate protein molecules and two strands of DNA that carry a piece of the genetic code, the blueprint for life. This is the longest segment of DNA ever seen at near-atomic resolution.

Nuclear Magnetic Resonance Spectroscopy

Nuclear magnetic resonance (NMR) spectroscopy uses high magnetic fields and radio-frequency pulses to manipulate the spin states of nuclei—including 1H, 13C, and 15N—that have nonzero-spin angular momentum. For a molecule containing such nuclei, the result is an NMR spectrum with peaks whose positions and intensities reflect the chemical environment and nucleic positions within the molecule. As applied to protein-structure analysis, the accuracy now achievable with NMR spectroscopy is comparable to that obtained with X-ray crystallography.

In several respects, NMR spectroscopy offers a technique complementary to X-ray crystallography and neutron diffraction. An important consideration is that NMR structures typically are obtained from proteins in solution, with no requirement that the protein be crystallizable. Not only does this lead to a protein-structure representation unconstrained by any crystal lattice, but it also allows structures to be determined for proteins that cannot be crystallized. The latter point is especially significant because a substantial fraction of all proteins are thought to contain long, disordered regions (>40 residues) that may prevent crystallization. In these cases NMR may



Environmental Molecular Sciences Laboratory's 800-mhz NMR spectrometer at Pacific Northwest National Laboratory

be the best, perhaps the only, method available to characterize the structures.

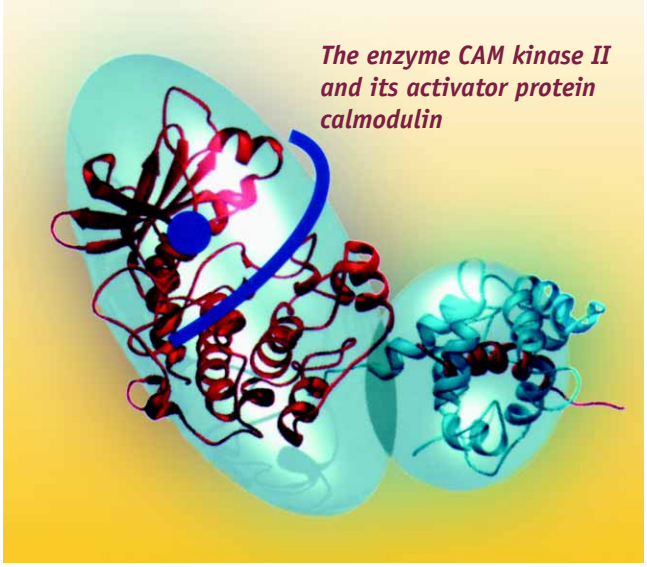
On the other hand, NMR historically has been limited in two important ways. First, protein-structure determination by NMR methods has been limited to relatively small proteins, that is, those smaller than about 40 kD. Second, data collection for a single protein structure typically has required weeks, as compared with the minutes or hours needed for X-ray crystallography. Recent advances on four fronts have significantly lessened these limitations, however. They are (1) development of instruments using higher magnetic fields and higher radio frequencies, thus improving sensitivity and resolution; (2) novel use of isotopic labeling, and (3) development of sophisticated experimental methods that can differentially manipulate nuclear spins, thus enabling studies of proteins up to 150 kD [K. Pervushin et al., *Proc. Nat. Acad. Sci.* **94**, 12366 (1997); R. Reik et al., *Proc. Nat. Acad. Sci.* **96**, 4918 (1999)]. The fourth factor is the significant advances in NMR probe technology that have improved sensitivity by three- to fourfold and substantially reduced data-collection times. Collectively, these advances are having a significant impact on experimental strategies being employed in protein studies; larger protein molecules are being analyzed, proteins with low solubility are being studied at lower concentrations (~0.2 mM), and data for well-behaved proteins of <40 kD are being collected more rapidly (in about 2 days). Other significant advances are being made in the automation of data analysis, reducing the time needed for protein-structure determination from weeks to days.

Neutron Scattering

Neutron scattering helps to resolve how the 3-D parts of protein machines fit together and how proteins communicate in dynamic regulatory and signaling networks. While nuclear magnetic resonance (NMR) spectroscopy and X-ray crystallography provide high-resolution structural information on individual subunits of protein and protein-DNA complexes, neutron scattering provides lower-resolution information on the shapes and arrangements of these subunits in solution, thereby complementing these other tools.

Although X-ray crystallography on its own has delivered a few spectacular examples of high-resolution structures of molecular machines (such as the structure of the ribosome), such structural achievements can take decades of work and require crystalline forms of the proteins studied. Further, if the protein is mechanically flexible, crystallizing the protein without distorting it can be difficult. Neutron scattering, however, analyzes protein complexes in solution, bypassing this potential complication. Together with NMR and X-ray crystallography, neutron scattering enables researchers to develop comprehensive high-resolution models of working molecular machines.

In addition to helping structural biologists fit together the interlocking pieces of complex molecu-



The enzyme CAM kinase II and its activator protein calmodulin

lar machines, state-of-the-art neutron scattering capabilities also can reveal the conformational dynamics of proteins. For example, neutron scattering has shown how the DNA in a nucleosome complex unwinds from the histone core as the solution chemistry changes. Similarly, neutron scattering can monitor conformational changes in an enzyme as a substrate binds. By combining high-resolution data on individual components (from NMR or X-ray crystallography) with neutron scattering's lower-resolution information on overall shapes and positions, scientists can view protein complexes in different functional states, along a signaling pathway, or at different stages of an activation mechanism.

The physical basis for neutron-scattering experiments lies in how neutrons interact with the atoms in a sample. Neutrons are electrically neutral particles scattered by atomic nuclei. The neutron-scattering power of different elements—and even different isotopes of an element—is a complex function of the properties of the compound nucleus that forms briefly between an incoming neutron and the atomic nucleus that scatters it. As a result, the neutron-scattering power does not vary smoothly with the atomic mass of the scattering nucleus. (In contrast, X rays are scattered by electrons, and thus the scattering power of X rays increases monotonically with the atomic number.)

The neutron-scattering signal from a biological macromolecule such as a protein or DNA molecule in solution is proportional to the "contrast" between the molecule studied and the solvent. This contrast is the difference between their neutron-scattering densities, which are calculated by summing the neutron-scattering lengths of all the atoms in the macromolecule or in the volume of solvent containing the macromolecule and then dividing by that volume.

One of the biggest differences in neutron-scattering power is between hydrogen and deuterium. In addition, hydrogen has a neutron-scattering length opposite in sign to those of most elements found in biological molecules. By manipulating the ratios of hydrogen to deuterium in the protein or DNA subunits of a sample, one can "tune" the scattering densities of different subunits relative to the solvent so that the subunits can be made to "disappear" or "appear" selectively. This contrast-variation technique enables researchers to study the shapes of individual protein or DNA subunits within large complexes and determine the relative positions of the subunits.

Neutron scattering shows much promise for structural analyses of protein complexes, but it faces several technical limitations. For example, techniques are needed for inexpensively producing relatively large quantities of soluble, deuterium-labeled samples, which requires robust protein-expression systems that give high yields in deuterated media. In addition, current applications are limited by the relatively low fluxes of neutron beams produced by reactors or accelerators. The development of more powerful neutron sources, such as the Spallation Neutron Source and the higher-intensity cold neutron source planned for the High Flux Isotope Reactor also at Oak Ridge National Laboratory, will give scientists access to state-of-the-art and eventually next-generation instruments.

Imaging Technologies

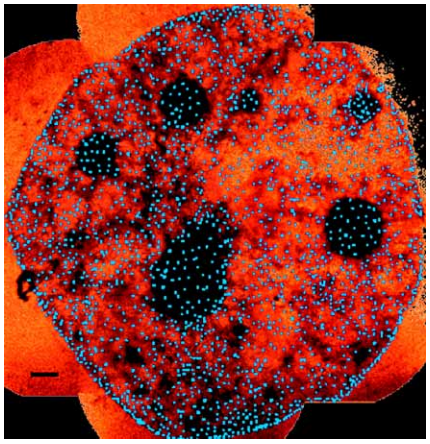
Understanding a complex living system will require a thorough comprehension of the interactions of cells and tissues in the organism. And understanding those cells will necessitate an integrated understanding of all functional units—signal transduction molecules, structural scaffolding, and genetic material. The molecular machinery of life must be studied at all size scales from atoms to complete organisms. Extensive information about the proteins that make up the cells' functional units can be obtained through the use of molecular biology, crystallography, and computational biology. But understanding their function within their natural environment—the cell—will require examining these proteins within the cell, through all phases of cell behavior.

Imaging is a very powerful unifying tool for such studies. Light-microscopic analyses of fluorescently tagged markers yield critical information about the location and behavior of proteins, as well as many details of protein-protein interactions. The information obtained, however, is limited by the light-microscope's resolution (150 to 200 nm). Better resolution can be obtained using electron microscopes, but that approach requires more elaborate cell-processing procedures and typically is limited to sectioned or very thin specimens. An array of imaging techniques will be needed if science is to understand the function of proteins in cells, the behavior of cells, and, ultimately, whole organisms.

Confocal Microscopy. At relatively low resolutions, confocal microscopy can produce three-dimensional (3-D) images of fluorescently tagged gene products to determine their distribution in the cell during different stages of the cell cycle or under various environmental conditions. Such information allows deep insights into cell and organelle biology. Furthermore, confocal microscopy permits analysis of the cell's 3-D architecture, which cannot be achieved by conventional light microscopy. The broad goal is to visualize cellular constituents and general cytoarchitecture in a state as close to native organization as possible.

The availability of laser-based confocal microscopes and the imaginative exploitation of green fluorescent protein from jellyfish have provided new tools of great diversity and usefulness. Watching a protein bind its substrate or its partners in real time with submicrometer resolution within a single cell is now possible. The importance of such processes as self-organization and the assembly of subcellular organelles is well recognized. Self-organization at the intermediate level of multimeric protein complexes is open to inspection.

X-Ray Microscopy. Soft X-ray microscopy, using X rays produced at synchrotron light sources, is an emerging biological imaging technique for the examination of intact, hydrated cells. The shorter wavelengths of X rays permit resolution 5 to 8 times better than that achieved by light microscopy, and the information obtained from the image contrast is highly quantitative in nature. Protein location can be determined at better than 50-nm resolution in whole cells, as shown in the image of the nucleus below, in which nuclear pore proteins located in the



Soft X-ray image of a human mammary cell

membrane surrounding the nucleus are labeled blue. Entire cells can be examined and information about the localization of a specific protein determined without extensive processing. X-ray cryotomography facilitates 3-D reconstructions of cells—a technique comparable to CT scans of the brain and other parts of the body—and precise localization of proteins in the cells. Future developments will enable studies of specific proteins in living cells using light microscopy, followed by determination of their ultrastructural localization using soft X-ray tomographic imaging. Single-cell studies with these techniques will combine the power of protein-specific, live-cell, fluorescent light microscopy with X-ray microscopy's unique high-spatial–resolution, whole-cell imaging methods for high-throughput analyses of protein function in cells.

Electron Tomography. At still higher resolutions, electron tomography is the most widely applicable method for obtaining 3-D information. Whereas the wavelength of visible light limits the resolution of light microscopy to hundreds of nanometers, the wavelength of intermediate-voltage electrons is only a fraction of an angstrom. Electron tomographic microscopy is therefore the only method suitable for examining such structures as many supramolecular assemblies, organelles, and cells that vary in structure from one to another. The method recently has been applied to the study of cryofixed and stained sections of mitochondria and secretion organelles. Furthermore, with the development of automated low-dose data-acquisition schemes, molecules and cells embedded in vitreous ice can be studied. This opens new horizons for investigating the functional organization of cellular components with minimal perturbation of the cellular context. The unmatched spatial resolution provided by electron microscopy complements the temporal resolution provided by light-microscopic techniques, which allow movements of molecules to be tracked in vivo.

In contrast to resin embedding and sectioning, preparation in vitreous ice yields a preserved state closer to the native state of biological samples. The sample is spread into a film only a few hundred nanometers thick across holes in a perforated carbon support and plunged into liquid ethane, thereby forming vitreous ice throughout the thin specimen layer [J. Dubochet et al., *Quart. Rev. Biophys*. **21**, 129 (1988)]. This procedure almost instantaneously immobilizes all cellular components in their native, frozen-hydrated environment and allows them to remain stable in the high vacuum of the electron microscope. In principle, image interpretation is also more straightforward because in stained samples—at least on a molecular scale—the relationship of stain distribution to underlying biological structure is not clear. Specimen contrast in vitreous ice is much lower, however, and the total electron dose is limited roughly to 2000 to 6000 e^-/nm^2 because of the samples' radiation sensitivity. Radiation damage results in structural changes at the molecular level and, at higher doses, in bubble formation. Using a computer to optimize focusing and tracking of the sample during the tilt series, however, has enabled the study of virus particles, lipid vesicles with or without cargo, and macromolecules without significant damage.

Theoretical analysis suggests that electron tomographic reconstructions to about 2-nm resolution should be possible [J. Bohm et al., *Proc. Nat. Acad. Sci*. **97**, 14245 (2000); R. Grimm et al., *Biophys. J*. **74**, 1031 (1998)]; this would allow the identity, location, and even conformation of many proteins to be seen. In the best published example, however, resolution was estimated at about 6 nm, sufficient to locate and identify only the largest multiprotein complexes [W. Baumeister et al., *Trends Cell Biol*. **9**, 81 (1999)]. Even better reconstructions, perhaps as good as twice the theoretical limit, are widely anticipated for the newest generation of

electron microscopes operating at 300 kV and equipped with field-emission electron sources, liquid helium–cooled specimen stages, and energy filters [A. J. Koster et al., *J. Struct. Biol.* **120**, 276 (1997)].

A serious problem with specimens thick enough to provide useful 3-D information is that many electrons lose energy in passing through the specimen. Energy filtering improves image contrast by selecting electrons within a narrow energy window, thereby minimizing loss of resolution and contrast due to chromatic aberration. When ice-embedded specimens are investigated, the best contrast and resolution generally are achieved by selecting electrons that have lost no energy. The long-term goal of these studies is to detect macromolecular structures in their native environment, thus providing insights into their cellular function. Whole-cell tomography using the next generation of intermediate-voltage, field-emission–gun, energy-filtered electron microscopes will provide detailed 3-D information about the distribution of gene products tagged with labels absorbing at specific electron energies.

Magnetic Resonance Microscopy. Confocal or optical microscopy (OM) and magnetic resonance microscopy (MRM) have developed as important tools for cellular research. MRM is noninvasive and nondestructive, and OM requires only the expression or uptake of fluorescently labeled molecules for detection. Both methods have their advantages and disadvantages. MRM provides access to several observable quantities that cannot be determined with OM alone (e.g., metabolite concentrations, chemical shifts, spin couplings, T1 and T2 relaxation times, and diffusion constants). These quantities have been related to a variety of such cellular events as tumor formation, programmed death (apoptosis), necrosis, and increased proliferation. Instruments combining OM and MRM allow live cells to be studied simultaneously using both techniques, providing a necessary link between cellular response and molecular information on proteins and other biochemicals involved in a certain cellular event. Two combined OM-MRM microscopes are under development at Pacific Northwest National Laboratory.

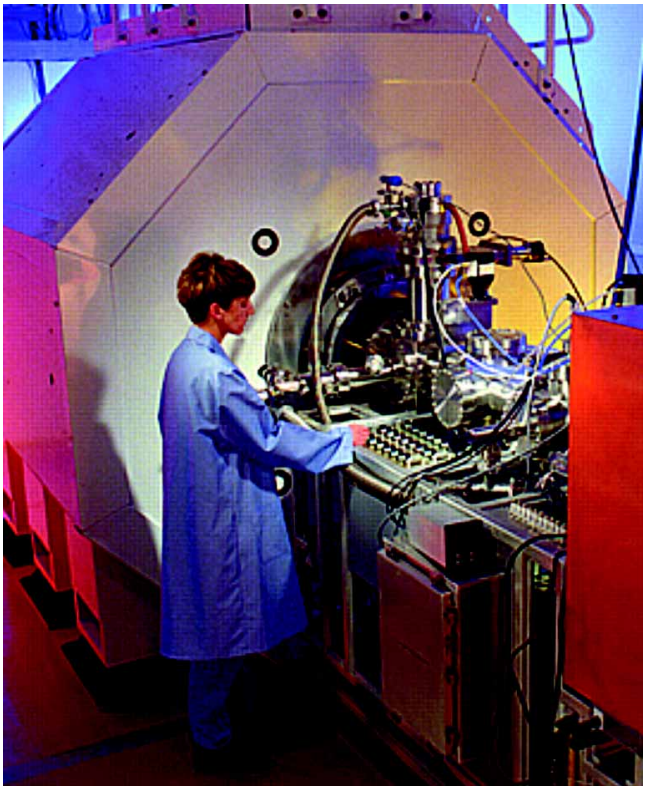
Mass Spectrometry

Over the past decade, mass spectrometry (MS) has become an important tool for the analysis of proteins. In its simplest form, MS sorts and measures the mass of individual ions (charged molecules). Ions can be formed from proteins using either electrospray (ES) or matrix-assisted laser desorption ionization (MALDI), each of which typically adds one proton (in the case of MALDI) or many protons (ES) to the protein. These positively charged protein ions can be analyzed directly to establish the protein's mass. Alternatively, the ions can be fragmented while inside the mass spectrometer by techniques such as collision-induced dissociation to provide more detailed information on the protein. Of particular relevance to the Genomes To Life program is the unique ability of MS to identify a protein unambiguously, establishing the amino acid sequence (the order in which these building blocks of proteins are arranged) and determining the presence of post-translational modifications that can impact the protein's function.

Scaling up MS from one-protein-at-a-time analysis to high-throughput proteome-wide analysis is of high importance to the Genomes To Life program. Currently, the majority of MS-based techniques for proteome analysis are linked with two-dimensional (2-D) electrophoresis. These 2-D separations typically combine isoelectric focusing

and sodium dodecyl sulfate polyacrylamide gel electrophoresis to separate the many proteins found in a cell. This approach offers two advantages. First, 2-D gel technology allows the visualization of a large number of proteins simultaneously. And second, comparing a 2-D gel from one organism (or cell) with that from another organism allows differences in expressed proteins to be observed clearly. A wealth of information regarding the isolated protein then can be obtained by excising individual spots, digesting proteins, and analyzing the resulting fragments by MS, typically using MALDI and a time-of-flight mass analyzer. The resulting MS data can be correlated with protein, genome, or expressed sequence tag databases. This technique will become increasingly popular as more and more genomic sequence data become available.

A major drawback of the 2-D gel approach, however, is that the entire process is quite labor- and time-intensive. The gel work alone can require hours to days of effort. Each protein has to be removed from the gel and prepared individually for MS analysis. Efforts to automate many of the spot-excision, digestion, and sample-preparation steps involved in 2-D gel and MS assays are being pursued. Other drawbacks include the limited dynamic range, quantification capabilities, and usefulness in analyzying hydrophobic proteins.



Mass spectrometer in Environmental Molecular Sciences Laboratory at Pacific Northwest National Laboratory

Recently, alternative techniques for analyzing entire proteomes have shown great promise by addressing many shortcomings of 2-D gel electrophoretic strategies. One such technique is isotope-coded affinity-tag peptide labeling. In this technique, quantitative differences can be measured readily in the levels of proteins expressed in a reference organism and in one grown with an isotope label (i.e., 18O or 15N). The isotopically labeled peptides are separated by liquid chromatography (LC) and then analyzed directly online by ES MS, providing full identification of the proteins. This technique, which has the advantage of eliminating the labor-intensive 2-D gel electrophoresis step, is potentially scalable to high-throughput techniques and is amenable to the analysis of minor expression products in complex mixtures ("large dynamic range").

Another MS-based approach that shows great promise is multidimensional LC coupled with MS. By use of 2-D capillary LC columns composed of a strong cation exchanger combined with a reverse-phase resin, samples of proteins in complex mixtures can be introduced directly into the MS using ES and the MS data can be produced in a fully automated fashion. This flowing 2-D LC approach offers substantial advantages in analysis time over conventional 2-D gel electrophoresis, and it has demonstrated a very wide dynamic range. Another key feature of this emerging technique is that it overcomes the protein-solubility problem often encountered in 2-D gel assays by including a proteolytic digestion step on the entire protein extract before sample loading.

Other MS-based approaches include the use of accurate mass tags derived from high-resolution capillary LC separation, combined with ES ionization and a Fourier transform ion cyclotron resonance (FTICR) MS. FTICR is a specialized MS that has capabilities for both high and accurate mass resolution (about one part per billion). FTICR allows a significant fraction of peptides obtained from a crude protein extract's complete proteolytic digest to be distinguished uniquely by mass alone. The technique thus can identify individual protein components from a complex mixture.

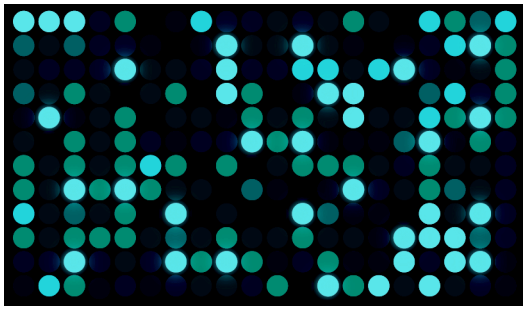
Still other promising new high-sensitivity techniques on the horizon require minimal sample preparation and therefore have potential for high-throughput proteome analysis. Microscale ("lab-on-a-chip") sample-preparation and separation technologies are being examined for online analysis of protein mixtures with MS to minimize sample-handling steps and realize high-throughput protein analysis. This high-sensitivity technique has the additional advantage of requiring minimum quantities of the sample and expensive reagents.

In another approach, the sample-separation steps are virtually eliminated when ion-ion recombination techniques are employed in a quadrupole ion trap (QIT) MS. In this technique, complete cell extracts are introduced into MS, and reactions within the QIT's trapped ion cell are used to simplify the mixture without physically isolating individual components. The resulting spectra are interpreted with the aid of computational techniques.

Although not as mature as those for proteomics, MS-based techniques also are being developed for analyzing protein complexes to rapidly assess the effects of minor protein modifications.

Microarray Technologies: DNA, Proteins, and Beyond

Historically, biochemical assays have focused on analyzing one reaction at a time. In the era of the New Biology, however, higher-throughput techniques are needed to make genomic-scale analyses practical. Microarrays are a promising technique that allows for massively parallel analyses by densely arranging miniscule samples on a glass chip or other solid surface. Most current applications involve analyzing samples with labeled probes and reading the results with a computerized image-analysis system, although mass spectrometry and other "label-less" detection techniques are becoming increasingly available.



Microarrays for simultaneous analyses of tens of thousands of samples at Oak Ridge National Laboratory

Much of the promise of microarrays lies in their small dimensions, which reduce sample and reagent requirements (samples are typically in the submicroliter range) and reaction times, while increasing the amount of data available from a single assay. In addition, through the use of different labels such as multicolor fluorescent tags, multiple tests can be conducted on the same array. The efficiency of microarrays is appealing, but their parallelism offers perhaps the most important benefit; microarrays enable many samples to be analyzed simultaneously, so standardizing data from multiple separate experiments is unnecessary and truly meaningful comparisons can be made.

Microarrays yield information on complex metabolic pathways, detailed genotypes, and the functional context of genes. One well-established use of DNA microarrays is to create transcription profiles, a measure of gene expression. Each microarray consists of a pattern of different known DNA sequences that is "probed" with fluorescently labeled mRNAs extracted from their cDNA complements or from cells. The mRNAs from expressed genes hybridize with the immobilized DNA on the chip. The fluorescence intensities reflect the amount of bound mRNA, which is in turn a relative measure of gene expression. Profiles can be generated readily to determine baseline expression levels, compare expression in cells under different conditions, and compare expression from different genotypes. DNA microarrays also are being used for such large-scale DNA sequence studies as genotyping single nucleotide polymorphisms and for investigating DNA-protein interactions.

Building on the success of DNA microarrays, protein arrays are being developed for high-speed assays of protein function, including protein-protein interactions and ligand-receptor interactions. Similarly, they can be used to screen for antibodies to use as reagents, or antibodies can be arrayed to simultaneously determine the presence and concentration of multiple analytes.

Another emerging technology involves using flow cytometry to analyze suspension arrays of fluorescent microspheres. In suspension arrays, individual array elements are defined by microspheres bearing different amounts of two or more fluorescent dyes, rather than a physical position on a flat surface. Not only can suspension arrays be handled like any other liquid, manually or in an automated system, they can be analyzed rapidly with a flow cytometer.

The success of microarray technologies has driven the development of new instruments and techniques for creating small but incredibly dense arrays; most arrays currently are prepared by robotic application of previously prepared samples or by light-directed in situ syntheses. Microarray technologies also have led to improved methods for rapidly reading and integrating the results of large-scale assays, although further advances are needed to keep pace with the massive amounts of data becoming available. Continued development of these tools will result in even greater miniaturization, sensitivity, and automation in the future.

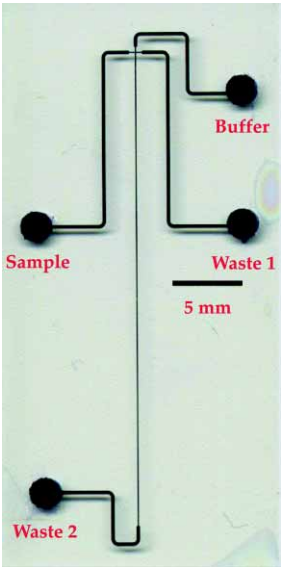
"Lab-on-a-Chip" Microfluidics

A promising analytical tool for analyzing proteins and protein complexes in the biology laboratory of the future is a microfluidic device commonly called a "Lab-on-a-Chip." These "laboratories" are fabricated using photolithographic processes developed in the microelectronics industry to create circuits of tiny chambers and channels in a quartz, silica, or glass chip. They direct the flow of liquid chemical reagents just as semiconductors direct the flow of electrons. These reagents can be diluted, mixed, reacted with other reagents, or separated by capillary electrophoresis or electrochromatography—all on a single chip.

These microfluidic circuits can be designed to accommodate virtually any analytic biochemical process. For example, a lab-on-a-chip for immunological assays probably would integrate sample input, dilution, reaction, and separation, whereas one designed to map restriction enzyme fragments might have an enzymatic digestion chamber followed by a relatively long separation column. Many features of these labs-on-a-chip make them well suited for

high-throughput analyses. Their small dimensions reduce both processing times and the amount of reagents necessary for an assay, substantially reducing costs. Just as microelectronic devices can be manufactured with many elements on a single chip, microfluidic devices can be fabricated with many channels, allowing for massively parallel chemical analyses at a reasonable cost. They are uniquely suited to small-scale analyses; sample volumes for a single experiment often are in the nano- to picoliter range, opening the door to the possibility of analyzing components from single cells.

Relatively simple labs-on-a-chip already are being used for some nucleic acid and protein analyses, but microfluidics technology may someday allow millions of automated biochemical experiments to be performed per day using miniscule quantities of reagents. Eventually, individual analyses may be replaced by protocols in which tens to thousands of analytical measurements are made in parallel, either on the same or multiple samples.



Microfabricated electrophoresis device at Oak Ridge National Laboratory. This "Lab-on-a Chip" electrophoresis device allows mixtures of DNA or proteins to be separated at 1% of the time required by conventional capillary electrophoresis while using much less sample.

Phage Display

Antibodies are well recognized as indispensable tools for recognizing and tracking target molecules. However, traditional methods for preparing antibodies are cumbersome and labor intensive. As a result, researchers are working to develop faster and easier ways to capitalize on the target-recognition qualities of antibodies. Phage display is a new method that enables researchers to quickly evaluate a huge range of potentially useful antibodies and then produce large quantities of the selected ones.

Phage display uses bacteria and bacterial viruses known as phage to produce and select synthetic antibodies that have all the target-recognition qualities of natural antibodies. In fact, these synthetic antibodies are produced using the same genes that code for the target-recognition or variable region in natural antibodies from mammalian systems. The phage are genetically engineered so that a particular antibody is fused to a protein on the phage's coat and the gene encoding the displayed antibody is contained inside the phage particle. This technology thus couples the displayed antibody's phenotype to its genotype, allowing the DNA that codes for the selected antibody to be retrieved easily for future use. Collections of these antibody-covered phage are called a library. Phage libraries each typically contain a billion different antibodies, a number comparable to that in human immune systems.

To select the phage with the desired antibody from a library, the phage are allowed to bind to the target molecule, which is attached to a solid surface. The phage with antibodies that recognize the target molecule bind tightly, and the remaining (unbinding) phage are simply washed away. (Phage display even permits researchers to select antibodies with different binding characteristics for a given target.) The DNA contained within the desired phage then can be used to produce more of the selected antibody for use in research or medical diagnostics.

Appendix B
DOE Partners in Genomes to Life Program

Office of Advanced Scientific Computing Research

The primary mission of the Office of Advanced Scientific Computing Research (ASCR) program is to discover, develop, and deploy computational and networking tools that enable researchers to analyze, model, simulate, and predict complex phenomena important to the U.S. Department of Energy (DOE). This mission is carried out by the Mathematical, Information, and Computational Sciences (MICS) Division. To accomplish its mission, ASCR fosters and supports fundamental research in advanced scientific computing—applied mathematics, computer science, and networking—and operates supercomputer, networking, and related facilities.

ASCR supports the Office of Science Strategic Plan's goal to provide extraordinary tools for extraordinary science and to build a research foundation in support of the plan's other goals. In the course of accomplishing this aim, ASCR research programs have played a critical role in the evolution of high-performance computing and networks.

High-performance computing resources are expected to support the objectives of Genomes to Life through the National Energy Research Scientific Computing Center (NERSC). Platforms at NERSC include an IBM SP RS/6000 system called Gseaborg, a 512-processor IBM RS/6000 SP system with a peak performance of 410 gigaflop/s, 256 gigabytes of memory, and 10 terabytes of disk storage (hpcf.nersc.gov/computers/SP).

ASCR also provides support for Advanced Computing Research Test Beds (ACRTs). The primary objective of an ACRT is to assess the potential of new computing technologies for advancing scientific applications critical to Office of Science missions. The assessment process can, at times, include making the platforms available for specialty computing applications. Some examples of ASCR accomplishments follow.

- Established First National Supercomputer Center. In 1974, DOE established the National Magnetic Fusion Energy Computing Center (predecessor to NERSC) and pioneered the concept of remote, interactive access to supercomputers. Before that time, scientists had to travel to the supercomputer, submit jobs, and wait for hours or days to see the output. The MICS subprogram developed the first interactive operating system for supercomputers, the Cray Time-Sharing System (CTSS), as well as a nationwide network to allow effective computer access to remote users. This revolutionary operating system also enabled users to monitor their jobs as they executed. When the National Science Foundation (NSF) initiated its Supercomputer Centers program in the 1970s, the CTSS operating system was adopted by the San Diego Supercomputing Center and the National Center for Supercomputing Applications to enable access to NSF's first Cray machines.

- Developed High-Speed Interconnects for Supercomputers. To provide a standard interface between supercomputers and such other devices as disk arrays, archival tape systems, and visualization computers, DOE laboratories developed a high-performance network interface. They also led a consortium of vendors to make it the industry standard for the highest bandwidth interconnects between computers and peripheral devices. This advance required the solution of many problems in high-speed signaling, data parallelism, and high-speed protocol design.

Information on ACRTs

- www.csm.ornl.gov/ccs/Falcon.html
- www-unix.mcs.anl.gov/chibaindex.html
- www.acl.lanl.gov/news/releases/99-001.html

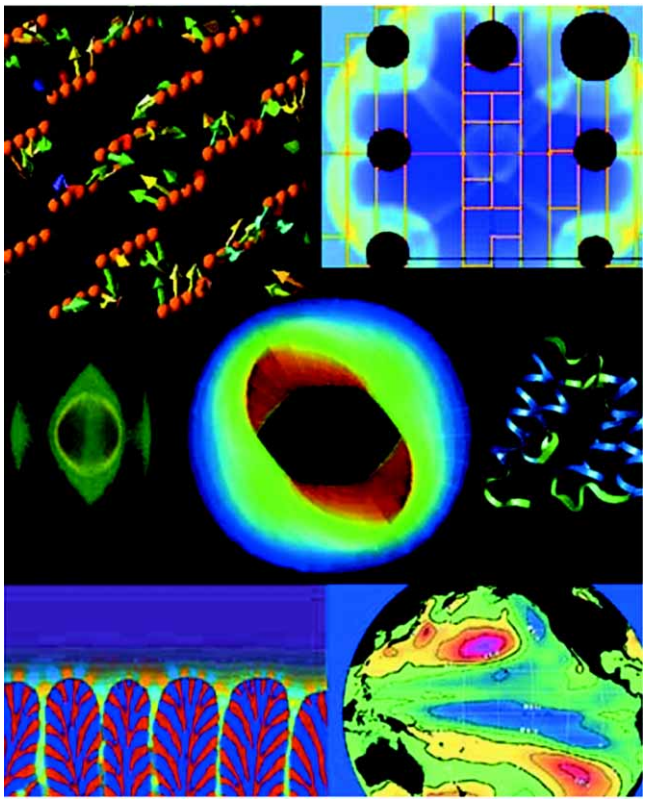
ASCR's mission and accomplishments

- www.sc.doe.gov/production/octr

Scientific Discovery Through Advanced Computing

- www.sc.doe.gov/images/news_photos SDAC_Overview_000330.pdf

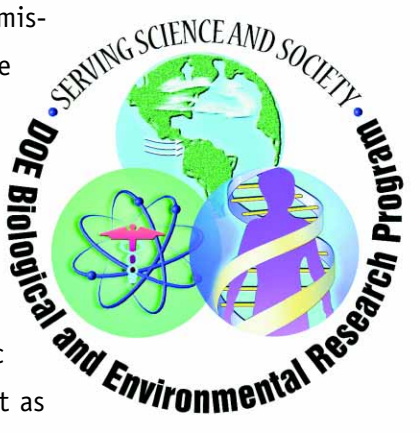
- Installed New Test Bed for Open-Source Software. A 512-CPU Linux cluster has been installed at Argonne National Laboratory's Mathematics and Computer Science Division. The cluster provides a flexible development environment for scalable open-source software in four key categories: cluster management, high-performance systems software (file systems, schedulers, and libraries), scientific visualization, and distributed computing. Its modular design makes the cluster easily reconfigurable for systems-management experiments, and its availability for testing open-source code and algorithms ensures broad use by researchers both within the laboratory and externally.



A display of NERSC-enabled research accomplishments (www.nersc.gov)

Office of Biological and Environmental Research

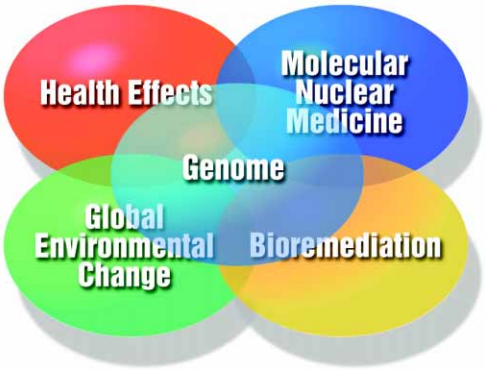
For over half a century since the establishment of the Atomic Energy Commission, the U.S. Department of Energy and its predecessor agencies have pursued biological and environmental research with an unwavering commitment to understand the health and environmental consequences of energy technologies and byproducts. To address these goals, the Office of Biological and Environmental Research (BER) relies on investigators supported in all three components of the nation's research community: multidisciplinary national laboratories, the academic community, and the private sector. Scientific diversity, always a hallmark of BER programs, has become even more important as science advances at the interfaces of such disciplines as biology and computational science.



For more details on the history of BER, see the booklet, A Vital Legacy (www.ornl.gov/hgmis/publicat/miscpubs/ober-lay.pdf).

National User Facilities

As a further service to our nation's biologists and environmental scientists, the Office of Science makes advanced instrumentation and other specialized resources available through its National User Facilities supported by the Office of Basic Energy Sciences and BER. Access to these facilities, listed below, enables the broader scientific community to increase the understanding of relationships between biological structure and function, study disease pathways, develop new pharmaceuticals, and conduct basic research in molecular biology and environmental processes.



BER programs cross traditional research boundaries to seek revolutionary solutions to energy-related biological and environmental challenges.

- **Advanced Light Source**
  www-als.lbl.gov/index.html
- **Structural Biology Center at the Advanced Photon Source**
  www.sbc.anl.gov
- **Environmental Molecular Sciences Laboratory**
  www.emsl.pnl.gov:2080
- **High Flux Isotope Reactor Facility**
  www.ornl.gov/hfir/hfirhome.html
- **Joint Genome Institute**
  www.jgi.doe.gov/tempweb
- **Los Alamos Neutron Science Center**
  lansce.lanl.gov/index_ext.htm
- **Mouse Genetics Research Facility**
  www.bio.ornl.gov/htpages/mgd/mouse_fac.htmlx
- **National Synchrotron Light Source**
  nslsweb.nsls.bnl.gov/nsls/Default.htm
- **Stanford Synchrotron Radiation Laboratory**
  www-ssrl.slac.stanford.edu/welcome.html

Appendix C
Web Sites of Research Programs and Resources Complementary to Genomes to Life

Web Sites of Research Programs and Resources Complementary to Genomes to Life

U.S. DEPARTMENT OF ENERGY

OFFICE OF SCIENCE

- www.science.doe.gov

Funding

- www.science.doe.gov/production/grants/grants.html

Office of Advanced Scientific Computing Research

- www.sc.doe.gov/production/octr

Office of Biological and Environmental Research (OBER)

- www.science.doe.gov/ober/ober_top.html

Carbon Sequestration Research

- cdiac2.esd.ornl.gov/

Environmental Molecular Science Laboratory

- www.emsl.pnl.gov:2080

Genome Programs

Human Genome Program

- www.ornl.gov/hgmis

Microbial Genome Program

- www.science.doe.gov/ober/microbial.html

Joint Genome Institute

- www.jgi.doe.gov

Microbial Cell Project*

- microbialcellproject.org

Ethical, Legal, and Social Issues

- www.ornl.gov/hgmis/elsi/elsi.html

Global Climate Change

- www.science.doe.gov/ober/esdrestopic.html

Low Dose Radiation Research Program

- www.lowdose.org

Natural and Accelerated Bioremediation Research (NABIR) Program

- www.lbl.gov/NABIR

Structural Biology Research Program

- www.science.doe.gov/ober/msd_struct_bio.html

OFFICE OF BASIC ENERGY SCIENCES (OBES)

- www.science.doe.gov/production/bes/bes.html
  Renewable energy, carbon sequestration, nanotechnology

OFFICE OF DEFENSE NUCLEAR NONPROLIFERATION

- www.nn.doe.gov
  Characterization and detection of potential biological warfare agents

OFFICE OF ENERGY EFFICIENCY AND RENEWABLE ENERGY

- www.eren.doe.gov/overview
  Renewable energy, hydrogen and ethanol production, organic acid synthesis, cellulose and lignin degradation

OFFICE OF ENVIRONMENTAL MANAGEMENT

- www.em.doe.gov
  Bioremediation research (organics)

OFFICE OF FOSSIL ENERGY

- www.fe.doe.gov
  Carbon sequestration

*Jointly funded by OBER and OBES; part of Genomes to Life program.

OTHER AGENCIES

ENVIRONMENTAL PROTECTION AGENCY

- www.epa.gov

FOOD AND DRUG ADMINISTRATION

- www.fda.gov

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

- www.nasa.gov

NATIONAL INSTITUTES OF HEALTH

- www.nih.gov

Human Genome Project: www.nhgri.nih.gov
Protein Structure Initiative:
www.nigms.nih.gov/funding/psi.html

NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION

- www.noaa.gov

NATIONAL SCIENCE FOUNDATION

- www.nsf.gov

Biocomplexity in the Environment Initiative:
www.nsf.gov/home/crssprgm/be/start.htm

U.S. DEPARTMENT OF AGRICULTURE

- www.usda.gov

PUBLICATIONS

Bringing the Genome to Life: Energy-Related Biology in the New Genomic World

- www.science.doe.gov/ober/berac/genome-to-life-rpt.html

Interagency Report on the Federal Investment in Microbial Genomics

- www.ostp.gov/html/microbial/start.htm

Microbial Genome Program Report

- www.ornl.gov/hgmis/publicat/microbial

***Science Special Human Genome Issue* (Feb. 16, 2001)**

- www.sciencemag.org/content/vol291/issue5507

***Nature* and *Nature Genetics Genome Gateway* (Feb. 15, 2001)**

- www.nature.com/genomics/human

SELECTED BIOINFORMATICS SITES

Celera Genomics

- www.celera.com

Computational Biosciences, ORNL

- compbio.ornl.gov

DNA Data Bank of Japan

- www.ddbj.nig.ac.jp

European Bioinformatics Institute

- www.ebi.ac.uk

KEGG: Kyoto Encyclopedia of Genes and Genomes

- www.genome.ad.jp/kegg

National Center for Biotechnology Information

- www.ncbi.nlm.nih.gov

Protein Data Bank

- www.rcsb.org/pdb

TIGR Microbial Web Page

- www.tigr.org/tdb

WIT at Argonne National Laboratory

- wit.mcs.anl.gov/WIT2