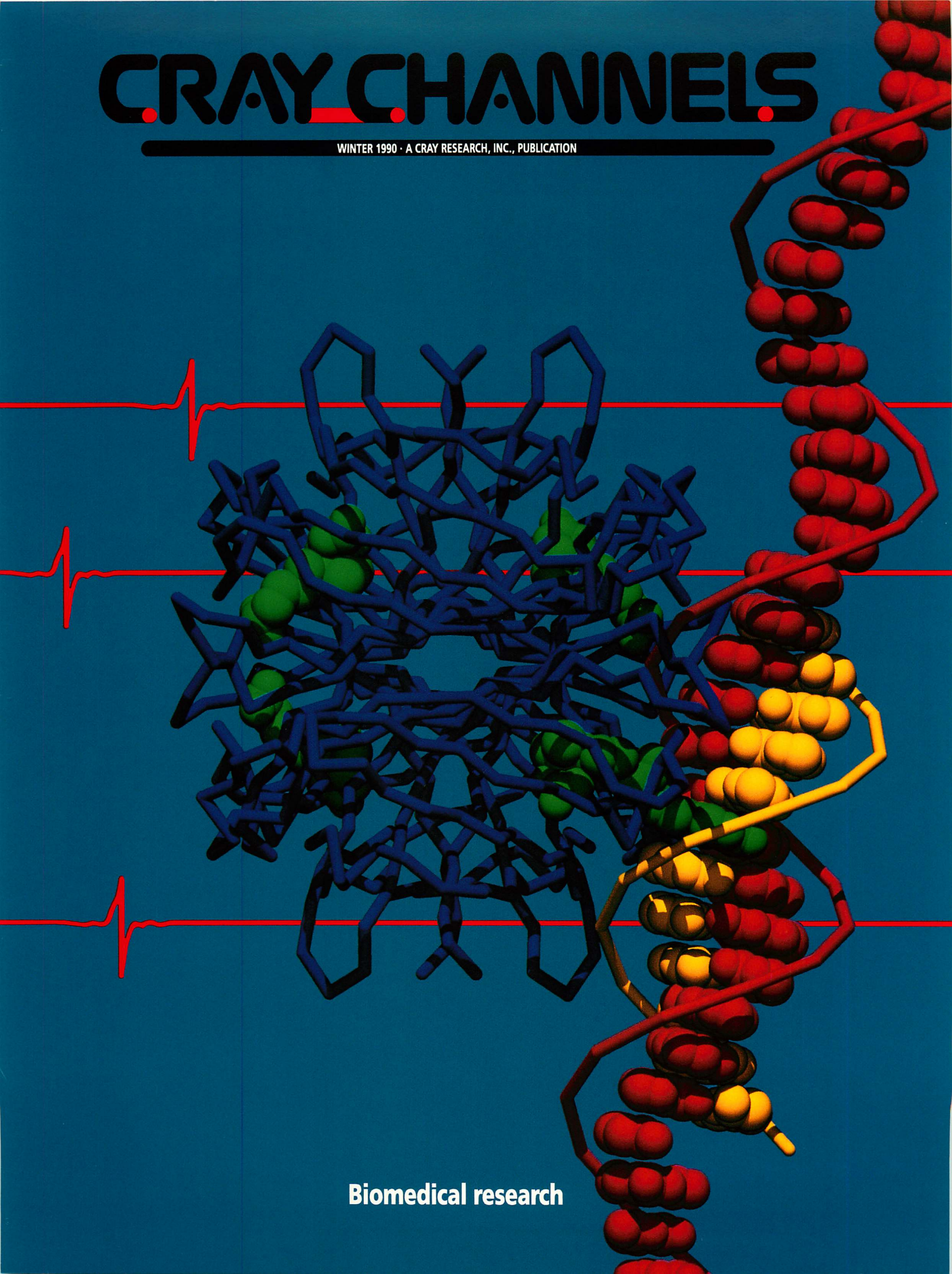
CRAY CHANNELS

WINTER 1990 · A CRAY RESEARCH, INC., PUBLICATION



Biomedical research

CRAY CHANNELS

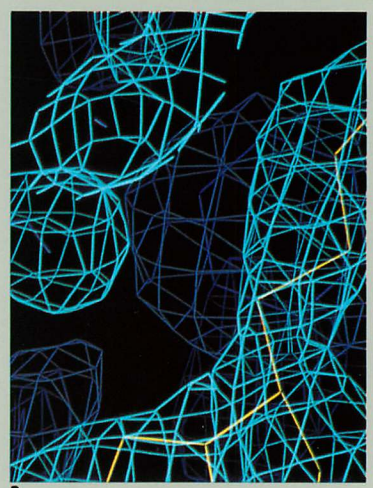
In this issue

By using supercomputers to explore the chemistry of life, biochemical researchers are opening new doors to improved health care. And researchers working in other areas of biomedicine are applying supercomputers to the design of prosthetic devices and to etiological research. In this issue of CRAY CHANNELS, we profile biomedical applications of supercomputers in commercial, university, and government research laboratories.

Researchers at the Du Pont and G. D. Searle companies are using computer modeling to study the biological activities of macromolecules. At the Research Institute of Scripps Clinic, researchers are using supercomputer technology to determine protein structures from nuclear magnetic resonance (NMR) spectroscopy. Researchers at The Ohio State University are using magnetic resonance imaging (MRI) technology to help chart the development of atherosclerosis. NASA researchers have turned their expertise in computational fluid dynamics toward improving the design of artificial hearts. Our regular departments describe enhancements to the Cray Ada Environment, along with a newly optimized convolution routine and profiles of Cray Research's Gigaflop Performance Competition winners.

As the complexity of biological systems yields to the computational power of supercomputers, medical and biochemical researchers are able to enjoy the cost- and time-saving benefits of large-scale computation. Everyone stands to benefit as this research leads to improved health care treatments and preventatives.

Features



2

Departments

CRAY CHANNELS is a quarterly publication of Cray Research, Inc., intended for users of Cray Research computer systems and others interested in the company and its products. Please mail feature story ideas, news items, and Gallery submissions to CRAY CHANNELS at Cray Research, Inc., 1333 Northland Drive, Mendota Heights, Minnesota 55120.

Volume 11, Number 4

Editorial staff
Ken Jopp, editor
Elizabeth Knoll, associate editor

Design and production
Barbara Cahlander
Eric Hanson
James Morgan
Cynthia Rykken

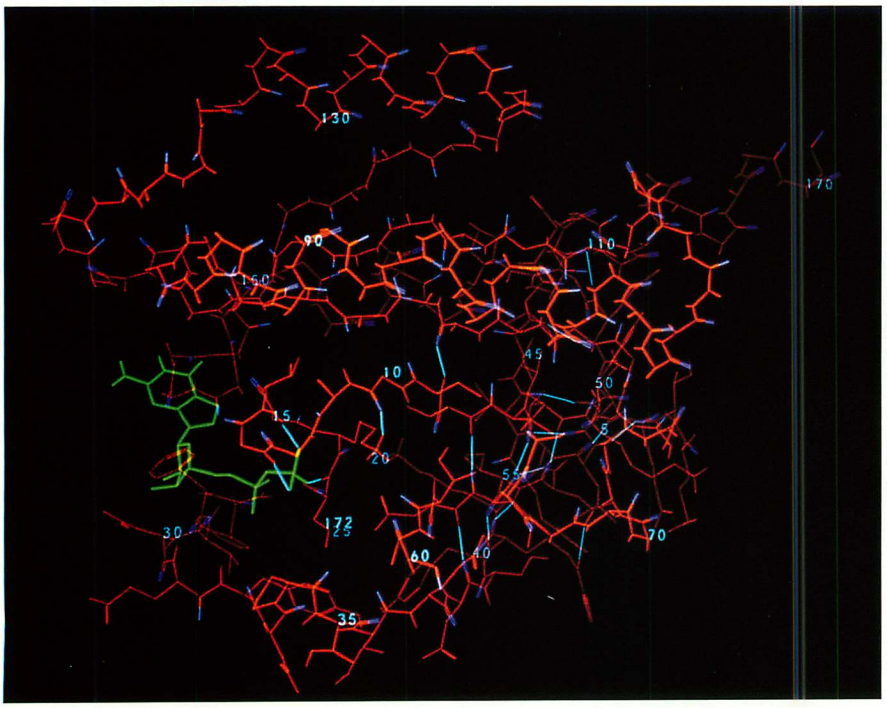


Figure 1. Model of the ras p21 protein showing guanosine diphosphate bound in the protein active site.

Computational challenges in structure-based drug design

John J. Wendoloski and F. Ray Salemme
E. I. du Pont de Nemours and Company
Wilmington, Delaware

An emerging understanding of the structural basis of biological function has made possible a new technology of rational molecular design. An area receiving intense attention is the rational design of new drugs and herbicides that regulate the activity of biological macromolecules. Typically, these molecules bind specifically to catalytic enzymes, cellular receptors, or nucleic acids that underlie physiological processes in living organisms. Effective rational design programs use a combination of modeling and intensive computational tools, but still rely critically on direct structural measurements for verification and direction. Both structural determination and rational drug design require using large amounts of supercomputer time. Much of the work described in this article was performed on the CRAY X-MP/28 computer system at the Du Pont Experimental Station.

Strategies for drug discovery

An established approach to pharmaceutical discovery uses screening methods to test a library of thousands of unrelated molecules in a biological assay system. Traditionally, these assays range from measures of whole animal behavior to assays based on microorganism survival. Lead compounds showing activity in assay screens are "analoged" by systematic chemical modification to produce new rounds of compounds. Statistical and structural comparisons of data from a family of compounds allow inferences to be made about which chemical and structural properties are important in generating the pharmacological response. Deductions based on such quantitative structure-activity relationships (QSAR) can drive synthesis of new leads or new rounds of analogs with improved properties. In a fundamental sense, these methods are aimed at generating a ghost image of the receptor site at which the drug acts, without knowing anything else about it.

Screening remains effective in drug discovery, particularly when a detailed understanding of the molecular basis of a disease allows identification and assay of a specific target molecule. At the same time, the more detailed characterization that usually follows enzyme or receptor isolation provides important information to direct the selection of candidate drugs. In cases where an enzyme is targeted for inhibition, a basic knowledge of the chemistry catalyzed can provide information sufficient to design useful compounds. These can be "transition state analogs," which bind to the enzyme active site more tightly than natural substrates and competitively inhibit activity, or "suicide substrates" that irreversibly react with the enzyme's catalytic machinery. A potential difficulty with a mechanism-based approach is that many enzymes share catalytic features, so that inhibitors generated in this way usually have to be analoged to achieve the required level of specificity. Nevertheless, this is an important approach that continues to provide new drugs and herbicides.

More recently, methods have been developed that attempt to design drug molecules from first principles, based on detailed structural knowledge of binding sites on an enzyme or cellular receptor. Although proteins are heteropolymers of amino acids whose chemical structures are inferred easily from the sequence of the encoding DNA molecule, they fold to form complex three-dimensional structures that presently are impossible to predict. However, by using x-ray crystallographic methods, the three-dimensional structure of a protein can be experimentally determined, providing a view of the enzyme or receptor binding site in atomic detail. As an example, Figure 1 shows the structure of the ras p21 protein, an enzyme whose genetically mutated form is a contributing factor in the growth of some cancers.[1] In principle, detailed structural information should be sufficient to design a drug to bind to the target molecule, but in fact this task is far from simple.

Binding sites on proteins usually form a pocket lined with functional groups that make specific

interactions with the normal substrate or effector. The objective of structure-based drug design is to generate, or alternatively, to retrieve from a structural data base, a small molecule ligand that will fit the binding site and make specific interactions with the protein. At present, no bona fide examples of the success of this approach have been reported that were not heavily based on pre-existing structural precedents. Indeed, the key to effective programs in structure-based drug design involves an integration of the methods of screening, QSAR analysis, data base searches, and computational approaches with iterative x-ray crystallographic studies that allow direct inspection and verification of the ways in which inhibitors bind to the target protein molecule.

Computational tasks in structure-based drug design

Figure 2 illustrates the information flow for a structure-based drug design program. This scheme includes inputs from empirical screens and molecular modeling studies. However, its distinctive feature is the use of x-ray crystallography to directly visualize protein-drug interactions as the definitive means of devising a rational strategy for modification. Many of the stages of the process outlined are computationally intensive. These include searches of structural data bases, molecular modeling studies to design new compounds, and particularly, the iterative computations required for crystal structure determinations of target proteins and inhibitor complexes.

Protein crystallography

A detailed description of the theory and practice of protein crystallography lies outside the scope of this article. However, the basic objective is to induce the molecules of interest to form a regular, three-dimensional lattice. The regularly arrayed molecules in a crystal lattice (Figure 3) scatter x-rays to produce a diffraction pattern that samples the spatial Fourier transform of the individual protein molecules. Structure computation basically involves mathematical inversion of the molecular transform to produce an electron density map of the protein.

Typically, x-ray data are collected as sets of multiple, two-dimensional digital arrays. Raw data for a protein structure determination easily can run

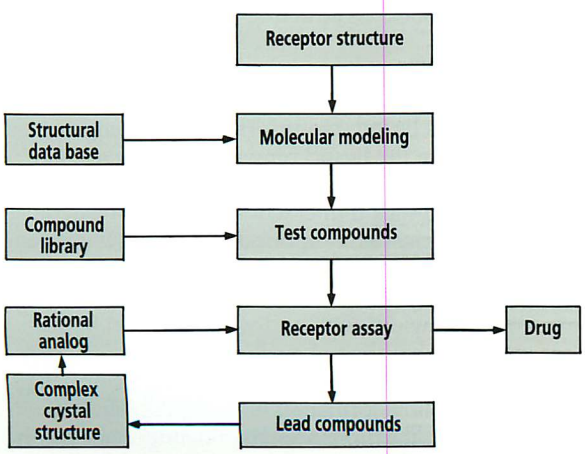


Figure 2. Information flow in a structure-based drug design program.

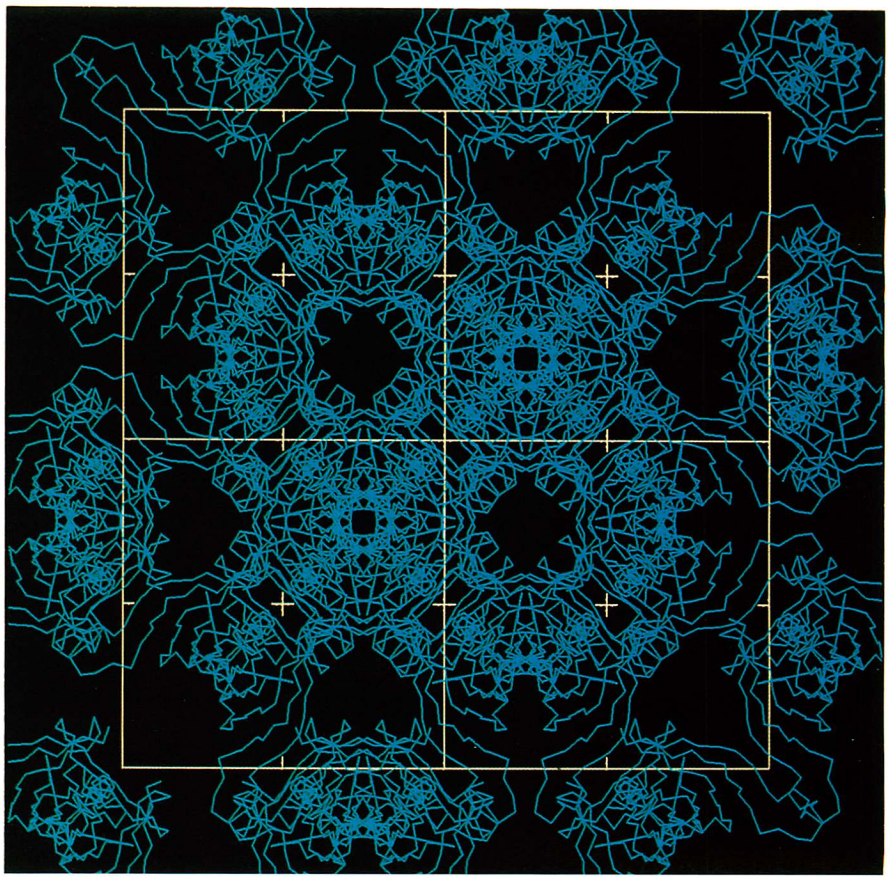


Figure 3. Molecular packing in the streptavidin crystal lattice.

to 2 or more Gbytes for a parent structure and about 500 Mbytes for each drug-complex data set. Because only the power spectrum of the spatial molecular transform can be recorded directly, the missing phase information required to reconstruct the scatterer must be recovered independently. One approach, the method of isomorphous replacement, involves collecting data from crystals of the native protein that have been doped with metals. The simple structure of the doping metal, embedded within the more complex crystal structure of the protein, can be solved by direct methods used for small molecule crystal structures. Once complete, the "heavy atom" structure provides an internal phase reference that can be used to phase the undoped protein crystal structure. Although the computations involved are substantial, they seldom constitute a rate-limiting step with modern computing equipment.

An alternative approach, useful when the structure being determined is similar to one known already but which crystallizes in a different space group, is the method of molecular replacement. Here the structure of the known molecule is used to compute a diffraction pattern that can be compared with that measured from the new protein crystal. If a correlation can be found, then phases provided from the input model can be combined with measured data to reconstruct the new structure. Finding a correlation can be an exceedingly large computational task, as the success of the method is exquisitely sensitive to case-specific details that must be investigated systematically.

Whatever the method of phase determination, inversion of the complex transform produces an electron density map of the protein crystal. The map can be interpreted using computer graphics tools that

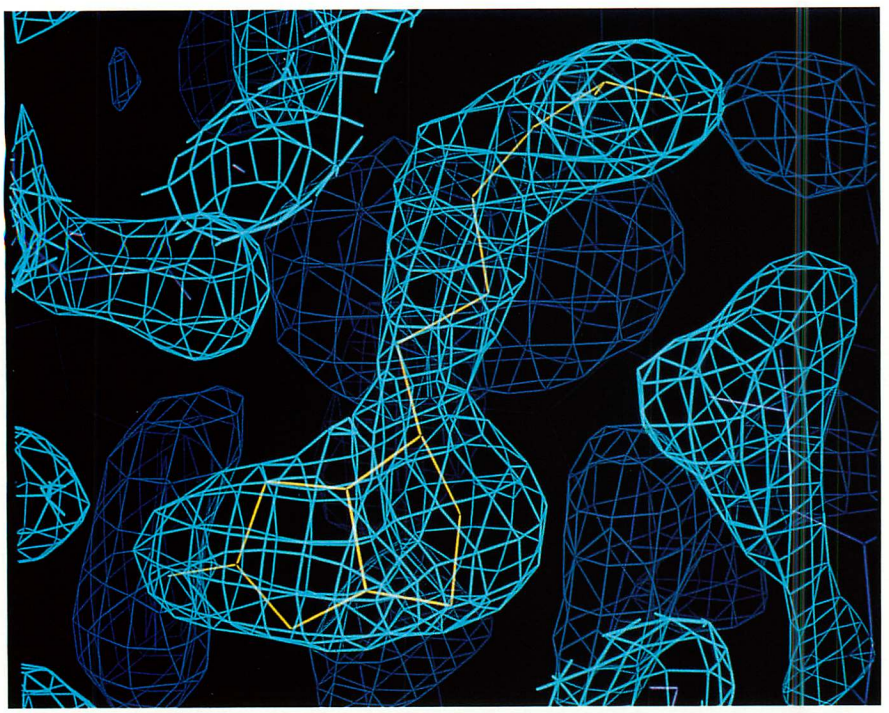


Figure 4. An electron-density map showing the biotin bound to its *binding site in streptavidin.*

make it possible to fit a stick-bond model interactively with complete conformational flexibility into the electron density (Figure 4). This initial model, which usually has only modest resolution, can be extended and improved using methods of crystallographic refinement. This process involves minimizing the differences between the observed diffraction intensities and values computed from atomic positions in the model structure. For most protein crystals, the ratio of observable data to variable parameters is not very large, so that it is useful to introduce information about the local geometrical features that are constant among protein structures. This is achieved by the introduction of energy functions or geometrical restraints, and results in refined structures that have standard geometry in their detailed features.

Protein refinement is extremely laborious and computationally intensive. This stems from the relatively poor accuracy of initial models of protein structure and the poor convergence of the refinement methods, which tend to become trapped in local minima. As a result, few proteins can be refined automatically to a correct solution. Instead, parts that are modeled incorrectly have to be readjusted manually using interactive computer graphics and electron density maps whose coefficients represent the differences between the observed data and the density predicted by the current model. Initial model building with structure fragment libraries and molecular dynamics methods of x-ray refinement are two major developments that improve this situation. In the former approach, extended fragments derived from a library of highly refined protein structures are used to assemble the new structure, many features of which may not be apparent in detail from inspection of the initial electron density map. Conceptually, this is similar to the idea of incorporating standard group geometry into restrained refinement; and, indeed, models constructed with fragments refine rapidly, owing to the elimination of most modeling errors at the outset.

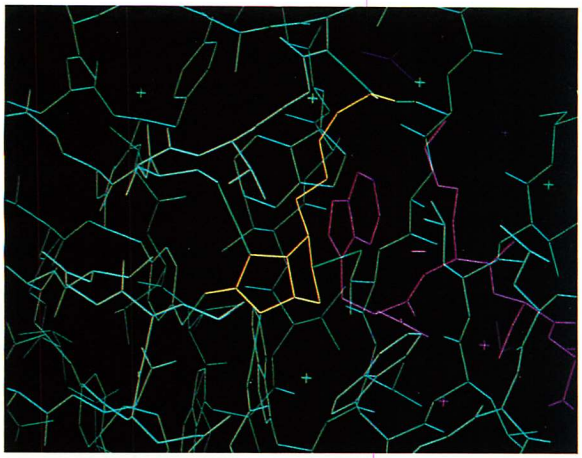
Molecular dynamics methods of protein refinement recently were described in CRAY CHANNELS.[2] The usual molecular dynamics simulations of proteins use a semiempirical potential function to describe pairwise atomic interactions and to define forces that govern the dynamics trajectory at normal temperatures. Molecular dynamics refinement basically involves the addition of a pseudopotential term, representing the difference between the Fourier coefficients observed from the crystal and those computed from the current model structure, to the usual interatomic potential. This, in effect, defines a trajectory for the system that eventually should arrive at the structure that best fits the data. The major advantage of the method is that it has a relatively large convergence radius, arising in part from its ability to break out of local minima when the simulations are run at high effective temperatures. Molecular dynamics refinement methods are very computationally intensive and have come into practical use only with the advent of supercomputers.

Despite the advantages of these advanced methods, neither is ultimately suited for automated protein refinement at the highest resolutions that most usefully reveal the details of inhibitor binding and protein-bound solvent structure. Thus, refinement remains a difficult procedure that must be performed independently for each receptor-drug complex.

Structural studies of drug binding

Determining how or where a structure will bind ligands is not normally possible from simple inspection. Moreover, even the application of the most sophisticated methods implicitly assumes that the binding site of the unliganded protein is unchanged when a substrate or inhibitor binds. In fact, this seldom seems to be the case. Fortunately, protein molecules generally retain full binding or catalytic activity in the crystalline state. This occurs because the molecules are interconnected only tenuously by weak ionic or hydrogen-bonded interactions, and otherwise are immersed in liquid water that fills continuous channels in the crystal lattice. Frequently, drugs or inhibitors simply can be diffused into the "native" protein crystal lattice and differences in x-ray scattering between the inhibited and native crystal directly transformed into a difference Fourier map that shows exactly how the drug is bound to the receptor molecule. Alternatively, it may be necessary to recrystallize the inhibitor-enzyme complex and solve its structure by molecular replacement methods.

The latter approach was used in a recent study of the tetrameric vitamin-binding protein streptavidin and its ligand, biotin.[3] This is an interesting system because the ligand binds with a dissociation constant of 10^{-14}M, making it among the strongest known protein-ligand interactions and of considerable utility in biological assay applications, where it is incorporated wherever a specific and irreversible link between biomolecules is required. From a comparison of the structures of streptavidin with and without biotin bound, it is apparent that the unusually high affinity of streptavidin reflects the participation of many factors that cooperate to allow formation of multiple hydrogen bonds between the protein and biotin. Important factors include the displacement of strongly bound water to



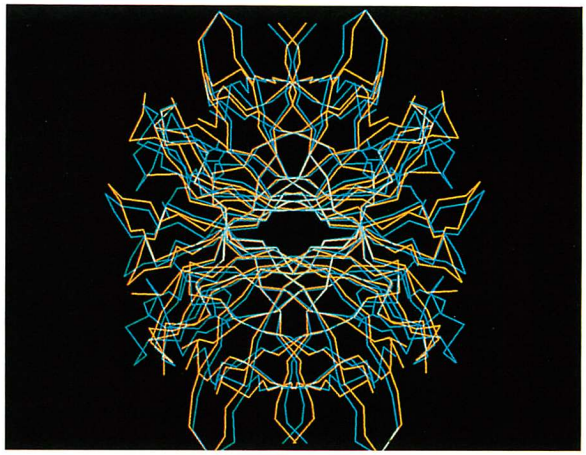


Figure 5. (Far left) The streptavidin binding site, showing the complexity of the interactions between biotin (in yellow) and surrounding residues.

Figure 6. (Left) A view showing a superposition of the streptavidin tetramer backbone with and without biotin bound.

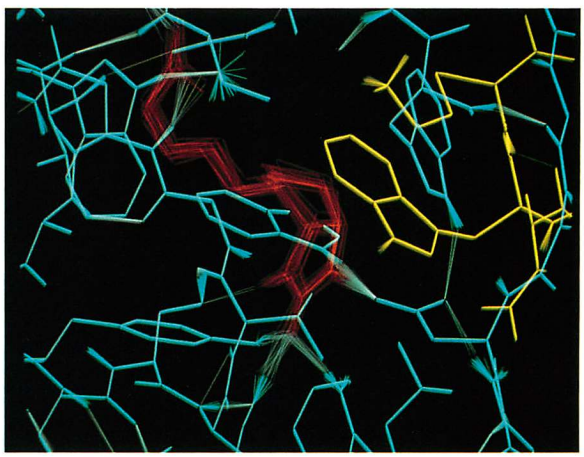


Figure 7. A simulation of biotin molecular dynamics in the surrounding environment of the streptavidin binding site.

reveal a polar binding site that binds to biotin ureido oxygen, whose oxyanionic form is, in turn, stabilized by an extended dipole array. These effects are enhanced by the ordering of a flap that sequesters the biotin from the surrounding aqueous environment (Figure 5). The whole binding process is accompanied by a substantial quaternary structural change in the tetrameric molecule (Figure 6). Measurements of the binding of biotin to streptavidin, or to a related protein avidin, suggest that the binding energy is essentially enthalpic and mainly involves the increased number of hydrogen bonds that the bound biotin makes owing to the special environment of the binding site. The important point is that the structure determination of the complex provides a detailed physical rationale of biotin binding affinity with features that can be tested explicitly by studying the binding and structures of analogs. At the same time, it is evident that virtually none of the important molecular effects or associated structural changes could be predicted accurately with available methods exclusively from a knowledge of the unliganded molecule.

Once a structural paradigm has been established for a class of inhibitors, a wide variety of modeling methods and data base search approaches can be applied to suggest rational modifications of the lead inhibitor or new compounds that might bind in analogous ways. Potential energy computations on the complex can provide detailed evaluations of the relative importance of specific interactions between the enzyme and inhibitor. In addition, free energy perturbation methods can provide estimates of binding free energies, although simpler simulations of complex molecular dynamics frequently can provide useful insight about ligand binding interactions (Figure 7).

Progress in structure-based design methods undoubtedly will accelerate as the accuracy and reliability of computational methods to assist rational design improve. In many cases, it seems clear that production of useful results will require increasing computational resources as required, for example, to simulate accurately the solvent environment around proteins. Nevertheless, it seems unlikely that purely computational methods will supplant crystallographic methods to determine protein ligand interactions in the near future. Instead, the most important computational advances in structure-based drug design are likely to be those that improve the speed and accuracy of protein structure determination.

About the authors

John J. Wendoloski is a member of the protein structure group at Du Pont, where he has been involved in applying ab initio methods and molecular modeling techniques to proteins and polymers. He received a B.S. degree in chemistry from the University of Scranton and a Ph.D. degree in chemistry from Yale University. Previously, he was on the staff of the National Resource for Computation in Chemistry.

F. Ray Salemme received a B.A. degree in molecular biophysics from Yale University in 1967. He earned a Ph.D. degree in chemistry from the University of California at San Diego in 1972 for x-ray structural studies of cytochrome c2. After 10 years at the University of Arizona in Tucson, Salemme joined Genex Corporation, where, as director of the Protein Engineering Division, he organized one of the first integrated research teams to engineer proteins using a combination of x-ray structural studies and computer-aided design methods. He joined the Du Pont Central Research Department in 1985, where he now heads a group whose objectives are the design and experimental development of engineered proteins and functional macromolecular assemblies.

References

1. de Vos, A. M., L. Tong, M. V. Milburn, P. M. Matias, J. Jancarik, S. Noguchi, S. Nishimura, K. Miura, E. Ohtsuka, and S. Kim, "Three-Dimensional Structure of an Oncogenic Protein: Catalytic Domain of Human c-H-ras p21," *Science*, Vol. 239, pp. 888-893, 1988.
2. Brünger, A. T., "Crystallographic Refinement by Simulated Annealing on Supercomputers," *CRAY CHANNELS*, Fall 1988.
3. Weber, P. C., D. H. Ohlendorf, J. J. Wendoloski, and F. R. Salemme, "Structural Origins of High-Affinity Biotin Binding to Streptavidin," *Science*, Vol. 243, pp. 85-88, 1989.