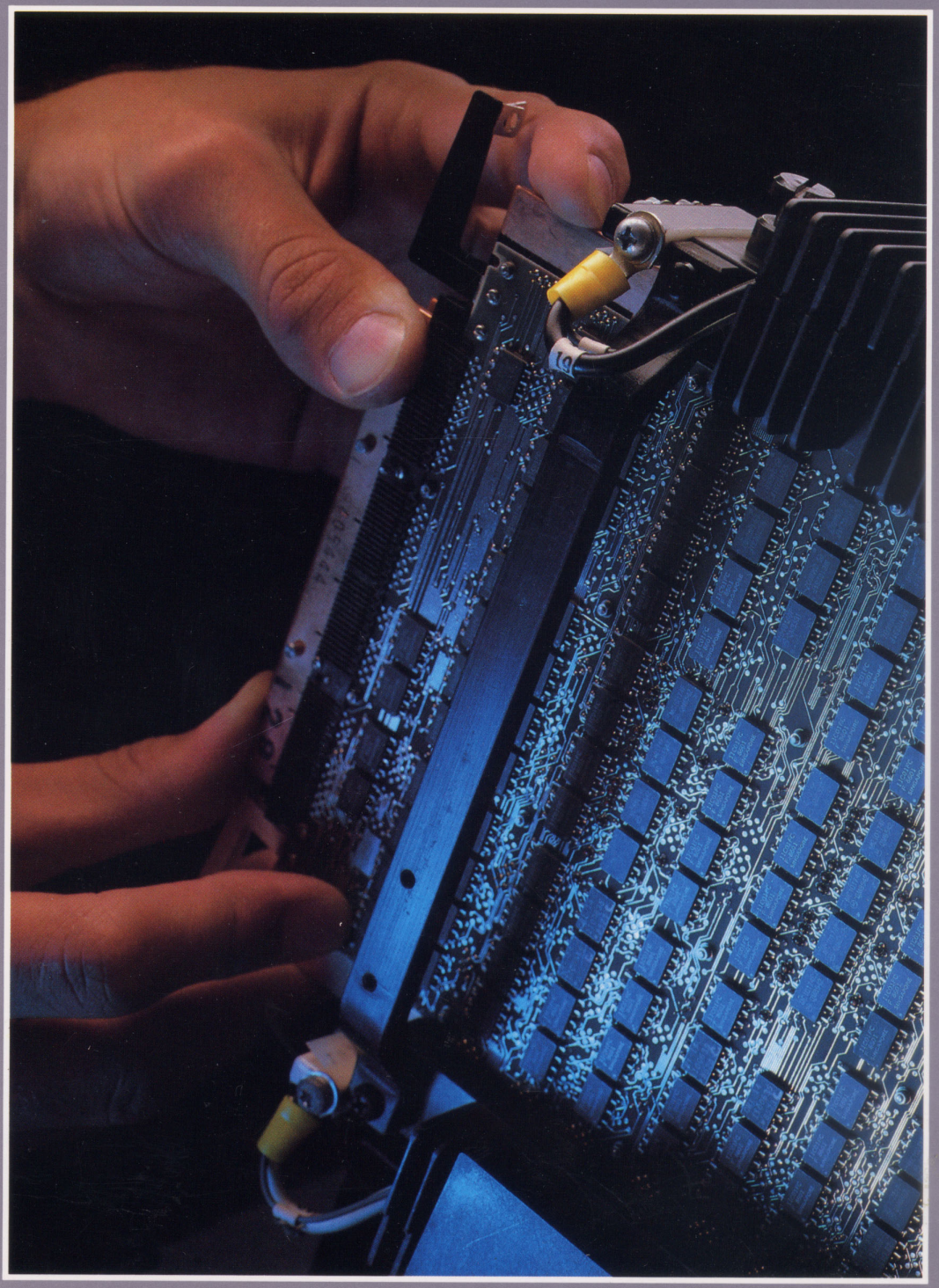
A CRAY RESEARCH, INC. PUBLICATION

CRAY CHANNELS

Fall 1987

FEATURE ARTICLES:

DEPARTMENTS:

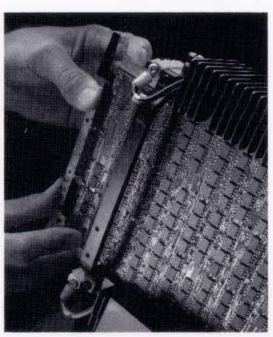


IN THIS ISSUE

Exploring the molecular worlds of chemistry and biochemistry once required laboratory experimentation using tools such as test tubes, beakers, and graduated cylinders. But as supercomputers grow in speed and power, they are becoming a beneficial and practical addition to the list of standard laboratory tools, as well as a means for reducing the time and cost involved in scientific investigation.

This issue of CRAY CHANNELS highlights the increasing use of Cray systems for molecular research. Researchers at the Naval Research Laboratory describe their methods for exploring chemical structures and Mayo Foundation scientists explain techniques for combining fluorescence methods with computational molecular dynamics simulations to model processes in macromolecules. We also look in on current efforts to optimize Monte Carlo programs for Cray systems.

As supercomputers are applied to new and diverse fields of study in a variety of operating environments, computer center managers share a need to integrate their Cray systems with other systems and with communication networks. This issue features some of the options available for achieving high-performance connectivity with Cray systems. Our regular departments showcase work at the Research Institute of Scripps Clinic, plus a new release of the Cray C compiler.

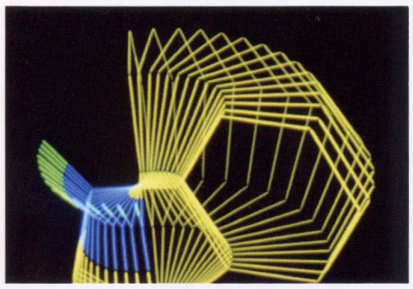


On the cover is a CRAY X-MP logic module that is being tested after assembly. After testing, the module will become an integral piece of a central processing unit in a CRAY X-MP computer system.

CRAY CHANNELS

A Cray Research, Inc. publication

Fall 1987



2

FEATURES

DEPARTMENTS

CRAY CHANNELS is a quarterly publication of Cray Research, Inc. intended for users of Cray computer systems and others interested in the company and its products. Please mail subscription requests, feature story ideas, and news items to CRAY CHANNELS at Cray Research, Inc., 608 Second Avenue South, Minneapolis, MN 55402.

Volume 9, Number 3

Editorial Staff
Tina Bonetti
Ken Jopp
Elizabeth Knoll
Carol Smith
John Swenson

Design and Production
Barbara Cahlander
Eric Hanson
Cynthia Rykken

Modeling activated processes in macromolecules

Christopher Haydock and Frank Prendergast
Mayo Foundation, Rochester, Minnesota

Molecular dynamics simulations have become important tools for biochemical research. Thanks to large-scale scientific computing systems such as Cray systems, experimental data that previously were interpreted with only heuristic and probabilistic models now are interpreted with computational models as well. This methodology is being used increasingly in clinical applications such as the design of drugs, in which the most promising candidates for synthesis must be selected from perhaps millions of possibilities.

This methodology also is being applied increasingly to fundamental studies of amino acids and proteins. The view that proteins are structures that exhibit motions occuring on a very broad time scale is now generally accepted. Evidence increasingly indicates that the functions of proteins as enzymes or drug-receptors are markedly influenced by the rates and amplitudes of these motions. Molecular dynamics simulations allow investigators to predict the dynamic behavior, or motions, of residues within proteins. The accuracy of these simulations has been confirmed by comparison of theoretical predictions with appropriate experimental data.

Thus, laboratory data have provided essential guidance in the development of molecular dynamics modeling. But the benefits have been mutual, as the successes of theoretical models have encouraged experimentalists to build more sophisticated instruments and study more complex systems.

Optical methods

Optical methods — microscopy and spectroscopy — can probe biological structure and function over an enormous range of size and complexity by exploiting the light-absorbing properties of a sample. In certain molecules, absorbed light is subsequently emitted at a longer wavelength by a process called *fluorescence emission*. This is the physical basis for fluorescence spectroscopy and microscopy. These methods are highly specific and environmentally sensitive.

Furthermore, because of the speed with which photons interact with molecules, and the time dependence of those molecules' responses to the physical effects induced in them by photons, optical spectroscopic techniques also afford measurement of dynamic processes in molecules. The time scale on which such dynamic processes can be probed is extremely broad, ranging from less than one picosecond (10^{-12} second) to greater than one second, depending on the particular optical technique used. For example, absorption of a photon by certain amino acids in proteins, most notably tyrosine and tryptophan, is demonstrated later by the fluorescence emission of a photon from the tyrosine or tryptophan. The emitted photon is of slightly lower energy than that originally absorbed, and is easily detected. The energy of the emitted photon, and the time dependence of total light emission from a large number of molecules that previously have absorbed a photon each, provide valuable clues about the structure and dynamics of the other components of the protein that surround the tryptophan or tyrosine moieties. Because the fluorescence process just described commences within a few femtoseconds (10^{-15} seconds) of the absorption of light (itself occuring in approximately 10^{-15} seconds) and continues for up to several nanoseconds (10^{-9} seconds), at least for tryptophan and tyrosine fluorescence, the dynamic interactions that influence the fluorescence can be tracked over a time range of femtoseconds to nanoseconds. Since molecular dynamics simulations currently depict mainly femtosecond and picosecond processes, it is not surprising that optical spectroscopic methods provide particularly valuable experimental data for comparison with the theoretically derived results.

Integrated approach

At the Mayo Foundation, our approach has been to use both fluorescence spectroscopy and molecular dynamics simulations to study the physics of molecular motion in relatively simple molecules. Our long-term objective is to study more complex macromolecules to learn how they work as enzymes, drug receptors or transport systems, especially to understand how molecular motions contribute to their functions. The studies of small molecules allow us to probe more precisely the determinants of the molecular events, which we hope will translate eventually into a better understanding of the more complex systems.

A full biomolecular dynamics simulation of a time interval lasting hundreds of nanoseconds is currently far beyond the computational capabilities of even the fastest Cray computer. However, infrequently occurring activated processes can be modeled readily. A variety of fluorescence

spectroscopy measurements can test theoretical models of biomolecular dynamics over a range of time scales from tens of femtoseconds to hundreds of nanoseconds. Thus, fluorescence spectroscopy can test not only conventional molecular dynamics, but also simulations of activated processes that occur as infrequently as once in a hundred nanoseconds. These activated processes are exemplified by the transitions between states with different local conformations in a simple amino acid, such as tryptophan.

In this regard, two fluorescence measurements of particular interest are the lifetime distribution and the emission anisotropy. The fluorescence lifetime is sensitive to the interactions of the tryptophan fluorophore with juxtaposing amino acid residues and solvent. Thus, each conformation of tryptophan and these neighboring groups may have a unique lifetime, typically somewhere between 0.1 and 10 nanoseconds. If these conformations all interconvert on the picosecond time scale, a mono-exponential decay of the fluorescence intensity is expected. In this case, the lifetime distribution is a single narrow peak positioned at the average value of all the lifetimes of the individual conformations. On the other hand, if a high-energy barrier separates two tryptophan conformations, for example, corresponding to the barrier that attends the ring flipping over, and these two conformations have significantly different average lifetimes, then a double exponential decay of the fluorescence intensity would be expected. The measurement and modeling of fluorescence lifetime distributions tests the computational methods for identifying significant energy barriers and simulating their crossing.

The second measurement of interest, fluorescence emission anisotropy, measures the average amount of tryptophan reorientation. Roughly speaking, the longer the delay between the time of light absorption by tryptophan and fluorescence emission, the more completely the orientation of the tryptophan will have been randomized. Accordingly, the emission anisotropy is specified as a function of this delay. Its time dependence can be measured for time delays from zero to several fluorescence lifetimes.

Theoretical considerations

Although our interest is in modeling activated processes of biological systems with thousands of atoms, the conceptual framework is supplied by the idealized model of a classical particle moving in a one-dimensional potential coupled to a heat reservoir. In the transition state theory (TST) of activated processes, the interaction with the heat reservoir serves only to establish the Boltzman equilibrium probability of finding the particle at any given height on the potential barrier. At a given temperature, heat reservoir interactions of any strength eventually will give this same equilibrium probability.

However, the strength of the interaction with the heat reservoir also has a direct effect on the rate of barrier crossing. H. A. Kramers introduced a theory of activated processes that includes these direct effects in the limiting cases of strong (overdamped) and weak (underdamped) heat reservoir interaction strength. The TST still provides a useful estimate of the rate of barrier crossing in the intermediate interaction strength regime. The parameters required by the TST or Kramers expressions for the barrier crossing rate can be estimated by the methods of molecular dynamics. The relative simplicity of these methods makes them a convenient stepping stone to the more detailed reaction pathway simulations.

A simple, yet instructive example of activated barrier crossing is provided by the free amino acid tryptophan. Upon inspection of the tryptophan chemical structure, it is apparent that the χ^1 and χ^2 sidechain torsion angles define significant stable conformations. The two torsion angles are shown in Figure 1. To model the fluorescence lifetime distribution and emission anisotropy of tryptophan, the transition rates between these stable conformations are needed. Indeed, a rotamer model has been proposed to explain the presence of multiple lifetime components in the fluorescence decay of tryptophan and its derivatives. Each lifetime component is associated with a particular stable conformation, referred to as a rotamer.

To obtain the TST or Kramers theory estimate of the transition rate, three measurements are needed: the free energy cost of moving along a suitably chosen reaction path, an effective mass, and coupling strength to the heat reservoir (viscosity). It is a reasonable guess that the rotamers will be connected by reaction paths running in either the χ^1 or χ^2 direction. To confirm this, we have calculated the potential energy surface as a function of both angles.

Computational results

If no explicit solvent is included, a high resolution surface can be calculated quickly by the method of adiabatic mapping. The adiabatic map is constructed by constraining both angles and then adjusting all remaining degrees of freedom for an energy minimum. The process is repeated

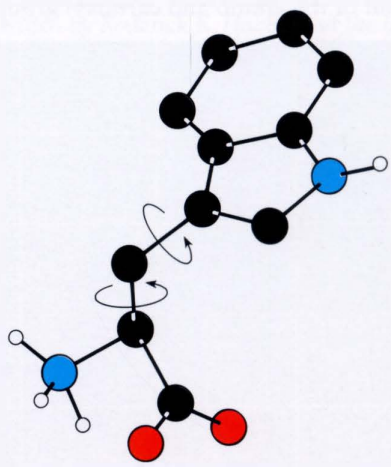


Figure 1. Pluto plot of the amino acid tryptophan. The terminal amino and carboxyl moieties at lower left polymerize the complementary terminals of other amino acids to form a protein. The sidechain indole, consisting of a six-membered ring fused to a heterocyclic five-membered ring, is oriented to the upper left. The torsion angle χ^1 is set at 180 degrees and χ^2 is set at 90 degrees. Carbon atoms are shown in black, nitrogen in blue, oxygen in red, and hydrogen in white (nonpolar hydrogen atoms are not shown).

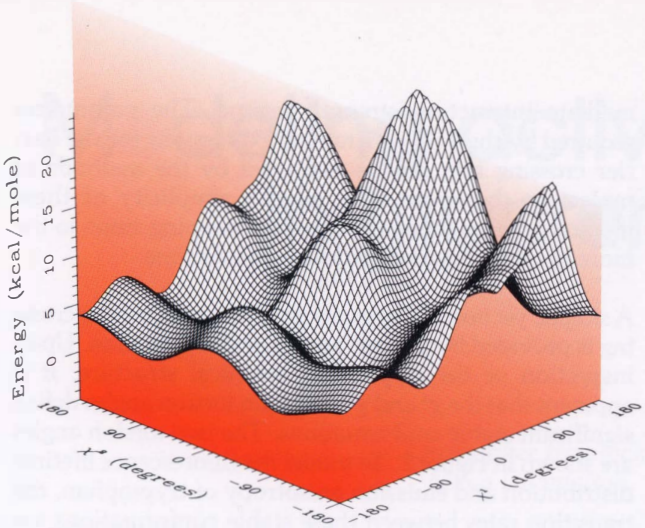


Figure 2. Adiabatic potential surface for rotational isomerization of the tryptophan indole ring. This potential map estimates the enthalpy required to move in any direction on the χ^1—χ^2 torsion plane.

for all points on a grid of selected angles. The surface formed by this set of minimized energies is shown in Figure 2.

These energies are a rough approximation of the reaction enthalpy, but are accurate enough for the present purpose. The stable rotamer conformations clearly are visible as minima in the surface. Figure 3 shows the enthalpy along the reaction path connecting the rotamer at (-60,-75) with that at (-60,+95). Note that these ordered pairs are (χ^1, χ^2). This reaction path moves nearly in the constant χ^1 = -60 degree direction. Figure 4 is a multiple exposure of 15 minimized tryptophan structures at 8 degree steps along this reaction path. The first structure, which has the plane of the rings projecting out toward the viewer, is at (-63,-96), and is thus in the first minimum in Figure 3. The fifteenth, in which the rings are seen nearly broadside, is at (-76,+16), and is at the top of the central barrier in Figure 3. The smearing out of the amino and carboxyl group positions seen in Figure 4 corresponds to small variations in χ^1 along the reaction path.

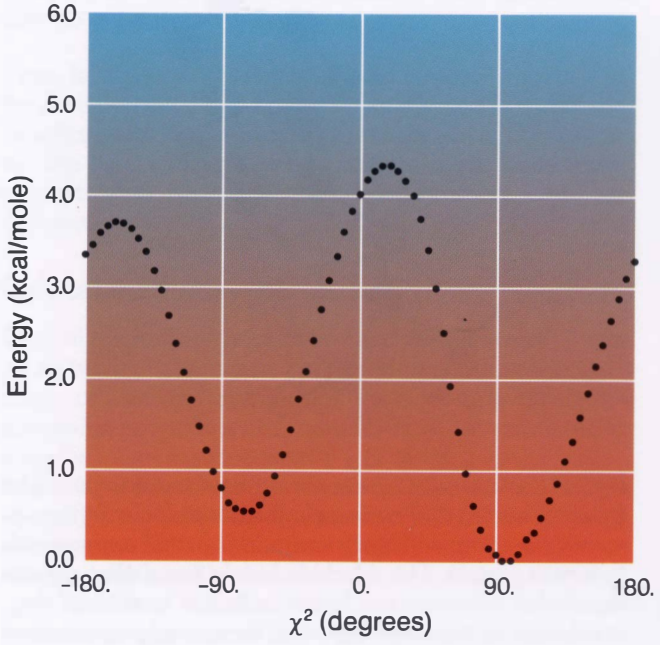


Figure 3. Adiabatic mapping of the energy required to rotate the tryptophan χ^2 torsion angle. Along this reaction path χ^1 remains nearly constant at -60 degrees.

In Figure 3, the TST estimate of the average time to cross the barrier at zero degrees is 20 picoseconds. However, extrapolation of the extensive results for tyrosine ring flipping suggests that Kramers theory in the overdamped limit is appropriate and that actual barrier crossing times are approximately one nanosecond, that is, the same time scale as the fluorescence lifetime. This in itself is an exciting result suggesting the importance of the rotamer model of tryptophan fluorescence. Extensive reaction pathway simulations with explicit solvent are needed to model accurately the transition rate between all of the tryptophan rotamers.

Finally, we present results on the transitions between local conformations of tryptophan-47 in variant-3 scorpion neurotoxin. The stable tryptophan conformations and the heights of the energy barriers between them are naturally expected to be much different than those found above for free tryptophan. As a first approximation, we computed the adiabatic potential energy as a function of the tryptophan-47 χ^1 and χ^2 torsion angles. These angles describe the rotational orientation of the indole ring, relative to the protein alpha-carbon backbone. The adiabatic mapping of the scorpion neurotoxin potential energy was calculated on a grid of 81 dihedral angles. At each grid point the χ^1 and χ^2 dihedral angles were constrained and the energy was minimized with respect to the remaining degrees of freedom by performing one picosecond of dynamics followed by 100 steps of Powell minimization.

This is a large calculation because both protein and solvent atoms must be included. By the method of stochastic boundary molecular dynamics, solvent is included, and protein atoms more than 12 angstroms distant are only implicitly included. The interactions of approximately 1000 protein and solvent atoms within a 24-angstrom

Figure 4. Multiple exposure of minimized tryptophan structures at selected points along the χ^2 reaction path in Figure 3.

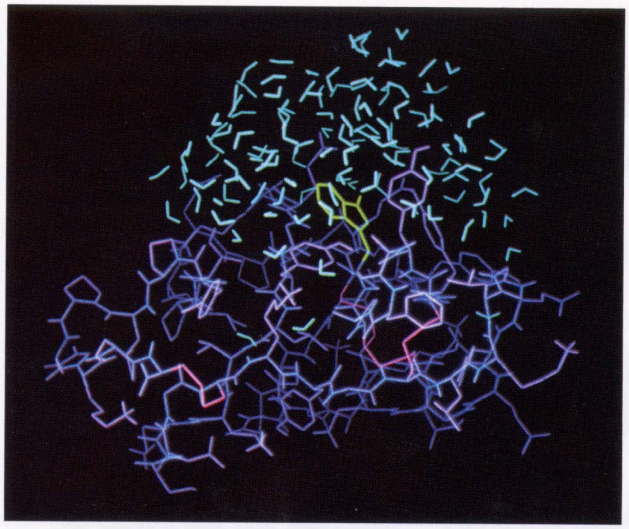


Figure 5. Variant-3 scorpion neurotoxin. The water molecules are contained within a 24-angstrom diameter stochastic boundary. The entire protein is shown. The tryptophan is located on the surface of the protein at the center of the water sphere.

diameter sphere centered on the indole are explicitly included (Figure 5). This calculation was done on the CRAY X-MP/48 computer system running the program CHARMM. The energy routines in CHARMM are fully vectorized and take advantage of the gather/scatter hardware of the Cray system. The required dynamics simulation ran about 200 times faster on one processor of a CRAY X-MP system than on the VAX 11/780. We used 10 hours of CPU time on the Cray system at Mendota Heights, Minnesota, to calculate this mapping. The resulting energy surface is shown in Figure 6. The 81 energy minima are fit with a bicubic spline for easier visualization of the surface. Note that the energy scale is an order of magnitude larger than the energy scale for the free tryptophan map in Figure 2.

This emphasizes the strong orientational constraints the protein environment places on tryptophan-47. Nevertheless, there are valleys in the map that the tryptophan orientation can explore at room temperature. Though the map is accurate enough to give only the crudest TST estimates

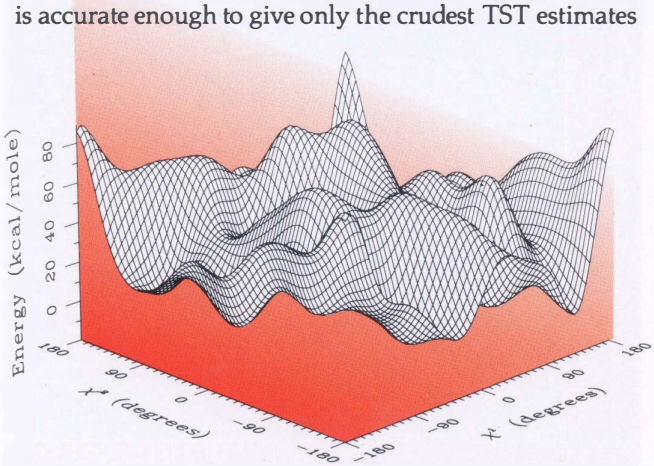


Figure 6. Adiabatic potential surface for rotational isomerization of tryptophan-47 of variant-3 scorpion neurotoxin. The surface has been fit with a bicubic spline to aid in its visualization.

of transition rates, it is quite adequate to locate the important conformational substates and the reaction pathways between them. Once these paths are known, accurate transition rates can be calculated by standard methods.

Though a single reaction pathway in a small molecular system might be modeled on a lesser machine, it takes a supercomputer to model multiple paths between multiple stable states of a protein. Such multiple paths must be considered in modeling the conformations of a fluorescence probe on the nanosecond time scale. Continued access to supercomputing systems, such as Cray systems, is allowing researchers to approach this challenging problem on a new level of molecular detail. □

About the authors

Christopher Haydock is a research associate in the Mayo Foundation department of biochemistry and molecular biology. He received his Ph.D. degree in physics from the University of Massachusetts at Amherst in 1982. His research interests are in the general area of modeling large molecular systems.

Frank Prendergast received his medical training at the University of the West Indies in 1968, and a M.A. degree in physiology from Oxford University in 1971. After a residency in medicine at the Mayo Graduate School, he received a Ph.D. degree in biochemistry from the Mayo Graduate School/University of Minnesota in 1977. He has been on the staff at Mayo since 1980 and is currently in the department of biochemistry and molecular biology.

Acknowledgments

Special thanks to the applications department at Cray Research, Inc., for computing time on the CRAY X-MP/48 system and supercomputing instruction.

All calculations reported here were done with the program CHARMM, which was developed at the Harvard University department of chemistry under the direction of Professor Martin Karplus. The molecular graphics were generated with the program HYDRA, written by Roderick E. Hubbard at the University of York, England. C.L. Brooks III of Carnegie-Mellon University has kindly shared his stochastic boundary molecular dynamics methodology and expertise.

References

A general view of protein dynamics may be found in (1). The other references provide an introduction to relevant current literature. The CHARMM progam is described in (2). Beecham and Brand succinctly review the literature on protein fluorescence in (3). A state of the art application of activated dynamics is described in (4). The stochastic boundary molecular dynamics method is described in (5).

1. Karplus, M. and J.A. McCammon, *Scientific American*, Vol 254, 1986, p. 42.

2. Brooks, B. R., et al., *Journal of Computational Chemistry*, Vol. 4, 1983, p. 187.

3. Beecham, J. M. and L. Brand, *Annual Review of Biochemistry*, Vol 54, 1985, p. 43.

4. Ghosh, I. and J.A. McCammon, *Biophysics Journal*, Vol. 51, 1987, p. 637.

5. Brooks, C. L. III, et al., *Biopolymers*, Vol. 24, 1985, p. 843.

Supercomputing in molecular structure analysis

Terry Richard Stouch, Judith L. Flippen-Anderson, and Keith B. Ward
Laboratory for the Structure of Matter, Naval Research Laboratory, Washington, D.C.

Several research projects at the Laboratory for the Structure of Matter at the Naval Research Laboratory (NRL) rely heavily on large-scale computing to investigate chemical structures. These projects include efforts to determine chemical structures by single-crystal x-ray crystallography, to investigate the energetics, structure, and dynamic properties of lipid molecules using empirically derived potential energy functions, and to investigate immunologically important regions on protein surfaces using fractal dimensions. These projects are computationally intensive and require large-scale computing resources.

X-ray crystallography

X-ray crystallography has come into increasing use in recent years due to the development of automatic diffractometers, major theoretical and analytical advances, and the increased power and availability of computers. Figure 1 is a rough plot of the number of papers that have been published in this field over the last 50 years. Prior to the days when computers were readily available to researchers, few chemical structures were determined by x-ray crystallography. In the 1950s researchers developed mathematical methods to solve the "phase problem," the integral problem in the field, and thus provided a means for effectively investigating a wide range of problems. One of the most widely used of the direct methods for solving the phase problem was developed at the NRL.[1] The "direct methods," coupled with better data-collection instruments and more powerful computers, resulted in the solution of many structures in the 1960s. Recently, large-scale computing has greatly facilitated the solution of a number of crystal structures.

The determination of a molecular structure using x-ray crystallography proceeds through several steps. First, a high-quality single crystal of the molecule of interest is

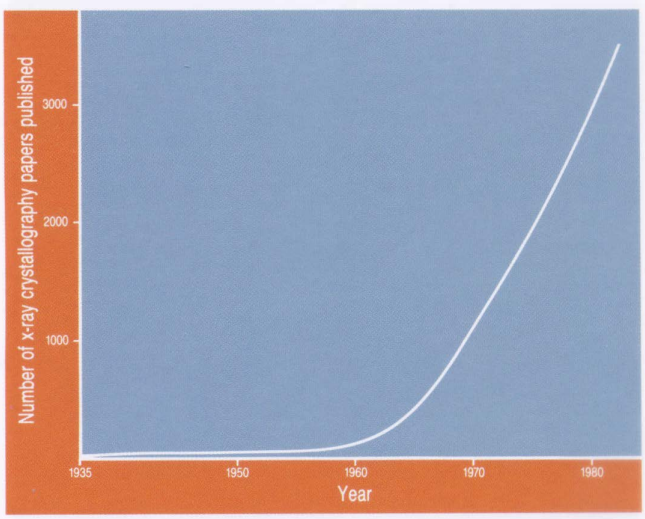


Figure 1. Number of publications reporting x-ray crystallographic structures solved plotted against year of publication.

bathed in a monochromatic beam of x-rays, which is diffracted by the crystal. Crystals are distinguished by their three-dimensional periodic structure. A single crystal is characterized by an ordered internal atomic arrangement and a definite chemical composition. The atomic spacings are of the same magnitude as the wavelength of x-rays used in diffraction experiments.

The patterns, or spectra, that result from the diffracted x-rays may be interpreted to reveal the atomic arrangements in the crystals. The intensities of the diffracted x-rays (also called reflections) comprise the raw data for the analysis. However, the relative phases of these diffracted beams, a quantity that is not obtained in ordinary diffraction experiments, also are needed. Determination of the relative phases constitutes the "phase problem."

Different approaches to this problem can be taken, based on the size of the molecule being investigated. For small and midsize molecules (less than a few hundred non-hydrogen atoms) the direct methods mentioned above usually are successful. In this procedure, phases are determined for a subset of the experimentally determined diffraction intensities. The phases then are refined and expanded to a larger set. Following this, a three-dimensional Fourier map of the electron density is calculated. If the phases are correct, the map will reveal the positions of all (or most) of the atoms in the molecule. If the phases are incorrect, the Fourier map will reveal little information. The probability that the starting phases will be correct varies with the size and complexity of the molecule and the quality of the observed data. For small, centrosymmetric structures most of the phases can be determined with a high probability and few sets of phases need be investigated. For large, non-centrosymmetric structures, a larger number of phase sets must be examined. Often one must test several thousand phase sets to find just a few that the mathematics indicate will have a high probability of being correct. While the phases for a simple structure can be determined in only a few minutes, more complex structures require considerably more time. Larger, more complicated (and often the most interesting) molecules can be very difficult to solve using any form of structure analysis. In such cases, a recourse for determining these structures is to apply large amounts of computer time to test many sets of phases. The processing speed of supercomputers makes it possible to carry out such investigations in a reasonable timeframe.

X-ray crystallography has been the definitive method for determining chemical structure for many years and is still used extensively for determining the structures of molecules that range in size from tens to thousands of atoms. The data supplied by these methods have been used in a broad range of applications that relate structure to function. An additional application has been to parameterize the empirical potential energy functions discussed below. Over the years, our laboratory has solved nearly 300 crystal structures ranging in size from small organic molecules comprising 10 to 20 atoms to proteins comprising many thousands of unique atoms. These structures include antimalarial, anticancer, and antiradiation agents; antibiotics, plant growth promoters, and others (Figures 2 and 3).

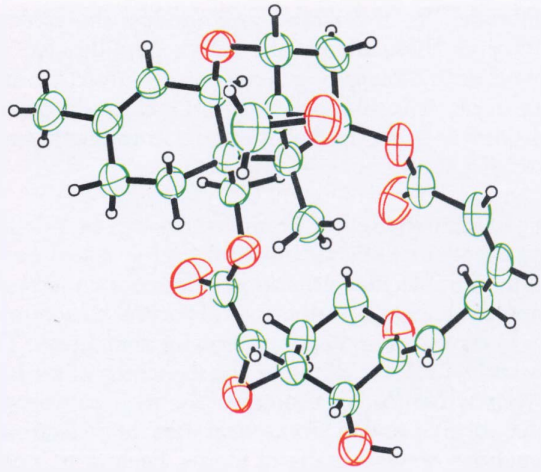


Figure 2. One representation of the structure of myrotoxin A, a chemical that is produced by a fungus and which belongs to a family of chemicals that have anticancer and antibiotic properties. The ellipsoids represent atoms. Carbon atoms are green; oxygen atoms are red; and hydrogen atoms are shown as small white circles. Chemical bonds are represented by the thick black lines and thermal vibrations of the atoms are depicted by the shape of the ellipsoids.

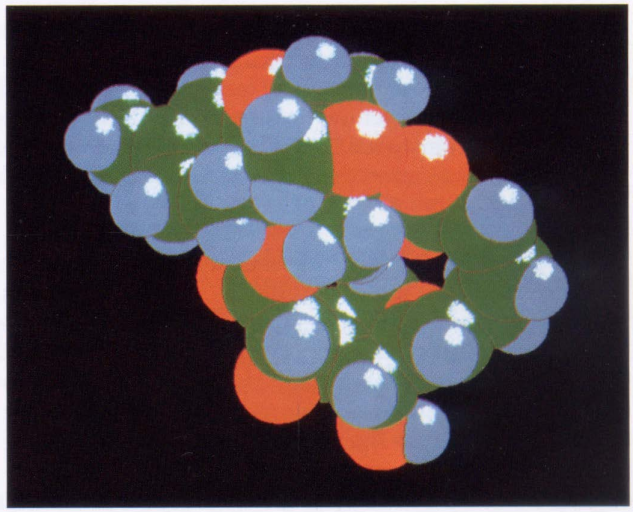


Figure 3. A different type of representation of the same view of myrotoxin A as seen in Figure 2. Here the atoms are represented more realistically as space-filling spheres. The color scheme is the same as in Figure 2.

Simulation of lipid systems

In collaboration with A.T. Hagler at the Agouron Institute, we are using computational methods to study the energetics and dynamics of lipid molecules and their crystal structures. Our overall goal is to determine the energetics and structures of membrane-active peptides in a lipid environment. Many biologically active peptides and proteins elicit their effect within the lipid-bilayer membranes of cells. The biological activity of peptides is directly related to their conformation, which is the spatial arrangement of the molecule's atoms. The conformation of a peptide is determined, in part, by its environment; a peptide in an aqueous solution usually will have a much different conformation than one in a hydrophobic solvent or lipid

membrane. To understand and modify the biological activity of these membrane-active peptides we must know their biologically active conformations. To investigate these conformations, we must include the peptides' environment — the lipid bilayer, adjacent water layer, and associated counterions — in our studies.

The potential energy of biomolecules can be calculated using a potential energy function (PEF).[2] As commonly used in biomolecular situations, PEFs contain terms that express the energy contributions of various structural features to the overall potential energy of a molecule. These features include, for example, the stretching of the bonds between two atoms, the bending of the angle between three consecutively bonded atoms, and steric and electrostatic interactions between pairs of atoms. Each term includes empirically derived parameters that determine the strength and energy of the structural feature represented by that term. For example, a bond between two atoms will have an optimum value. Any deviation from that value will require an increase in energy. A simple harmonic term that calculates the increase requires the value of the optimum length and the "force constant" for the bond stretching. These two values comprise the parameters of the bond stretching term for that particular bond.

The quality and accuracy of the results of a simulation depend completely on the terms of the PEF and the parameters used in those terms. The terms included in a PEF determine which structural features and molecular forces are included in a simulation. The parameters of each of those terms determine the value and weighting of each structural feature.

A properly formulated PEF and parameter set allows detailed studies of molecular interactions and energetics. The dynamics of a molecule also can be studied because the force on individual atoms can be directly calculated from the PEF. Using this force, Newton's equations of motion can be applied to calculate a classical dynamic trajectory of the atoms in the simulated system.

Many studies have been performed on the biomolecular simulation of proteins. Much effort has been expended on developing reliable PEFs and parameters for this class of molecules. Until recently, lipid molecules have received little attention. To study lipid-protein interactions, we must have PEFs and parameters that describe reliably the structure and energetics of lipid molecules. Parameters that currently are unknown or insufficient are being developed in our laboratories through a variety of means. For example, electrostatic parameters are being calculated with the aid of the ab initio quantum mechanical program Gaussian-82, developed in John Pople's group at Carnegie Mellon University and which we are running on the CRAY X-MP/48 computer system at the San Diego Supercomputer Center.

To test the PEFs and parameters, we are performing molecular dynamics simulations of the crystal structures of several molecules that are either lipids or are closely related to lipids. The experimentally determined crystal structures provide a measure of the quality of our PEFs and parameters. If the PEFs and parameters do not accurately reflect the interactions and forces that are active within this class of molecules, then little chance exists that the crystal structures will be accurately reproduced by our simulations. If the simulations do reproduce the crystal structures, this will indicate that the PEFs and parameters do, in fact, reflect the forces that determine the structure of these molecules.

Our studies to date indicate that currently available parameter sets are not sufficient to represent accurately all of the interactions seen in lipid molecules. As mentioned above, we currently are involved in developing new parameters to fill in these gaps. As our confidence in our PEFs and lipid parameters increases, we will study larger and more complicated systems, including large assemblies of lipid molecules (Figure 4).

The crystal simulations require considerable computational resources including CPU time, memory, and bulk storage of the results. They currently are being performed on the CRAY X-MP/24 computer system at the NRL using the DISCOVER biomolecular simulation package of Biosym Technologies, Inc. Several hours of CPU time on the Cray system are required for even short simulations. These simulations also require a large fraction of the system's four million words of memory. The demands on computational resources are high. We expect, however, that the resulting increased understanding of lipid structure and of the interactions between proteins and lipids will be worthwhile. The knowledge gained by these studies will be invaluable in the understanding of many important bi-

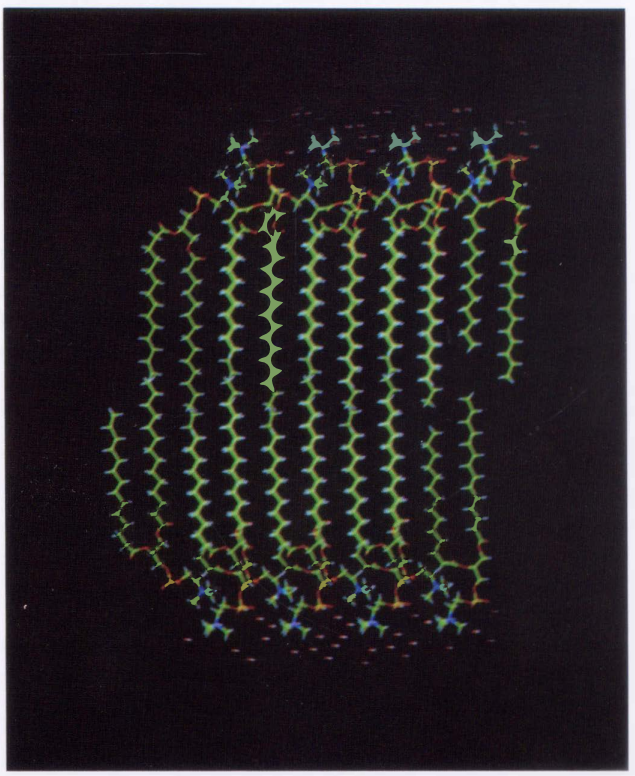


Figure 4. A small portion of a lipid bilayer with associated water layers. In actual biological systems and in our computer simulations this small segment would be replicated in two dimensions to produce an extensive two-dimensional lipid bilayer capped on either side by a layer of water.

ological processes and in the development of new and modified pharmaceutical agents.

Fractal dimension as an index of protein surface roughness

Antigenic determinants are sites on a protein or peptide surface that bind with antibodies. Identification of these sites is important to the development of synthetic vaccines. Since this identification is currently a tedious process requiring exacting laboratory work, we hope to apply computational methods to aid in this process and reduce the time required to produce synthetic vaccines.

Several properties of protein surfaces are believed to correlate with the positions of antigenic determinants. These include surface mobility, solvent-accessible surface area, and surface roughness. Currently, we are investigating the usefulness of the fractal dimension of a protein surface as a measure of surface roughness.

The solvent-accessible surface area of a protein has been found to be a natural fractal.[3,4] Because of this, the roughness of that surface can be described by the Hausdorff dimension, D, which takes on values of 0.2 for a smooth surface to 0.3 for a very rough surface.

The Hausdorff dimension of a surface can be obtained through calculations of surface area. We calculate the surface area using a modified Connolly algorithm. This algorithm, in effect, rolls a probe sphere over the protein surface and sums its contacts with the surface in order to determine the surface area. By definition, over a limited range, the area of a surface that displays a fractal nature will depend on the size of the "yardstick" that is used for the measurement. The yardstick in our calculation of protein surface area is the probe size employed in the Connolly algorithm. This probe size can be varied over a wide range. If the resulting surface area summations are plotted against the corresponding probe sizes, the slope of the plot will be zero at high and low values of the probe radius. At intermediate values, the slope of the plot will have a nonzero value equal to 2 minus the Hausdorff dimension (Figure 5).

The Hausdorff dimension can be calculated for isolated portions of the surface as well as for the entire surface. In this way, a quantitative measure of smoothness and roughness of specific regions of the surface can be calculated.

We are collaborating with researchers at the United States Army Medical Research Institute for Infectious Diseases on the development of synthetic vaccines for several proteins. Currently, researchers at the institute are applying experimental methods to determine the actual antigenic determinants of these proteins. Once those have been determined, we will correlate their positions with several calculated properties of the corresponding areas of the surfaces of these proteins, including the fractal dimension, to assess the usefulness of these properties for predicting sites of antigenicity.

Several large-scale computing systems at the NRL have been of great utility to our laboratory for the elucidation and investigation of chemical structures and properties. Computational science using large-scale systems is an invaluable tool for generating information that otherwise would not be available.□

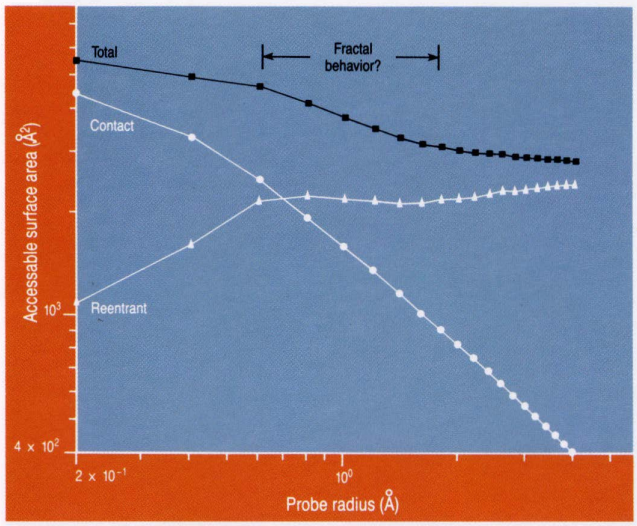


Figure 5. A plot of the surface area of the neurotoxin B protein versus the size of the probe spheres used to calculate the area. The change in the slope of the total surface area may relate to the fractal nature of the protein surface.

About the authors

Terry Richard Stouch received his B.S. degree in biochemistry and his Ph.D. degree in chemistry from Pennsylvania State University. He is an Office of Naval Technology postdoctoral fellow at the NRL. His research interests include computational investigations of the structure, energetics, and dynamics of biological molecules, quantitative structure-activity relationship (QSAR) studies of pharmaceutical agents, chemometrics, and pattern recognition and artificial intelligence applications in chemical research. His research efforts make him a heavy user of the NRL's CRAY X-MP/24 computer system.

Judith L. Flippen-Anderson came to the NRL after receiving her M.S. degree in physical chemistry from Arizona State University in 1966. She is an x-ray crystallographer specializing in the solution of small to midsized organic and biological molecules. She has solved and published over 100 structures, and has worked on all of the NRL's large-scale computing systems, including the Texas Instruments ASC, the CRAY X-MP system and Thinking Machines, Inc.'s Connection Machine.

Keith B. Ward received his B.S. degree in physics from Texas A & M University and his Ph.D. degree in biophysics from Johns Hopkins University. He presently is a research biophysicist at NRL.

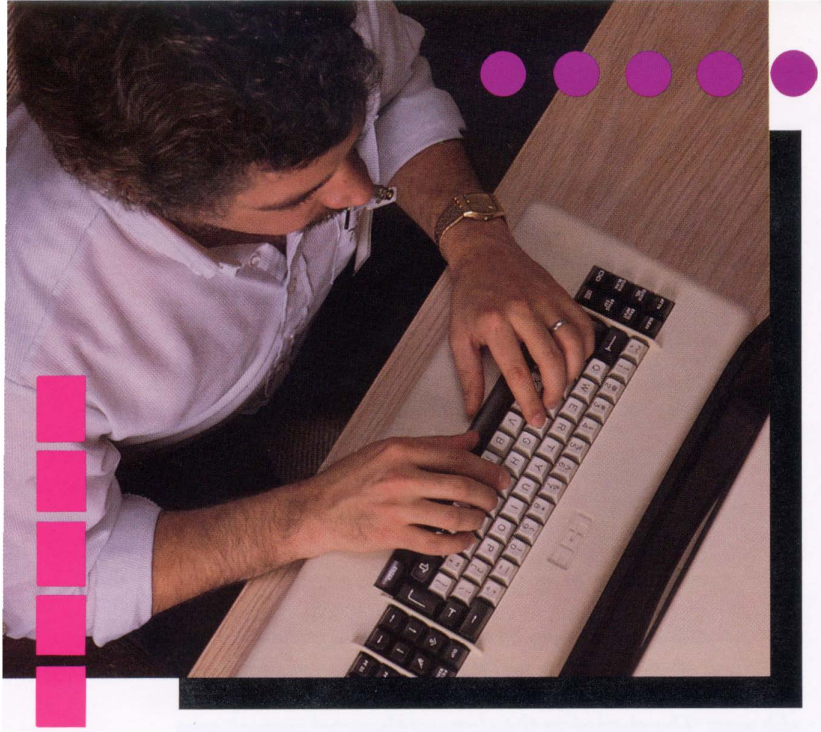
References

1. Karle, J. and I. Karle, *Acta Crystallographica*, Vol. 21, 1966, p. 849.

2. McCammon, J. Andrew and Stephen C. Harvey, *Dynamics of Proteins and Nucleic Acids*, Cambridge University Press, New York, 1987.

3. Lewis, M. and D. C. Rees, *Science*, Vol. 230, 1985, pp. 1163-65.

4. Avnir, D. and D. Farin, *Nature*, Vol. 208, 1984, pp. 261-3.



Connectivity: supercomputer networking has never been easier

Managers of computer centers that include Cray systems have many requirements in common. Among them is the need to integrate their Cray systems with other systems and communication networks in a way that delivers top performance to every level of the computing environment. At the same time, this integration must be managed to ensure that the total supercomputing solution is reliable, flexible, and easy to use. Cray Research is dedicated to providing customers with a variety of options for achieving high-performance connectivity to Cray systems.

Cray computer systems have been integrated into many types of computing environments consisting of single or multiple supercomputers, other vendors' mainframes, minicomputers, workstations, graphics devices, and high-speed data transfer facilities. This article describes Cray station software, TCP/IP, SUPERLINK/MVS, hardware options, and other networking tools and strategies available to Cray customers.

Cray station software

Cray Research's station software products allow a Cray system to appear as a natural extension of a user's familiar operating environment. Using station software and commands patterned after those of the front-end operating system, design input can be conducted interactively on a front-end system and then transferred to the Cray system for analysis. Following analysis, results can be transferred back to the front-end system and explored interactively. Users also can access the Cray operating system interactively, entering commands and receiving immediate response. In many environments, Cray station software allows users at remote sites to work as if they were connected directly to the Cray system.

Cray station software not only allows users to submit jobs to the Cray system, but also enables them to closely monitor and control running jobs. In addition, users and user jobs can transfer files and datasets easily and efficiently between the Cray computer system and front-end networks, taking advantage of data reformatting and conversion. System managers also can use a front-end system running station software to control and operate the Cray system.

Standard Cray station software is available for IBM MVS and VM, DEC VAX/VMS, Apollo AEGIS, CDC NOS and NOS/VE, Data General RDOS, AT&T UNIX System V, and the 4.2 release of the Fourth Berkeley Software Distribution. Station software for Unisys, Honeywell Bull, and Data General AOS operating systems is currently available from third-party sources.

The VAX/VMS Station provides an example of the variety of features provided by station products. More than 40 commands are available to users, including facilities to drop, kill, or rerun Cray jobs; release datasets or files held on the Cray job queue; and display information on Cray files or datasets, batch and interactive jobs, entries in the dataset staging queue, and tape devices and jobs.

Users at DECnet remote nodes also can access VAX/VMS Station facilities. In addition, users at non-DEC workstations can use an attached VAX/VMS Station as a gateway to the Cray computer system. System management tools and utilities also are available to allow a VAX system manager to alter configuration data that controls either attached or remote station operations.

Cray station software communicates with both COS and UNICOS, the two Cray operating systems. The Station Call Protocol (SCP) and the UNICOS Station Call Protocol (USCP) run on the Cray computer system to provide protocol-level communication with front-end station software.

TCP/IP

Cray Research licenses, develops, and supports the U.S. Department of Defense (DOD) Transmission Control Protocol/Interconnect Protocol (TCP/IP) suite that allows Cray systems to participate as peers in TCP/IP network environments.

The TCP/IP protocol suite, available with the UNICOS operating system, is a proven product available on most UNIX system implementations and most major computer systems. It provides file transfer applications, virtual terminal access, and networking tools upon which users can

build distributed applications. Cray Research's implementation of TCP/IP provides both local and remote peer-to-peer connectivity of Cray computer systems to other Cray systems as well as to other vendor equipment using commercially available hardware and software packages.

TCP/IP provides communication among physically diverse networks such as HYPERchannel, Ethernet, ARPANET, and many others. When a local area network (LAN) is linked to a wide area network (WAN), TCP/IP allows communication with hosts on the larger network. Regardless of whether the communicating computer systems are linked by coaxial cable, radio waves, telephone lines, or other means, TCP/IP allows users to log in remotely and transfer files across the network. Differences in text representation are automatically resolved during transmission.

TCP/IP users can log in to the UNICOS operating system interactively and use the full range of Cray software to edit program files, compile and load programs, and analyze

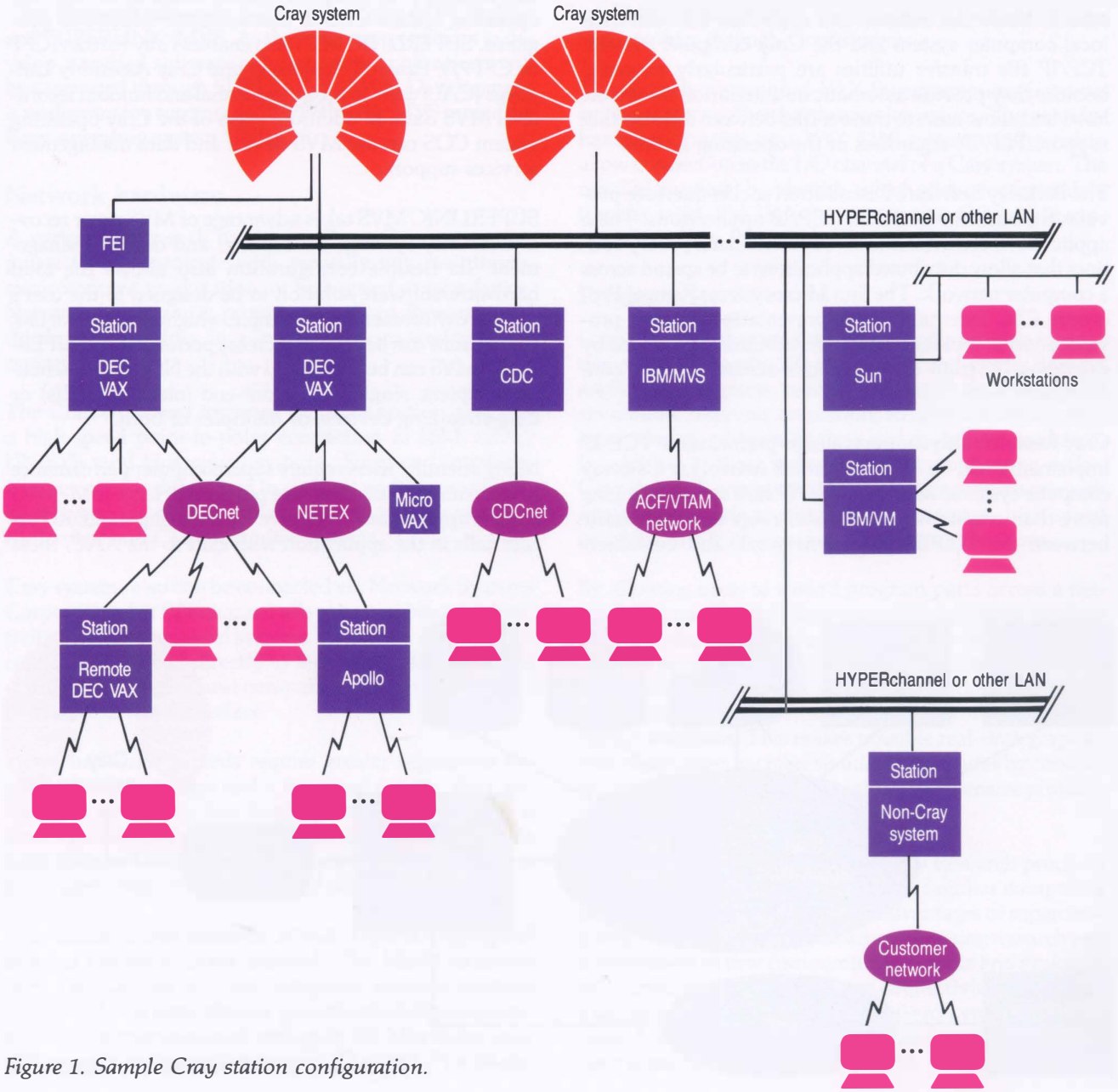
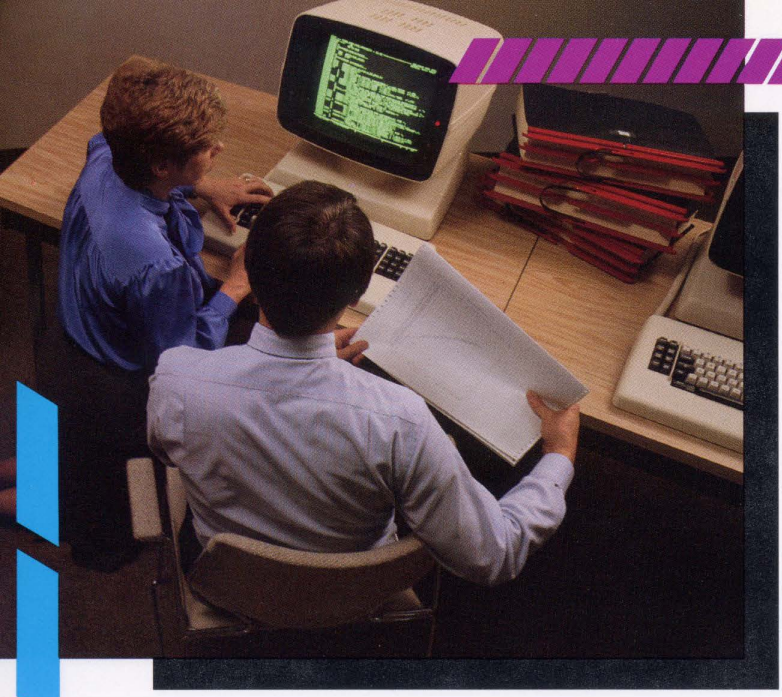


Figure 1. Sample Cray station configuration.



results. Users can transfer and copy files between their local computer system and the Cray computer system. TCP/IP file transfer utilities are particularly powerful because they provide automatic authorization on remote hosts and allow users to transfer files between all hosts that support TCP/IP, regardless of the operating system.

The Berkeley Software Distribution socket interface provides the basis for UNICOS TCP/IP applications. These applications are available through standard library routines that allow distributed applications to be spread across a computer network. The Sun Microsystems Remote Procedure Call/External Data Representation product provides another tool for building distributed applications by extending function call semantics across network connections.

Cray Research also supports a high-performance TCP/IP implementation on HYPERchannel networks. Gateway computer systems, which are TCP/IP host systems running more than one network interface, route network traffic between the HYPERchannel network and customer-supplied networks. Cray Research cooperates with several vendors to ensure the convenient and cost-effective availability of TCP/IP gateway systems. Systems offered by Digital Equipment Corporation, Sun Microsystems, Pyramid, Apollo, Amdahl, and Silicon Graphics can operate as TCP/IP gateways to Cray computer systems.

SUPERLINK/MVS

Complementing Cray Research's MVS Station is SUPERLINK/MVS, a new communication link designed to enhance resources available to users in IBM environments. SUPERLINK/MVS offers distributed applications processing through the Application-to-Application Communication (AAC) feature, which supports true process-to-process communication. SUPERLINK/MVS thus can support the development of a wide range of new applications.

Distributed processing through the AAC feature of SUPERLINK/MVS allows programs designed for the Cray operating system COS to communicate with MVS programs. SUPERLINK/MVS also enables Cray Fortran (CFT or CFT77), Pascal, C language, and Cray Assembly Language (CAL) users to access sequential and random record-level MVS data. In addition, users of the Cray operating system COS can use MVS device and data management services support.

SUPERLINK/MVS takes advantage of MVS error recovery, security systems, accounting, and dataset management. Its flexible configuration also allows the total hardware/software solution to be designed to the user's specific environment. For example, a higher number of link connections can be used to increase performance. SUPERLINK/MVS can be configured with the NSC HYPERchannel adapters, single Cray front-end interfaces (FEIs) or data-streaming devices, or multiples of both.

Many scientific users require supercomputer performance to run computation-intensive portions of large IBM/MVS Fortran applications efficiently. By replacing standard Fortran calls in the application with calls to the AAC, these

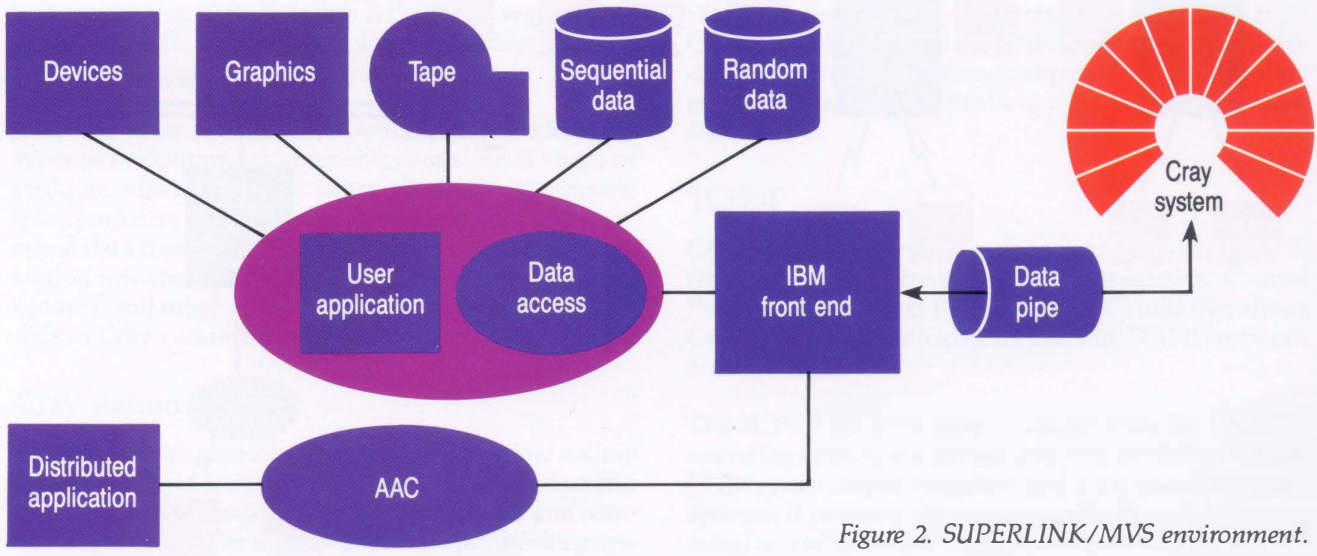


Figure 2. SUPERLINK/MVS environment.

users now can distribute program processing between the Cray and IBM computer systems. Through use of AAC, a process can create, start, and stop another process, and processes can transfer information.

Random and sequential record-level access lets COS users access MVS data one record at a time rather than an entire dataset at a time. This functionality provides a high-performance data path between the Cray computer system and IBM peripherals, which allows COS users to access MVS data locally without dataset staging, access any part of a dataset, improve throughput over Cray MVS Station transfer rates, and access datasets too large for Cray system storage.

IBM/MVS device support makes tape, disk, mass storage, and other types of MVS peripherals local to COS jobs. Thus, MVS datasets are not copied to Cray disks, and COS application programs can execute immediately without waiting for data disk transfer.

Cray Research currently provides both the MVS Station and SUPERLINK/MVS. As the company continues to develop a unified software solution, new applications will be supported through SUPERLINK/MVS. Development is underway for support of SUPERLINK/MVS under the Cray operating system UNICOS.

Network hardware

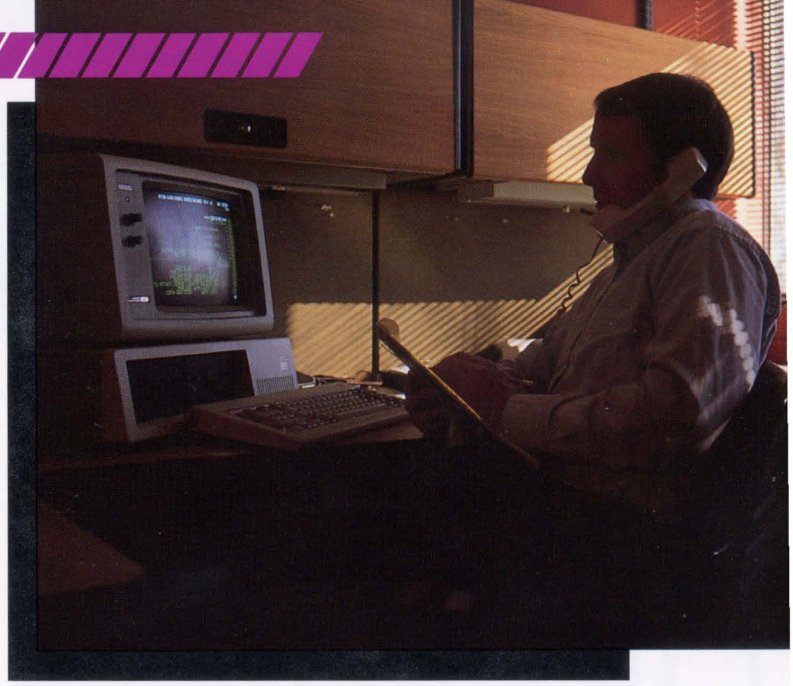
A variety of network hardware allows Cray supercomputers to communicate with virtually any mainframe, minicomputer, workstation, or high-speed storage device. Network hardware is provided by Cray Research and third-party vendors in a range of price/performance options.

The Cray front-end interface (FEI) connection provides a high-speed point-to-point connection to IBM, CDC, UNISYS, and Honeywell systems. Front-end interfaces compensate for differences in channel widths, word size, logic levels, and control protocols between other vendors' equipment and Cray computer systems.

Cray systems also can be connected via Network Systems Corporation HYPERchannel adapters, enabling connectivity to many front-end systems. Computer systems that cannot be attached directly to a channel adapter can be connected through a host computer system running more than one network interface.

When networking needs require greater separation between the Cray system and a front-end system, the Cray Research fiber-optic link is an option for connecting a front-end interface up to .621 miles (1000 meters) from the Cray system. The fiber-optic link provides complete electrical separation of the connected devices.

Cray Research also offers the HSX-1, a special high-speed external communication channel. The HSX-1 connects very fast devices to Cray computer systems running UNICOS 3.0 or later releases, providing full duplex point-to-point communication at rates up to 100 Mbytes/sec over distances up to 70 feet (22 meters). The HSX-1 facilitates



intelligent manipulation and control of high-speed data and graphics through Cray systems.

The recently announced VAX Supercomputer Gateway provides yet another hardware alternative. The gateway hardware consists of a VAX 8250 system configured to allow connection to the I/O channel of a Cray system. The new link increases bandwidth, which allows for faster data transfer rates. The VAX Supercomputer Gateway is supported by Cray Research's VAX/VMS Station.

Productivity

Cray Research communication products provide interactive access to nearly 500 third-party and public domain application programs, covering fields such as aerodynamic simulation, reservoir simulation, structural analysis, electronic design, physics, energy, and molecular modeling. In addition to these industry-specific application codes, Cray Research communication products provide network applications such as distributed processing, high-speed graphics, and high-volume data base applications.

By allowing users to spread program parts across a network, placing computation-intensive portions of applications on the Cray system, distributed processing can enhance productivity considerably. In a similar way, users can create programs that have communicating processes existing on the Cray system and on a network of engineering workstations. This makes possible real-time graphics that allows users to create simulated structures by controlling and monitoring a computationally intensive program in real time.

The connectivity provided with Cray Research products allows Cray system users to work in familiar computing environments while enjoying the advantages of supercomputing. And as technology advances, ongoing research and development of new communication media and protocols will further enhance Cray system connectivity. As a result, a broad range of users will continue to benefit from improved productivity and derive maximum benefit from their supercomputing investment. □

Vectorizing and macrotasking Monte Carlo neutral particle algorithms

Daniel B. Heifetz
Princeton Plasma Physics Laboratory
Princeton, New Jersey

This article describes programming techniques used to vectorize and macrotask Monte Carlo algorithms for simulating the transport of mass, momentum, and energy by atoms and neutral molecules in plasmas and high vacuums. These techniques also apply to Monte Carlo simulation of the transport of neutral particles such as neutrons and photons, and to Monte Carlo integrations in general, as used in computational chemistry, for example.[1]

Neutral particle transport in controlled fusion experiments

The application treated here involves computing neutral particle transport in plasmas magnetically contained in controlled fusion experiments.[2] The goal of such experiments is to create ionized gases hot enough (100 million° C) and dense enough to produce significant energy through the reaction

$$D + T \rightarrow {}^4He(3.5\ MeV) + n(14.3\ MeV).$$

Plasma confinement in controlled fusion devices is not perfect, and a steady current of ions and electrons leaves the core plasma. Ions striking the walls recombine with electrons to form atoms and molecules that re-enter the plasma. These neutral particles play an important role in fusion experiments because they are unaffected by the magnetic field that confines the plasma, and hence transport mass, momentum, and energy across magnetic flux surfaces, significantly influencing the confinement of the plasma.[3]

Hydrogenic atoms, in particular, interact with the plasma in a number of ways. They can be ionized by plasma electrons or ions, a process analogous to the absorption in neutron and photon kinetics. An atom also can transfer its one electron to an ion by passing near enough to the ion to form momentarily a quasi-molecule. This quasi-molecule immediately breaks apart, due to the high energy of the two nuclei, with the electron now possibly binding to the original ion. This is called charge transfer. The net effect is to change the velocity of the atom, which corresponds to the process of scattering in neutron kinetics.

Mathematically, the kinetics of neutral particle transport is described by a linear Boltzmann equation

$$\frac{\partial f}{\partial t} + v \cdot \nabla f = K(f), \qquad (1)$$

where $f = f(x,v,t)$ is the distribution of neutral population as a function of position x, velocity v, and time t, and where the collision operator, K, describes the absorption and scattering processes of the neutral particles in the plasma.[4]

The linear Boltzmann equation also can be put into an integral form, with a kernel $C(x,y)$, which is the probability that a neutral particle from point y will charge-transfer at point x. In a discretized solution of the Boltzmann equation, for a slab or circularly cylindrical geometry, integration using C becomes multiplication by a matrix $C = (c_{ij})$, where c_{ij} is now the probability that an atom starting in zone j charge-transfers in zone i.

Given a source, S, of atoms, the distribution of the first generation of charge transfers is given by $C * S$, the next by $C * C * S = C^2 * S$, and so on. The total charge-transfer rate is the sum over all the generations, which can be expressed as

$$(I + C + C^2 + \ldots) * S = (I - C)^{-1} * S. \qquad (2)$$

The Monte Carlo method is a good choice and in many cases the only one for solving equation (1) in three dimensions. Because the degree of confidence in the results of Monte Carlo solutions is approximately proportional to the square root of the number of test samples made, a great incentive exists for incorporating vectorization and macrotasking into Monte Carlo algorithms.

Vectorizing Monte Carlo algorithms

The basic approach to exploiting vector computers is to compute Monte Carlo test samples, or test flight trajectories, in loops over groups of test flights, keeping in arrays information about the flights such as their positions and velocities. This is possible because of the independence of the trajectory calculations. Two obstacles impair vectorizing these loops: test flight tracking algorithms contain many conditional branchings that inhibit vectorization, and, as test flights finish, holes are left in the flight information arrays that must be "filled" to achieve efficiently vectorizable loops.

The first difficulty is addressed by limiting conditional statements to as few loops as possible. In particular, numerically expensive calculations should be collected into loops free of "IF" statements so that the loops will vectorize. The gaps left by completed test flights can be accounted for in a number of ways, which are described further in the following section.

Example: the Green's function solution

The Green's function solution, equation (2) to the Boltzmann equation, has been implemented in a highly vector-

ized program. The main part of the calculation is computing the matrix C. Decomposition of LU then can be used to determine the solution $(I - C)^{-1} * S$ to equation (2). To compute C, test flights are launched in batches in each zone j, and are followed only until their first collisions with the plasma. If a collision is a charge transfer, then the test flight's weight is added to the total, c_{ij}, in the zone i where the flight collided. Thus, the Monte Carlo tracking calculation is relatively short; the infinite sum in equation (2) replaces tracking through all subsequent generations of charge-transferring atoms. A flow chart of the loops in the vectorized calculation of the Green's function matrix C is shown in Figure 1.

The computationally expensive operations in the program involving trigonometric, exponential, and logarithmic functions are contained in vectorized loops. Two loops, the repacking loop discussed below, and a loop that calculates neutral particle reflection, are not vectorized. This last loop involves sampling from multi-dimensional distributions.

The speedup, σ, from vectorization can be estimated using a version of Amdahl's relation:

$$\sigma = \frac{1}{(1-\alpha) + \alpha/w} .$$

Here α is the fraction of the calculation done in vectorized loops, and w is the averaged speedup of those loops over their scalar versions. The individual vectorized loops in the Green's function calculation showed speedup factors of 4 to 11 on a CRAY X-MP/22 computer system. The speedup averaged over all the vectorized loops for a typical calculation using 20 cylindrical zones was about $w = 5.8$. Since approximately 78 percent of the computation of C took place in these loops, the overall speedup during this calculation was a factor of $\sigma = 2.8$. The fraction $\alpha = 0.78$ could be increased, for example, by using a similar reflection model. Amdahl's relation shows that α is already large enough so that a small increase in α will give a large increase in the speedup, σ.

Accounting for gaps

As the calculations of test flights finish, for greatest efficiency they should be deleted from the arrays containing test flight information. In the present implementations of the Green's function calculation, the flight information arrays are repacked "back to front" as gaps appear. A search for gaps begins at the head of the array, and when the gap is found, it is filled with data on the last active flight in the loop. This minimizes the amount of shuffling required to repack.

A second approach to accounting for gaps is to use a flag array: LFLIGHT(I) = 1 if flight I is active, and 0 if not. Loops of the following form then can be vectorized under CFT 1.14:

```
      DO 50 I=1,K
      IF (LFLIGHT(I) EQ.1) THEN
      ...
      ENDIF
50    CONTINUE
```

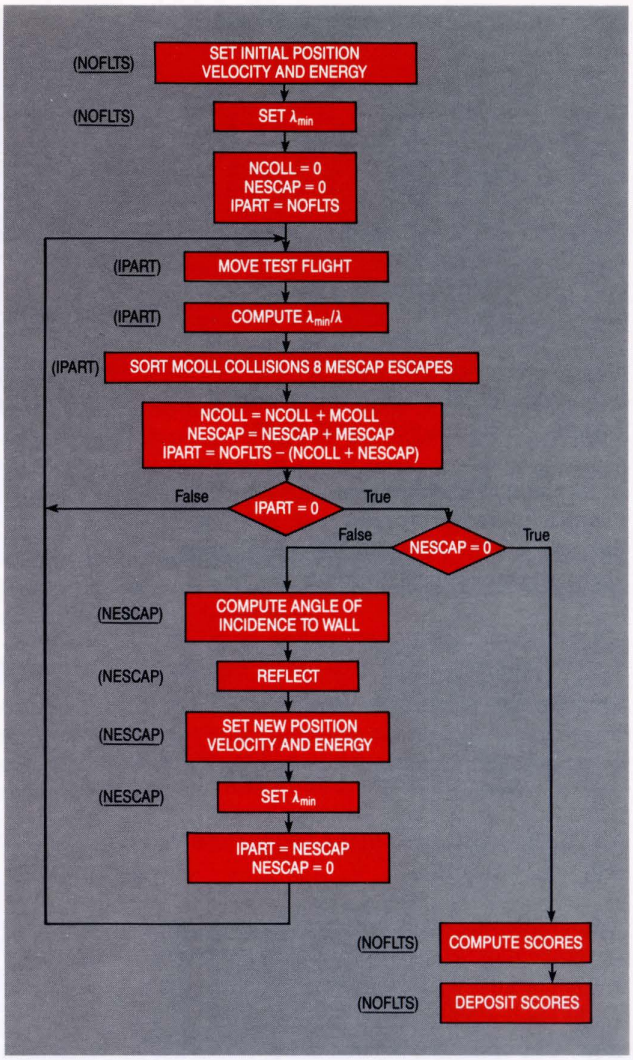


Figure 1. A flow chart of the loops in the vectorized calculation of the Green's function matrix C. NOFLTS is the number of test flights, NCOLL and NESCAP are the number of test flights that have ionized and escaped respectively, IPART is the number of active flights, and λ_{min} and λ are the distances that the test flights move. The sizes of the loops are indicated in parentheses, and are underlined if the loop is vectorized.

Macrotasking Monte Carlo algorithms

Many Monte Carlo algorithms have been vectorized. However, introducing vectorization into an already existing program typically requires extensive rewriting. Vector hardware as implemented in general purpose computers does not seem to be "natural" for Monte Carlo algorithms because of their conditional statements. It would be better to make parallel calculations at the subroutine level, that is, to macrotask, with different tasks tracking independent test flights.

Example: implementing macrotasking of the DEGAS code

Asynchronous macrotasking, as implemented on the CRAY X-MP and CRAY-2 computer systems,[5] has been incorporated in the DEGAS neutral particle transport code.[6] The DEGAS code computes a three-dimensional solution of the linear Boltzmann equation (1) for the transport of neutral particles in plasmas and high vacuums.

```
      PARAMETER (NPTASKS=4)
      DIMENSION ITCA(3,NPTASKS),NSEED(NPTASKS)
      EXTERNAL PROFILE
            ...
      CALL LOCKASGN(LOCKVAR,ISTAT)
      DO 10 I = 1, NOTASKS
      ITCA(1,I)=3
10    CALL TSKSTART(ITCA(1,I),PROFILE,NSEED(I),LOCKVAR)
```

Figure 2. Code used to start tasks.

It has 55,000 Fortran lines, and would be difficult to substantially vectorize. However, macrotasking was implemented in DEGAS with the addition or modification of only approximately 20 lines.

To compute a profile consisting of NOFLTS test flights, NOTASKS tasks are used, each consisting of a partial profile of NOFLTS/NOTASKS flights:

```
      SUBROUTINE PROFILE(KSEED,LOCKVAR)
      TASK COMMON/COMTRACK/
            ...
      DO 10 J=1,NOFLTS/NOTASKS
10    CALL TRACK
```

The arguments to PROFILE are the random number seed KSEED, and a lock identifier LOCKVAR. The TASK COMMON/COMTRACK/ contains the arrays of test flight tracking information, and their Monte Carlo scorings. The tasks are started using TSKSTART,[7] as shown in Figure 2.

The array ITCA is the task control array. Note that the random number seed NSEED is passed to the task through an array, so that each task will use a different address to retrieve that task's random number seed.

The routine LOCKASGN is used to assign a lock identification for the single lock used in the algorithm. The lock is used in the tasks at the routine DEPOSIT, where the results of the tasks, stored separately in TASK COMMON/COMTRACK/, are added to the COMMON totals over all tasks:

```
      SUBROUTINE PROFILE(KSEED,LOCKVAR)
            ...
      CALL LOCKON(LOCKVAR)
      CALL DEPOSIT
      CALL LOCKOFF(LOCKVAR)
```

This is necessary to avoid two tasks simultaneously modifying the same COMMON variable.

The macrotasking is completed with a loop that waits for all the tasks to finish before ending the calculation

```
      DO 30 I=1,NOTASKS
30    CALL TSKWAIT(ITCA(1,I))
```

Performance

The macrotasking DEGAS program has been run extensively on the four-processor CRAY-2 system at the National Magnetic Fusion Energy Computer Center in Livermore, California, using the Cray Time Sharing System (CTSS). One measure of macrotasking is processor overlap, defined as the sum of the individual processors' CPU times divided by the time during which at least one processor was in use. The maximum theoretical processor overlap during the flight tracking profile of the DEGAS calculation, where over 95 percent of its computing time is spent, approaches the number of processors used. This overlap is actually less, since tasks vary in length due to the statistical nature of the calculation, and because of intrinsic limitations in the system, such as memory bank delays.

The CTSS scheduling algorithm favors macrotasking programs, sending related tasks to the top of the processor queues when any one task gains a processor. A range of overlaps from 1.40 to 3.75 has been observed using the CRAY-2 system during the flight tracking section of the calculation. The maximum overlap of 3.75 was achieved during a test with no other programs running. The processor overlap in DEGAS varies with the number of CTSS jobs running simultaneously, and the sizes of those jobs. The lowest overlaps were observed when DEGAS was competing with a number of other large programs that probably were not macrotasking. This behavior depends critically on system-defined scheduling parameters.

A more practical measure of macrotasking is the ratio of the wall-clock time for the calculation while unitasking over the macrotasking wall-clock time. This ratio is generally bigger than the processor overlap. For example, the same problem was run twice in similar system conditions, using one and four tasks. The processor overlap for the macrotasking job was 2.0, but by wall-clock time, the unitasking job took over 3.5 times as long as the macrotasking job.

The enhanced wall-clock speedup occurs because even though the tasks may not receive processors simultaneously, they may receive processors before the lead task exits its processor (Figure 3). This determines the wall-clock speedup time. The wall-clock performance appears not to degrade as quickly as processor overlap with the increase in competition for processors. This again depends heavily on the system's program scheduling algorithm.

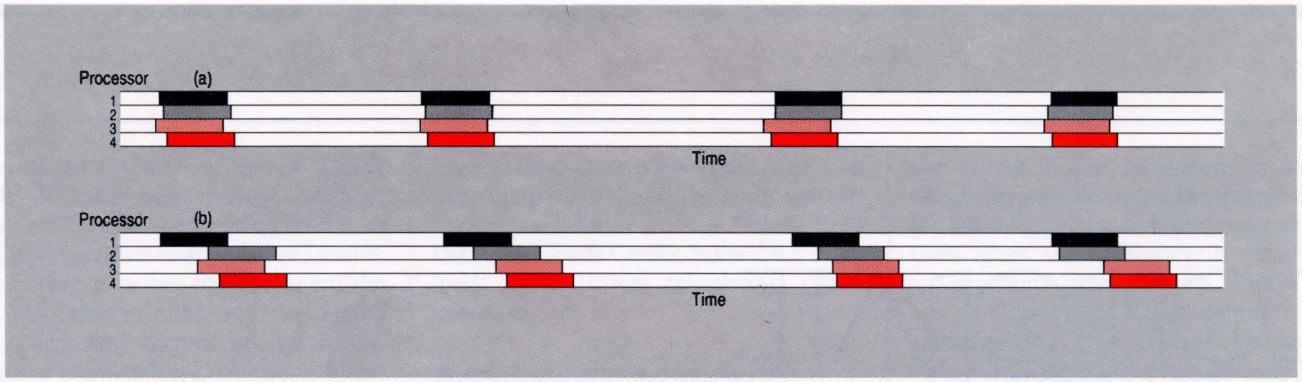


Figure 3. Two possible sequencings of four tasks from a single program, both showing good performance in reducing the wall-clock computation time, but with (a) good and (b) bad processor overlaps. Even though the tasks in example (b) do not receive processors simultaneously, all of them may get a processor before the lead task exits its processor. Thus, both cases will finish in roughly the same wall-clock time.

Program development

While the modification required to introduce macrotasking in the DEGAS program was minimal, a few time-consuming bookkeeping chores were needed to complete the implementation. For example, a careful inspection of the Fortran coding was made to verify that the task routines were truly re-entrant. While extensive program editing was necessary, no essential changes in the algorithm were needed. This is in contrast to the modifications typically needed to vectorize an existing scalar Monte Carlo algorithm, in which the extensive changes necessary in the logical structure can introduce programming errors.

An unsolved difficulty is that each TASK COMMON generated by a task adds 10 to 15 percent to the size of the executable code. The reason is that DEGAS collects scorings in the TASK COMMON/COMTRACK/ for all the elements in its geometric grid, because it is necessary to know the entire neutral particle population $f(x,v)$. This is unlike a typical calculation of neutron transport, in which only a small number of quantities are calculated in a single computation. While tolerable with four tasks, this growth may become prohibitively expensive if larger numbers of tasks are used. An alternative, less expensive storage approach, using TASK COMMON "buffers," is currently being developed.

Note also that the macrotasking algorithm can be extended to include a recursive generation of tasks; new tasks can be spawned as test flights "split" in the course of their flights. Such a scheme should be very attractive on future machines with large numbers of processors.

The problem of reproducibility

The CFT random number generator produces a single sequence of pseudo-random numbers. Separate tasks sample asynchronously from this sequence, and different calculations from the same input will agree only statistically. A systematic approach to achieving exact reproducibility between calculations is to use trees of pseudo-random numbers, called Lehmer trees.[8] Two linear congruent pseudo-random sequences are used to generate left and right successors to an element X in the tree

$$L(X) = a_L X + c_L \ (Mod\ m) \text{ and } R(X) = a_R X + c_R \ (Mod\ m),$$

where the constants a_L, a_R, c_L, c_R, and m are chosen to ensure that the tree's branches of pseudo-random numbers are reasonably independent.

The present DEGAS algorithm uses an original seed together with three left successors to generate separate sequences of right successors for each task. Using more complicated task structures will exercise the full power of the Lehmer tree.□

About the author

Daniel Heifetz received his Ph.D. in mathematics from Columbia University in 1982, where he did research in the representation theory of Lie groups. He is currently on the research staff at the Princeton Plasma Physics Laboratory. His present interests are in plasma-material interactions and in the interaction of computer hardware with the design of numerical algorithms.

References

1. Alder, Ceperley, Pollock, "Quantum Mechanical Simulation of Liquids," *Accounts of Chemical Research*, Vol. 18, 1985, p. 286.

2. Bromberg, J.L., *Fusion: Science, Politics, and the Invention of a New Energy Source*, MIT Press, Cambridge, 1982.

3. Heifetz, D., "Neutral Particle Transport," *Physics of Plasma-Wall Interactions in Controlled Fusion*, D. Post and R. Behrisch, Eds., Plenum, New York, 1986.

4. Tendler, M. and D. Heifetz, "Neutral Particle Kinetics in Fusion Devices," *Fusion Technology*, Vol. 11, No. 2, 1987, pp. 289-310.

5. Nelson, L. and P. Rigsbee, "Multitasking at Cray," *CRAY CHANNELS*, Summer 1985.

6. Heifetz, D., et al., "A Monte Carlo Model of Neutral Particle Transport in Diverter Devices," *Journal of Computational Physics*, Vol. 46, 1982, p. 309.

7. *Multitasking User Guide*, Cray Research Publication SN-0222.

8. Fredrickson, P., et al., "Pseudo Random Trees in Monte Carlo," *Parallel Computing* Vol. 1, 1984, p. 175, and Bowman, K.O. and M.T. Robinson, "Studies of Random Number Generators for Parallel Processing," *Proceedings: Second Conference on Hypercube Multiprocessing*, SIAM, 1987.

CORPORATE REGISTER

Cray gains customers worldwide

Cray Research announced in July that the DuPont Company ordered a CRAY X-MP/24 computer system with SSD Solid-state Storage Device to replace its CRAY-1/A system. DuPont will use the system to explore research areas including molecular modeling, life sciences, electronics, catalysis, polymer sciences, and chemical process simulation. "We are looking to the Cray system for insights in simulation that will provide research breakthroughs," said John Taylor, manager of DuPont's scientific computing division.

The greater memory and performance capabilities of the CRAY X-MP/24 system will enable DuPont researchers to solve more complex problems and model larger molecules, processes, and reactions, according to Taylor. The new system will be installed in the fourth quarter of 1987 at DuPont's Experimental Station near Wilmington, Delaware, where 40 percent of their research is done. DuPont currently ranks first in research and development expenditures in the chemical industry and ranks sixth among U.S. manufacturers in research spending.

Cray Research announced in August that German automobile manufacturer Daimler-Benz AG installed a CRAY X-MP/24 computer system with SSD Solid-state Storage Device. The system was installed in the second quarter of 1987 at Daimler's research facility in Stuttgart, West Germany. The system will be used for vehicular research, development, and design, primarily in the areas of structural analysis, combustion, aerodynamics, crash simulation, and acoustics. This system is the tenth Cray supercomputer installed in West Germany and the seventh Cray installation in the automotive industry.

The Scripps Clinic and Research Foundation in La Jolla, California, ordered a CRAY X-MP/14se computer system to be used for drug design and to decipher the shapes and dynamics of peptides, proteins, and other biological materials. "The addition of the Cray supercomputer represents a dramatic increase in our capabilities," said Dr. Richard Lerner, director of the institute. "Now, the rate at which our investigations progress will not be defined by the limitations of our machinery. We will have a computer that can keep pace with us."

Scripps Clinic researchers also plan to develop application software for Cray Research as part of an agreement between the two organizations. Scripps Clinic, founded in 1924, is one of the largest private, nonprofit, biomedical research institutions in the country. Cray Research announced the order in September.

Cray Research announced in June that the U.S Army Strategic Defense Command purchased a CRAY X-MP/14 computer system. The computer was installed in the second quarter of 1987 at the U.S. Army Strategic Defense Command Simulation Center located in Huntsville, Alabama. A large base of scientists and engineers will use the system to research and develop technology for strategic defense applications.

Boeing Computer Services, acting as prime contractor for the state of Alabama, ordered a CRAY X-MP/24 system with SSD Solid-state Storage Device, Cray Research announced in June. The Alabama Supercomputer Center in Huntsville will host the Cray system, which will be installed in the fourth quarter of 1987. Alabama industry, government agencies, and state universities will use the system for scientific and engineering research. High-speed remote access to the system will be provided by the Alabama Supercomputer Network, which operates statewide.

The Unisys Corporation ordered a CRAY-2 supercomputer, acting as prime contractor for the United States Air Force Weapons Laboratory. The system was delivered in the third quarter of 1987 to Kirtland Air Force Base, Albuquerque, New Mexico. The laboratory will use the CRAY-2 system to support Air Force applications, including the Strategic Defense Initiative. A sophisticated communications network will link the system to other Air Force units. Cray Research announced the order in June.

Volkswagen A.G. ordered a CRAY X-MP/14 system with SSD Solid-state Storage Device, Cray Research announced in July. The system, which is the 11th Cray supercomputer in West Germany, will be installed at Volkswagen's research and development center in Wolfsburg, West Germany. Volkswagen AG will use the Cray system for vehicle research, development, and design. Specific applications will include structural analysis, combustion modeling, computational aerodynamics, and crash and acoustic simulation.

Also in July, Cray Research announced that the Department of Energy's Idaho

Operations Office ordered a CRAY X-MP/24 computer system to support the scientific and engineering efforts of the Department of Energy. EG&G Idaho, Inc., acting as prime contractor for the U.S. Department of Energy, will install the computer system at the Idaho National Engineering Laboratory, Idaho Falls, in the third quarter of 1987.

Leading Edge Technologies, Ltd. of Australia ordered a CRAY-1/M computer system, which will be used by petroleum, aerospace, automotive, biotechnology, and other high technology companies. Government, education, and research organizations throughout the country will also use the system. The system will be installed in Melbourne during the first quarter of 1988, pending export license approval. Melbourne was chosen as the computer site because it represents the heaviest concentration of companies and industries in Australia with large-scale computing needs. Cray Research announced the order in August.

The University of Kiel in West Germany ordered a CRAY X-MP/18 computer system, Cray Research announced in September. The system will be used for basic research, particularly in astrophysics and oceanography. The system will be installed in the fourth quarter of 1987 at the university's computer center, pending export license approval.

Cray Asia/Pacific opens Singapore office

Cray Asia/Pacific, Inc., the Pacific Rim business center of Cray Research, has opened an office in Singapore to serve the subsidiary's southern region. The office is managed by Henry Lau Li Hien, district sales manager of Cray Asia/Pacific, Inc., and is located at #30-30 Raffles City Tower, 250 N. Bridge Road, Singapore 0617; telephone: (65) 3365151.

Cray opens New York City office

Cray Research has opened a new sales office in New York City's World Trade Center to serve the investment banking, insurance, and securities industries. The office is part of the company's eastern region and is located on the 79th floor of the World Trade Center, Suite 7967, New York, N.Y. 10048; telephone: (212) 466-6546.

Release 3.0 of Cray C compiler now available

Cray Research has improved the Cray C compiler with release 3.0 for CRAY X-MP and CRAY-1 computer systems running COS 1.16 or UNICOS 3.0. Based on the Portable C Compiler from AT&T, release 3.0 of the Cray C compiler features

- Direct generation of relocatable object code, thus bypassing the need for an assembler. For some codes this feature has reduced compile time by 50 percent.
- An optional instruction scheduler. When selected, this feature can improve execution time 10 to 20 percent.
- Further improvements in character code generation, in-line functions, and loop vectorization. The Cray C compiler now vectorizes loops with pointers.
- An interface to standard Cray tools for symbolic debugging

The Cray C compiler supports a large standard library of functions and an ever-expanding base of software application programs. Additionally, C programs can include calls to CAL and Fortran routines to maximize performance.

The C preprocessor, *cpp*, is included as part of the Cray C compiler. *Cpp* allows macro substitution, conditional compilation, and the inclusion of named files in the compilation process.

Use of the Cray C compiler requires licensing. Contact the nearest Cray Research sales office for additional information.

Cray ships 200th system

On September 25, Cray Research shipped its 200th supercomputer — Serial 426. To recognize the shipment of the CRAY X-MP/24 system, a small ceremony was attended by several people from Cray Research and representatives from the customer, a classified government organization. Cray Research shipped its 100th system in the winter of 1985, after 13 years of business. And after only another two and one-half years, the company has doubled that number.

Cray discontinues development project

In September, Cray Research discontinued the computer system development project headed by Steve Chen, a senior vice president. As a result, Chen resigned from Cray Research to reconstitute the project on his own.

"The project grew significantly beyond our original vision, both in terms of technological risk and budget, and we believe it no longer met the objectives or style of Cray Research," Chairman John A. Rollwagen said. "We are refocusing on the fundamentals that have made Cray Research an industry leader," he added.

Chen's project, internally called the MP project, was established late in 1985, and focused on both hardware and software design of a high-level parallel processing architecture. Over the past two years, the group had moved into large-scale research in a number of areas.

Ewald named vice president, software development

Cray Research also announced that Robert H. Ewald, formerly vice president of commercial marketing, has been named vice president for software development. In his new post, Ewald will report to Lester Davis, executive vice president, development. Ewald joined Cray Research in 1984 after seven years in computing and communications at the Los Alamos National Laboratory, Los Alamos, New Mexico. "We think he is the best person to bring us up to a new level of software support with new products and new capabilities for Cray," Chairman John A. Rollwagen said.

APPLICATIONS IN DEPTH

CHARMm, QUANTA simulate molecular behavior

CHARMm software models dynamic behavior and other characteristics of molecular systems. Available for use with CRAY X-MP, CRAY-1, and CRAY-2 computer systems, CHARMm uses empirical energy functions and performs a variety of computations to develop molecular structures. The program was written and developed at Harvard University, and is now maintained, enhanced, and supported by the Polygen Corporation. Polygen's QUANTA graphics display program runs CHARMm on a local graphics workstation and allows chemists to view three-dimensional models of structure and properties generated by CHARMm.

With its analysis capabilities, model-building, energetics, and mechanics features, CHARMm enables researchers to simulate realistically the behavior of molecules. It can be used to calculate a wide range of molecular properties, from simple peptide conformations to dynamic atomic motions of multi-component protein units. CHARMm manipulates a variety of systems, from a single molecule to a full crystal composed of a protein and several hundred solvent molecules.

CHARMm's model-building features enable users to construct molecules using either polar atom or full atom representations. They can be treated as isolated molecules, molecules in solvents, or molecules in repeating unit cells of crystalline solids.

CHARMm's energetics and mechanics features allow users to perform four major operations involving the energy of a system: minimization, conformational searching (or energy mapping), molecular dynamics, and normal mode analysis. The program's empirical energy function is based on separable internal coordinates and pairwise nonbonded interaction terms. The energy function is a summation of terms for bond and bond angle potential, torsion potential, improper torsion terms, van der Waals interactions, electrostatics, hydrogen bonding, and potentials such as harmonic atom constraints and dihedral constraints. All energy calculations can be performed on a constrained or an unconstrained system.

The minimization techniques used by CHARMm adjust the system coordinates to lower the energy, which in turn locates minimum-energy configurations. CHARMm can determine global minimum-energy configurations for small systems or local minimum-energy configurations for large systems.

Conformational search procedures systematically vary specified internal coordinates — typically dihedral angles — to produce energy curves and surfaces. The energy may be minimized at each point on the curve by constraining the internal coordinate varied.

CHARMm performs molecular dynamics simulations of molecules in vacuum, solvent, or crystalline environments. It carries out dynamics simulations of molecules on full or limited molecular systems in which uninteresting degrees of freedom are eliminated from the energy calculations.

CHARMm also has a data analysis facility that collects data and generates tables to efficiently store the variety of data generated by CHARMm. Data are manipulated and analyzed to furnish information about potentially active structures. CHARMm compares and reports the structural, dynamic, and energetic properties of molecules.

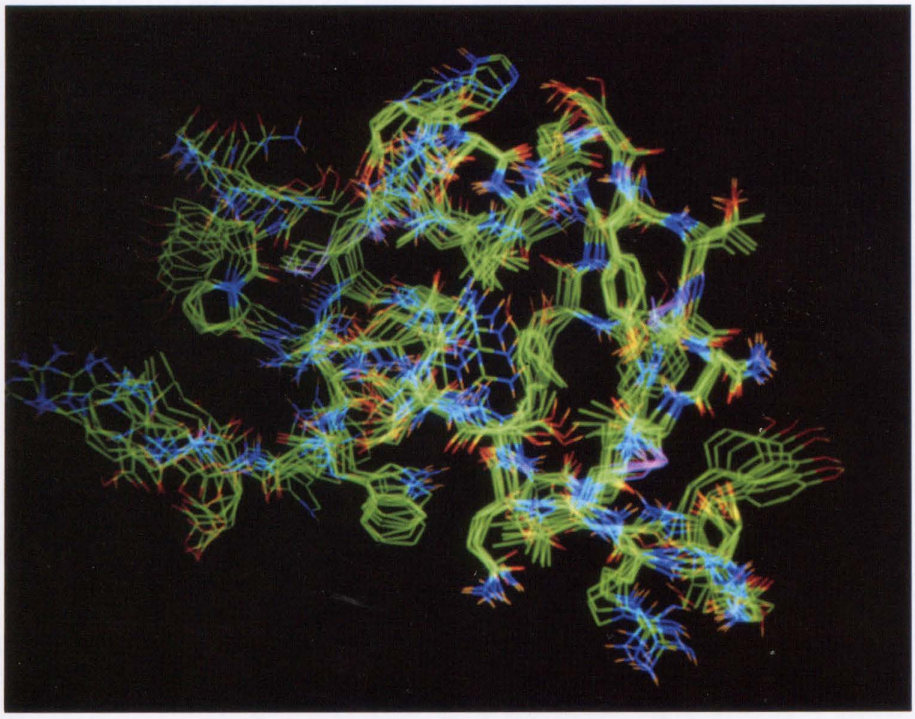
QUANTA, a graphics display tool for molecular modeling, enables chemists to view three-dimensional models of structures generated by CHARMm software. For rapid display and analysis of structures, CHARMm and QUANTA communicate by means of shared data files.

QUANTA is designed to link local graphics workstations with a central supercomputer running CHARMm. QUANTA formulates input for CHARMm, sends information to control the supercomputer simulation, and displays the results.

Features of QUANTA include

- ☐ Three-dimensional display and real-time manipulation of molecules
- ☐ Animated display of minimization and dynamics calculations
- ☐ Flexible selection and coloring of atoms
- ☐ Color highlighting of atomic or user-defined properties
- ☐ Docking of molecules with real-time distance and energy calculations
- ☐ Comparison of molecules using statistical methods
- ☐ Optional display of hydrogen bonds and molecular surfaces
- ☐ Protein engineering: construction and modification of polypeptides

For more information about the use of CHARMm and QUANTA on Cray computer systems, contact Bruce Gelin at Polygen Corporation, 200 Fifth Avenue, Waltham, MA 02254; telephone: (617) 789-2888.



Epidermal growth factor simulated by CHARMm dynamics.

AMBER optimized for Cray systems

AMBER version 3.0 is a macromolecular simulation package highly optimized for use with Cray systems. AMBER is used to study the structure, dynamics, and interaction energies of biological macromolecules and their interactions with small molecules.

Researchers in the chemical, pharmaceutical, and biotechnology fields apply AMBER to problems such as examining the factors responsible for enzyme catalysis and modeling the effects of anti-tumor drugs on the structure of DNA. The program also has been used to simulate the effects of amino acid substitutions in genetically engineered proteins.

Macromolecular simulation requires a model of the system, usually based on structural information obtained experimentally. Software within AMBER allows users to build realistic models that include water and ions present in real systems. The fundamental interactions among atoms in the system are described in AMBER by analytical equations that yield energy as a function of the positions and nature of the atoms.

Separate modules within AMBER perform four main actions on the model system. Energy minimization results in a stable, low-energy structure. Molecular dynamics simulates the thermal motions of the atoms at a given temperature. Normal modes analysis determines the fundamental vibrational modes of the system. Finally, free energy perturbation calculates realistic free energy differences due to small changes in the system, such as the mutation of an amino acid in an enzyme.

AMBER 3.0 was written by U.C. Singh, P.K. Weiner, J.W. Caldwell, and P.A. Kollman. It is furnished under license, and is available to nonprofit institutions for a nominal fee. The program is dis-

tributed in a Fortran source form, as well as VAX/VMS executables. The computationally intensive modules of AMBER have been used under the Cray operating system COS and the Cray Time Sharing System (CTSS).

For more information on using AMBER with Cray computer systems, contact Peter Kollman, Department of Pharmaceutical Chemistry, University of California, San Francisco, CA 94143; telephone: (415) 476-4637.

DISCOVER solves molecular design problems

Cray Research is working with BIOSYM Technologies, Inc., of San Diego, to implement their DISCOVER software program on Cray systems. The program is expected to run on a single processor of a Cray system at approximately 150 times the execution speed of a VAX 11/780. As a result, scientists will be able to use DISCOVER on very large problems with greater efficiency.

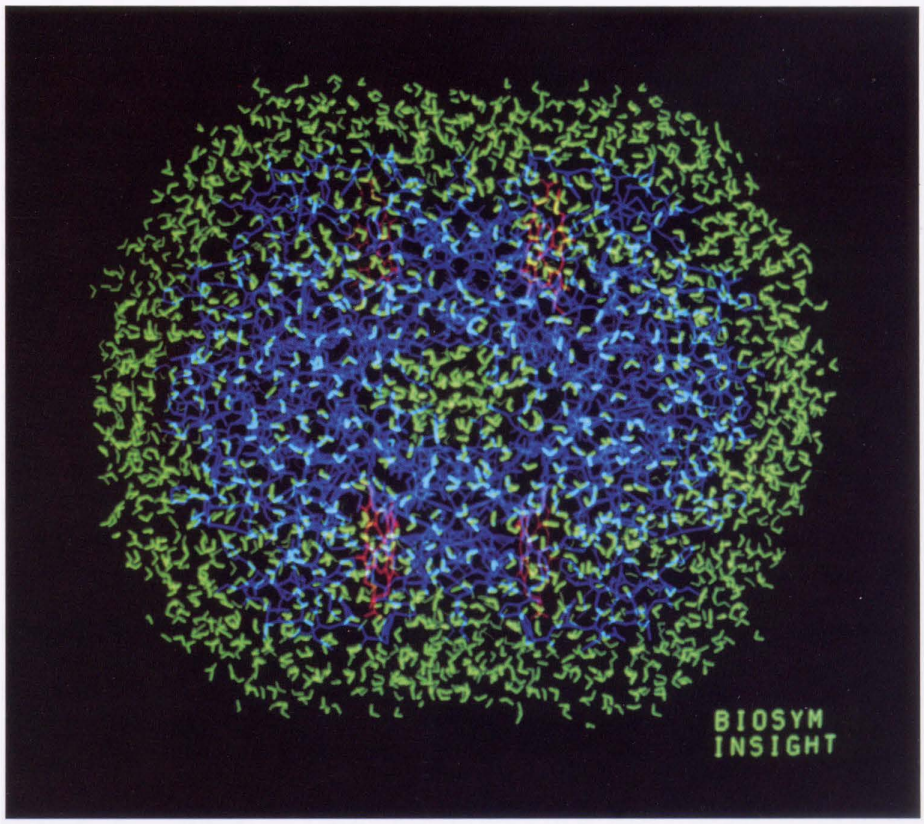
DISCOVER is a rigorous full-featured molecular simulation program for applications in computer-assisted molecular design. DISCOVER incorporates a broad spectrum of molecular mechanics and dynamics algorithms and methodologies that have demonstrated applicability to drug-design problems. It features an English-language command structure and an array of built-in strategies designed to boost ease of use and efficiency for chemists studying structure-related effects.

DISCOVER accepts information from Brookhaven and other molecular databases and includes an interactive molecule binding/editing facility. The program is being used at academic and industrial sites for molecular design problems in areas including agricultural chemicals, food, genetic engineering, and pharmaceuticals.

Some of DISCOVER's molecular mechanics functions include

- ☐ Flexible geometry energy minimization with a choice of energy optimizers including steepest descents, conjugate gradients, quasi-Newton Raphson, and a full Newton Raphson
- ☐ Forcing of selected torsion angles and atom-atom distances to user specified values during energy minimization
- ☐ Simulation of solvent molecules
- ☐ Crystal energy minimization
- ☐ Calculation of atom-atom interaction energies for a subset of the residues in the system
- ☐ Calculation of interaction energies between a selected residue and all residues within a specified distance
- ☐ Automated assignment of force field parameters for geometry refinement
- ☐ Harmonic constraint or fixing of selected ranges of residues or atoms, including relaxation of side chains while backbone atoms remain fixed

DISCOVER's molecular dynamics functions include simulations at a constant temperature, inclusion of solvent molecules, and harmonic constraint, or fixing, of selected ranges of residues or atoms during simulation.



The structure of alpha-4 hemoglobin in the presence of water and ligands, as simulated by DISCOVER software.

Results of calculations can be visualized using INSIGHT, BIOSYM's three-dimensional, real-time molecular modeling program. INSIGHT can selectively display and manipulate molecules and their components, with emphasis on large molecules.

For more information about DISCOVER, contact Christine Sheppard, BIOSYM Technologies, Inc., 9605 Scranton Rd., Suite 101, San Diego, CA 92121; telephone: (619) 458-9990.

GRADSCF designed with Cray systems in mind

GRADSCF, an ab initio quantum chemistry program created for the Cray system environment, provides energies and properties for a variety of molecules of chemical interest.

The program is available for CRAY-1 and CRAY X-MP systems running the Cray operating system COS. GRADSCF was ported recently to a CTSS en-

vironment, and a UNICOS version of the program is anticipated by the end of 1987.

The program predicts equilibrium structures, locates saddle points on multidimensional potential surfaces, and calculates harmonic force constants and vibrational frequencies. These methods also have been used to predict the intensities of infrared vibrational bands.

The most prominent feature of GRADSCF is the capability to calculate analytically the gradient of the potential energy with respect to the nuclear coordinates for systems in which atomic orbitals are expanded in terms of a Gaussian basis set, and in which the wavefunction is restricted to several types of self-consistent field (SCF) functions. Functions of the program include

- Evaluation of an SCF wavefunction and energy at a specified geometry
- Location of an optimum geometry for a molecular system that lies at an energy minimum, subject only to implied symmetry constraints present in the initial structure
- Evaluation of a vibrational spectrum of the molecule by calculating the matrix of second derivatives, from which the force constants and vibrational frequencies are derived
- Location of stationary points that are not necessarily minima with respect to the energy, and thereby find saddle points, or transition states in chemical reactions

Some of the one-electron properties the program provides include a Mulliken population analysis and the dipole and quadrupole moments of the molecule. If requested, the program also will provide the polarizability tensor for the molecule.

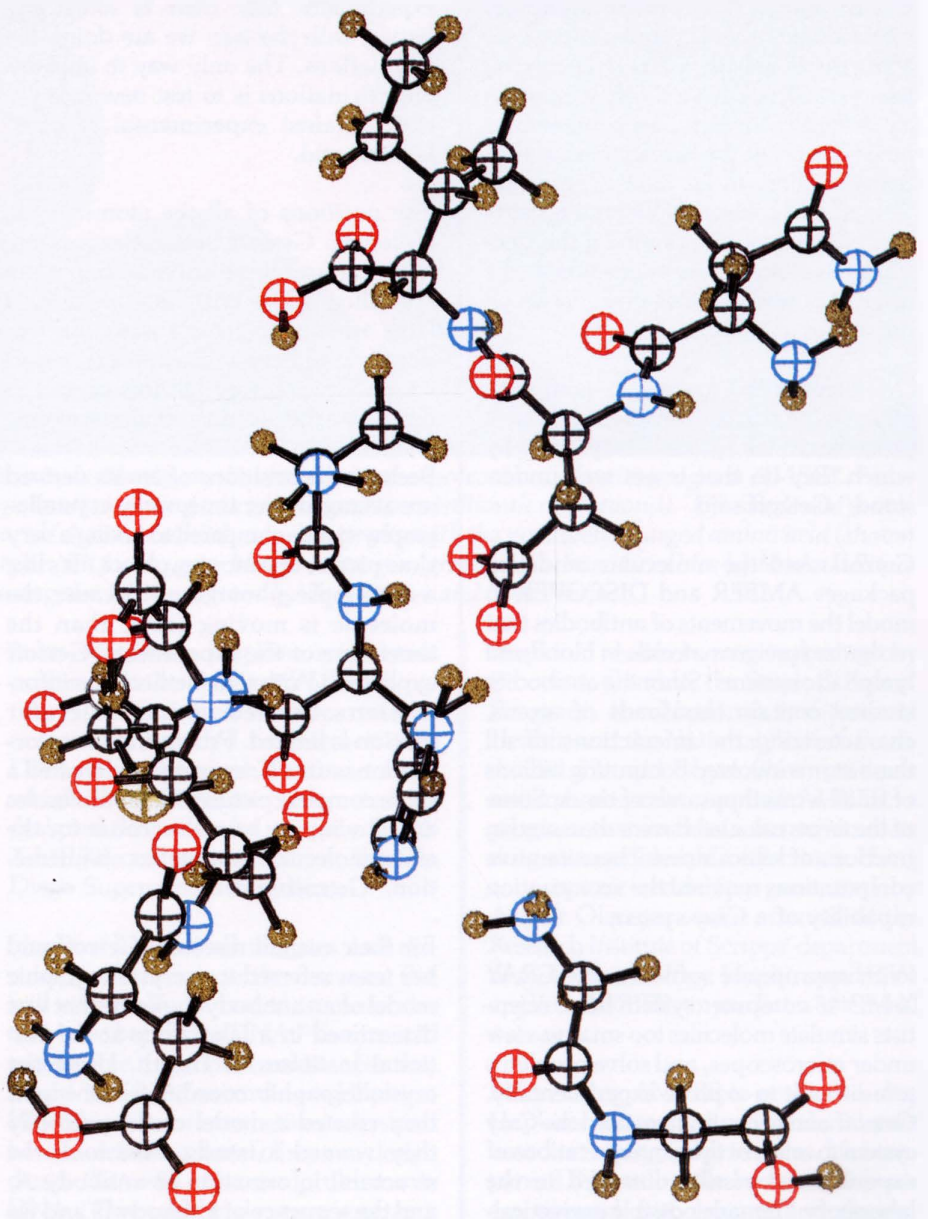
GRADSCF has been installed at Cray system sites in Europe, the United States, and Japan. For more information about using GRADSCF with Cray computer systems, contact Andrew Komornicki at Polyatomics Research Institute, 1101 San Antonio Road, Suite 420, Mountain View, CA 94043; telephone: (415) 964-4013.

MOPAC probes chemical reactions

MOPAC is a general-purpose, semi-empirical molecular orbital program available for use on Cray systems. The program was created for the study of chemical reactions involving molecules, ions, and linear polymers. MOPAC implements the semi-empirical Hamiltonians MNDO, MINDO/3, and AM1. The fully integrated program combines the calculations of vibrational spectra, thermodynamic quantities, isotopic substitution effects, and force constants.

The program originally was developed for use on a VAX 11/780 computer, but as the size of the systems being studied increased, it became essential to convert the program to run on supercomputers, such as Cray systems. As a result, MOPAC has been vectorized to make use of the Cray architecture.

For more information about using MOPAC on Cray computer systems, contact James J. P. Stewart, Frank J. Seiler Research Laboratory, USAF Academy, Colorado Springs, CO 80840-6528; telephone: (303) 472-2655.



Active site in α chymotrypsin produced from MOPAC simulation.

USER NEWS

Scripps researchers investigate molecular interaction

Understanding how protein molecules interact with other types of molecules is a puzzle Elizabeth Getzoff is piecing together with a CRAY X-MP/48 computer system at the San Diego Supercomputer Center. At the Research Institute of Scripps Clinic in La Jolla, California, Getzoff and colleagues Victoria Roberts and Gary Liao are examining the electrostatics of molecular interaction — the result of positive and negative charges on the atoms of proteins.

"Molecules need to interact with each other in the correct ways for biological processes to take place. The process by which they do that is not well understood," Getzoff said.

Getzoff used the molecular modeling packages AMBER and DISCOVER to model the movements of antibodies that recognize foreign materials in blood and lymphatic systems. Since the antibodies studied contain thousands of atoms, characterizing the interactions of all these atoms involved computing billions of small terms thousands of times. Some of the terms calculated were measured in fractions of kilocalories. These iterative computations required the vectorization capability of a Cray system.

With appropriate software, the CRAY X-MP/48 computer system helps scientists simulate molecules too small to view under microscopes, and solve problems too difficult to explore experimentally. Getzoff and her colleagues used the Cray system to analyze their interpretations of experimental results obtained in the laboratory. "It made possible energetically quantifying a solution that we had proposed in a qualitative, pictorial way," she said.

"If our computations cannot explain the experiments, then there is something wrong with the way we are doing the calculations. The only way to improve the calculations is to test new calculations against experimental results," Getzoff said.

The positions of all the atoms in the molecules Getzoff and colleagues are studying have been solved using x-ray crystallography. With this method, a large amount of protein was collected, purified, and crystallized. Next, x-rays were diffracted through this crystal to determine the positions of all the atoms.

Because the positions of atoms derived are averaged over time, x-ray crystallography can be compared to taking a very slow picture of a moving object. "It's like a time-lapse photograph because the molecule is moving faster than the timeframe of the experiment," Getzoff explained. With this method, the information obtained about molecular motion is limited. Processing the information with a Cray computer created a more complete picture. "Since molecules are moving, we have to account for the effect molecular motion has on interaction," Getzoff said.

For their current research, Getzoff and her team referred to a crystallographic model of an antibody molecule that was determined in a laboratory at the National Institutes of Health. Using the crystallographic coordinates as input, they created a model of the antibody they wanted to study. "We took the structural information of antibody A, and the sequence of antibody B, and we made a structural model for antibody B," Getzoff explained. The team used this model as the basis for their computational calculations.

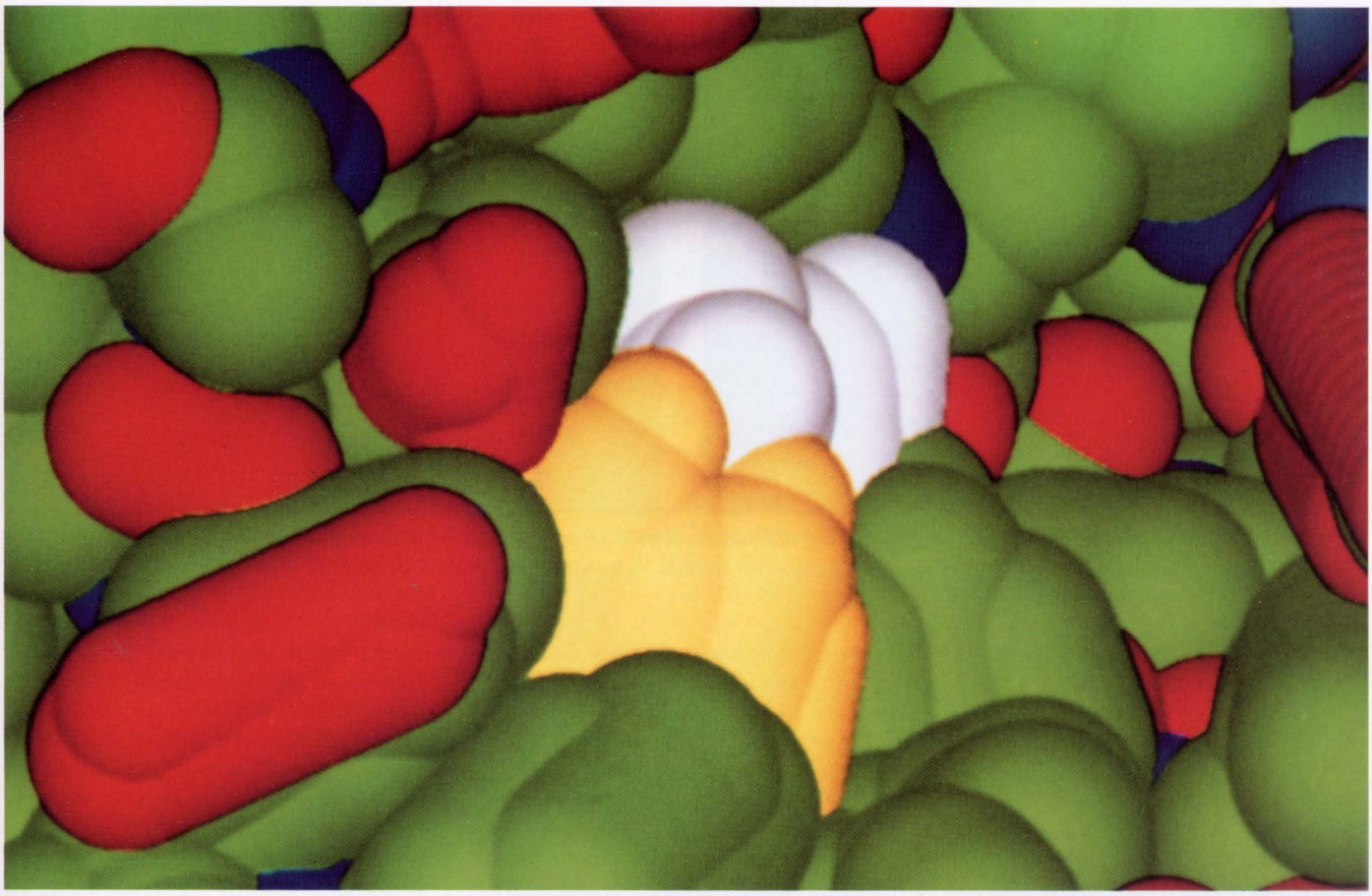
She said the Cray computer system helped her perfect the model and understand its stability. The computer also simulated the molecular motions over time. "You can make a crude model and then allow the computer to make the model better."

Of the two antibodies modeled, one was normal; the other differed by one amino acid, a building block of protein. This single change is called a single-site mutation, because one change has taken place in the genetic material that provides a code for the antibody. As discovered by researchers in Matthew Scharff's laboratory at the Albert Einstein College of Medicine, the normal antibody was capable of recognizing a known compound, while the altered antibody could not recognize the compound.

"We're trying to understand why this single change stopped the binding from occurring. It was a mystery why this antibody with one change — far away from the binding site — no longer bound. We think it results from a rearrangement of one of the amino acids in the antibody, so it points in a different direction," she said.

Next, the researchers plan to model electrostatic interactions in the normal and altered antibodies. Then they will characterize movements of the antibodies' side chains to determine their influence on electrostatic recognition.

For Getzoff and her associates at Scripps, the results of this research will be only one step toward completing the puzzle of molecular interaction. "It is a case of



Close-up view of movements of amino acids in an antibody, as simulated with the molecular dynamics program AMBER. Fourteen positions of each sphere represent atomic motion during one picosecond (one trillionth of a second). The single site mutation occurs at the negatively charged amino acid (shown in white), which interacts with the positively charged amino acid (shown in yellow). Remaining colors indicate atom type: red for oxygen, blue for nitrogen, green for carbon. Graphics by Michael Pique of the Research Institute of Scripps Clinic using the CPK program by Tom Porter and Tom Ferrin.

finding what aspect of this research to look at next. The results may answer the question specifically for this antibody, but we have other molecules we want to look at."

Their answers will help scientists predict how protein molecules will interact with unknown complexes. Once scientists can comprehend how the body makes antibodies to recognize foreign molecules, and how these molecules interact, they can apply this knowledge to uncover why molecular defects cause diseases.

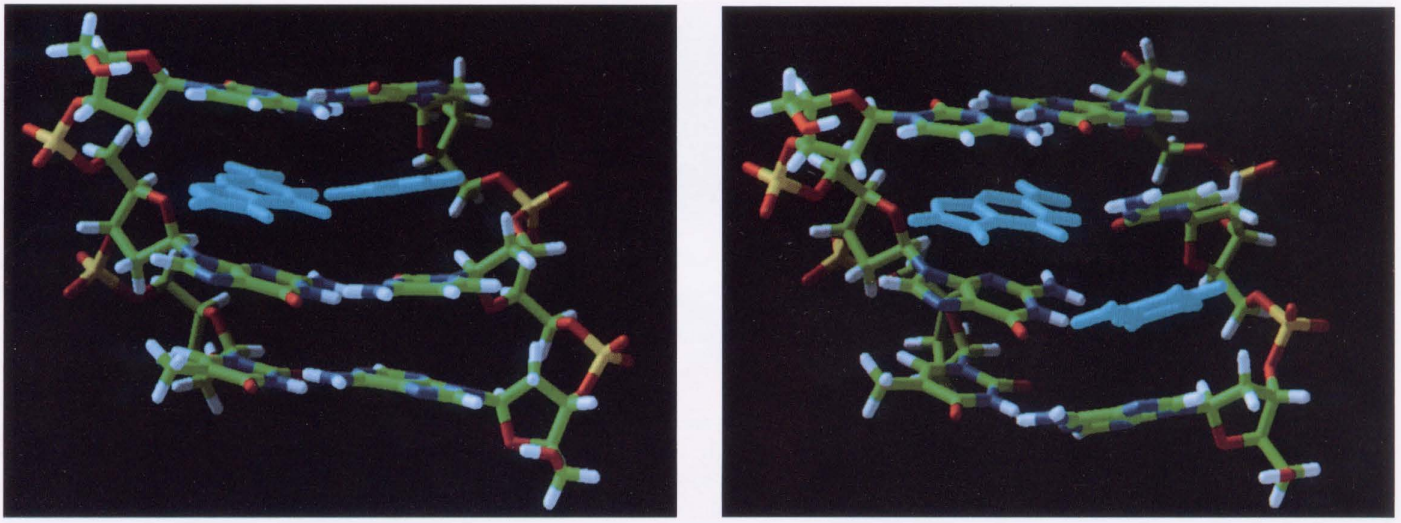
Scientist simulates drug-DNA complexes with Cray system

Paul Hopkins is a soldier in an army of scientists working to understand molecular recognition. Hopkins, an assistant professor at the University of Washington, has designed small artificial molecules intended to interact with DNA, the chemical database from which all organisms are constructed. He designed the molecules using the CRAY X-MP/48 computer system at the San Diego Supercomputer Center.

Hopkins believes all illnesses and diseases, from cancer to the common flu, will eventually be traced to molecules that have bonded to DNA. "Even a cold bug is able to produce substances that bind with DNA," according to Hopkins. When biochemists can better understand how molecules interact, they can design drugs that will discover problems in DNA. "One could make some spectacular drugs," Hopkins says.

Hopkins' research was supported by Procter & Gamble, the Searle Scholars' Program, the National Institutes of Health, and the National Science Foundation. He and Regan Shea, graduate research assistant, applied AMBER, a program that enables users to simulate drug-macromolecule interactions. Using a computer rendering generated by Arthur Olson, a senior scientist at the Research Institute of Scripps' department of molecular biology, Hopkins and Shea applied AMBER to predict a chemical outcome.

Before using AMBER to predict results, Hopkins designed an artificial molecule based on biological principles. He compares DNA to braille because the rods are covered with bumps and depressions. "If you can make a molecule that



Computer-generated illustration of DNA interacting with a small molecule in the well-known intercalation mode (left), and the new sequence-recognizing mode proposed in this research (right). Graphics generated by Michael Pique and Arthur Olson of the Research Institute of Scripps Clinic.

fits into the bumps, it can read the braille," he says.

Hopkins' science is called "molecular recognition," which is the study of how molecules recognize each other. When molecules meet, shape and electrical charge are two of many factors that determine their reaction. Molecules that bond with DNA can be grouped into three classifications. One class, the *intercalators*, bond to the DNA interior. The *groove binders* are curved molecules that "snuggle up to the braille," as Hopkins describes it. Third, the positively charged *electrostatic binders* react with the negatively charged DNA. However, getting shapes to match is only part of the battle. When molecules meet, other factors determine their reaction, such as hydrogen bonding and a range of solvent effects.

Hopkins created a molecular model similar to an intercalator. "It was like an intercalator, but different enough that people said it looked odd to them," he says. Working with Olson, Hopkins took a set of coordinates from DNA with an intercalator attached. Olson helped him remove the interior-bonded intercalator to create an isolated DNA with no drugs attached. His assistance enabled Hopkins to create molecular graphics to obtain the molecules' coordinates. Next, they bonded a more generic intercalator to the DNA to become a reference DNA-drug complex — their benchmark.

After creating this reference complex, they processed their results on the Cray system. While the program was not yet at a stage to design molecules from scratch, they could compare the graphically generated complexes to known complexes, showing some similarites and differences.

In a second experiment, Hopkins and associates formulated a new drug of their own design. The Cray system computed the energetics of the simulation. When compared to their benchmark, the new drug-DNA complex was found to be energetically equivalent. Hopkins used the Cray system for practical reasons. "We could have used the VAX, but that would have taken much longer. What would have been days or months on a VAX was a matter of minutes or hours on the Cray system. The Cray system made it possible for us to be imaginative, to make mistakes and try again," he says.

When scientists tested the results in the laboratory, the molecules did not bind to DNA. However, this result in no way doomed the concept. Says Hopkins, "This does not mean the computations were wrong; the computer results were a simplification and provide only a direction. That the laboratory experiments failed suggests we have not yet fine-tuned the experiments."

He explains that while it is easier to create synthesized molecules on paper, it is quite difficult to concoct the molecules in a laboratory. "The whole proposition is risky business. Many have told me I'm brave to do such a thing," he says. Since there are so many unknown factors, the risk Hopkins faces is the likelihood that the substances he simulates may not bind to DNA in the laboratory. But, he adds, the risk is worth the gain. "It's a high payoff area. In any event, you learn something," he says.

Meanwhile, Hopkins must put his research on hold until his findings can be applied in the laboratory. "I would like to work on this again, but the project needs a delay for other areas to mature. The time is not really ripe for this type of drug design," he says.

When the time does ripen, the fruits of Hopkins' research will be worth the wait. "I am firmly convinced that drug-DNA complexes of the type I have proposed and studied by computer will one day be formed in the laboratory," says Hopkins. When computer-simulated molecules can be produced chemically, the results may enable scientists to synthesize drugs with enough power to stamp out cancer and other illnesses.

Cray system aids search for room-temperature superconductivity

It has been deemed the hottest topic on the scientific frontier — a race scientists are running at breakneck speed.

The discovery of high-temperature superconductors by J. Georg Bednorz and K. Alex Muller in Zurich, and Paul Chu's group in Houston, has challenged scientists from research institutions around the world to find room-temperature superconductors. These scientists are searching for metallic materials that will conduct electricity without resistance and without the need for supercooling.

Arthur Freeman, Morrison Professor of Physics at Northwestern University, is using a CRAY-2 computer system to simulate the electronic properties of superconductors. "Simulating all of these properties helps us understand materials and why they are superconducting," he said.

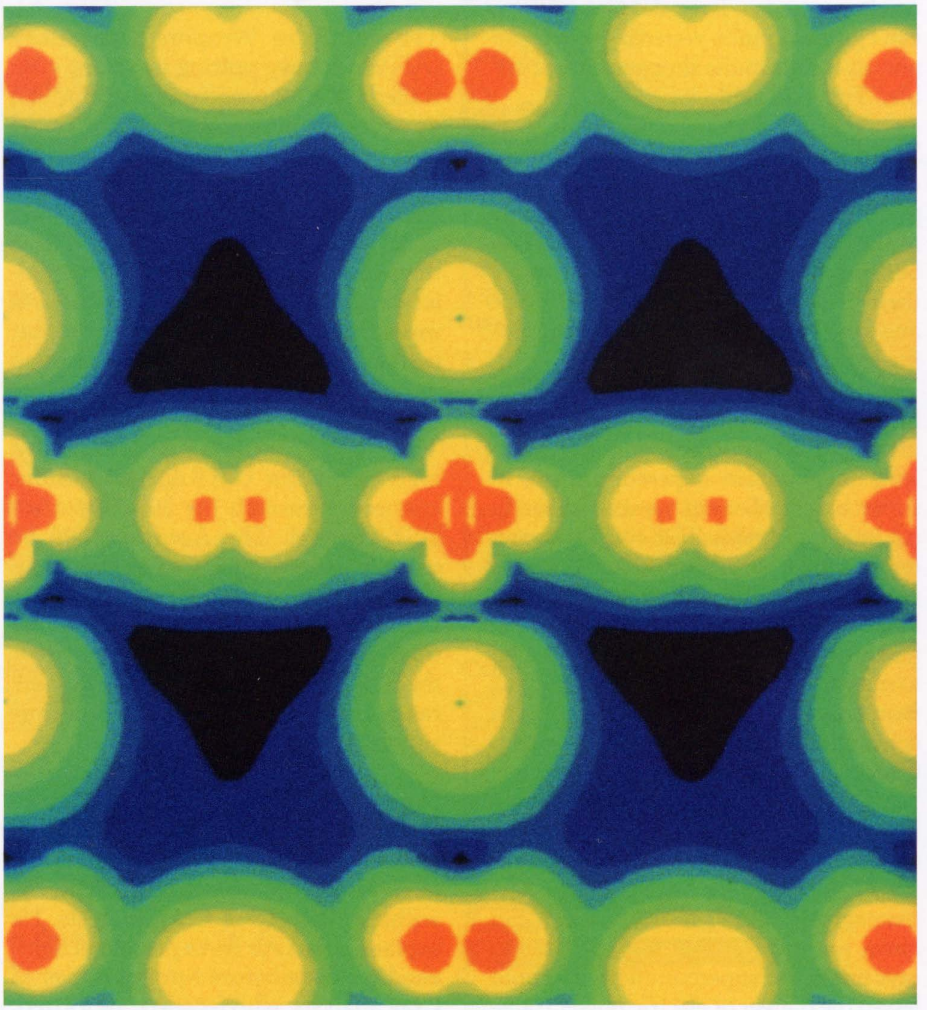
With support from the National Science Foundation, Freeman and collaborators used two CRAY-2 systems for their research, one at the University of Minnesota, the other at the NASA Ames Research Center in California. Assisting Freeman were Jaejun Yu and Sandro Massidda at Northwestern University and Dale Koelling at Argonne National Laboratory.

The computer simulations give Freeman's group a bird's-eye view of the electron structure of superconductors. "In these simulations we actually see inside, on an atomic scale, the distribution of the electrons in the material. All properties are determined by the electrons," he explained. This view enables the researchers to understand both the physics and chemistry of these materials in their normal states, and may help to provide a theoretical explanation for record-high transition temperatures observed, now well over 100° K.

To solve the complex electronic calculations that account for all the interactions of the electrons and nuclei of the atoms, Freeman and his collaborators developed their own *full-potential linearized augmented plane wave method* to obtain highly precise solutions to the time-independent quantum mechanical equations of electron motion. With this method, Freeman compared the electrical properties of two high-temperature superconducting materials: $YBa_2Cu_3O_7$ (yttrium-barium-copper-oxide) and $La_{2-x}Ba_xCuO_4$ (lanthium-barium-copper-oxide).

In Y-Ba-Cu-O, both chains and planes of copper and oxygen exist. In La-Ba-Cu-O, only planes of copper and oxygen exist. Freeman links the former material's higher-temperature superconductivity to its combination of planes and chains. He decribes planes as the surfaces formed when a material has chains of copper and oxygen atoms alternating along both x and y axes. When the oxygen atoms are entirely removed from the x axis, alternating chains remain only on the y axis.

"The results tell us these are very unusual materials. Metallic conduction is only in the copper oxide metallic planes and metallic chains," Freeman said. He also found that the materials are highly anisotropic, meaning their properties are highly directionally dependent. These findings differ from the original theories about high-temperature superconductivity. "The traditional electron-photon mechanism for superconductivity is insufficient — it's too weak and needs to be augmented," he commented.



Contour map of the electronic charge around copper and oxygen atoms in the superconductor $YBa_2Cu_3O_7$. In this vertical plane, copper atoms are aligned vertically through the bottom, center, and top of the image, connected to adjacent oxygen atoms.

According to Freeman, the rapid development and availability of supercomputing power in this decade has strongly impacted condensed matter physics. Freeman attributes the success of his full-potential linearized augmented plane wave method in solving these problems to the speed and memory of the CRAY-2 system.

More than 10 million words of memory were needed to obtain highly precise results. "This problem was made for the CRAY-2 system. We're getting more truth, better solutions, and because of larger memory, we can do it a lot faster. When doing research in such an exciting area, you don't want to obtain results six months or a year from now; we want results yesterday. Even the Cray system can't give us results yesterday, but it can give us results in the shortest time possible," Freeman said.

Speed is essential to physicists racing to discover room-temperature superconductivity, and the practical benefits make winning even sweeter. "The benefits are very clear, which is why there is enormous activity worldwide and why so many companies are jumping into it," Freeman said.

Higher transition temperatures would mean simpler and cheaper cooling systems to maintain superconductors in supercomputer components and medical imaging devices. Of course, room-temperature superconductors would require little cooling. Since the liquid helium needed to cool many existing machines is expensive and complicated to maintain, improved superconductors would be a cost-effective alternative to conventional low-temperature superconductors.

In fact, the benefits of room-temperature conductivity would be limitless. For example, high voltage lines generate electromagnetic fields for 150 feet that can harm livestock and humans. Freeman says improved superconductivity could lessen the danger of electric power transmission, since wires could be buried underground. Also, power transmission could be less expensive, since room-temperature superconductors could transmit power a thousand miles without loss of energy. In general, devices that were once impractical to create using liquid helium could be developed with less complication.

Freeman expects scientists to discover room-temperature superconductivity within the next two years, and predicts some of the benefits will emerge within five years. Meanwhile, Freeman will continue to simulate superconductors on the CRAY-2 system, hoping to be the first to find solutions. "No one remembers the second one to make a discovery," he said.

Air Force Weapons Laboratory reaches new heights

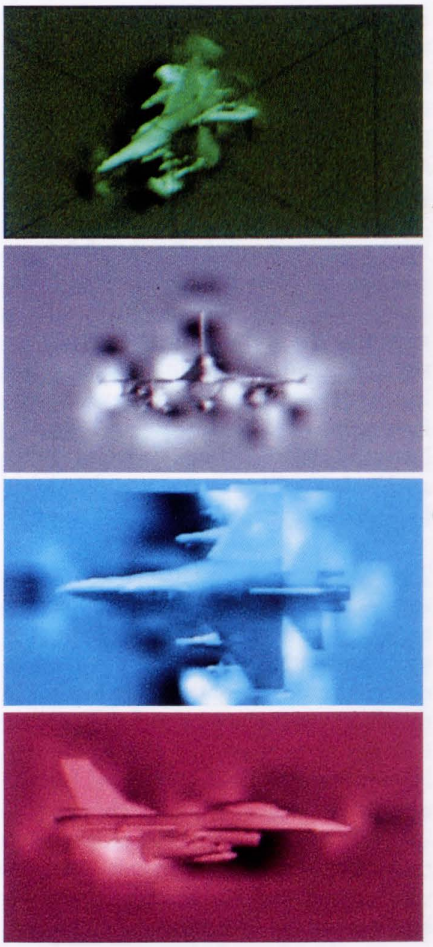
The Air Force Weapons Laboratory (AFWL) is busy putting its recently installed CRAY-2 supercomputer through its paces. Applying multitasking on all four of the system's processors, the lab at Kirtland Air Force Base in New Mexico achieved a peak integration rate of 3.5 million cells/sec for NEWTUN, a three-dimensional Eulerian fluid dynamics program. The integration rate is a measurement of the size of a grid that can be solved over one Courant time step in one second.

The CRAY-2 system is supported with a high resolution ULTRA frame buffer connected through the 100-Mbyte/sec HSX channel. The Air Force will be using its supercomputer to produce real-time simulation and visualization (24 Courant time integrations per second) of three-dimensional flow phenomena for the F-16 and other aircraft. Grid sizes will range from 125,000 cells for real-time simulations to 40 million cells.

Nazareno L. Rapagnani, chief, Communication-Computer Systems Technical Office at AFWL, said the high performance would not have been achieved without the help of Cray analysts. He explained that Air Force personnel tightened the code before turning it over to Cray analysts, who recoded, vectorized, and multitasked NEWTUN.

"It took about one Cray man-month and the work has really paid off. With their help we have realized a factor of 40 improvement on the CRAY-2 over the CRAY-1 system. There is no doubt in my mind that many Cray systems are underutilized. As you can see, the extra effort put in to make NEWTUN use the hardware was well spent," Rapagnani said.

Those interested in learning more about NEWTUN should refer to *Fast Interactive Graphics and the Numerical Electronic Wind Tunnel*, AFWL publication number 87-33. It may be obtained by writing to Nazareno L. Rapagnani, chief, Communication-Computer Systems Technical Office, Air Force Weapons Laboratory, AFWL/SC, Kirtland Air Force Base, NM 87117-6008.



A preliminary calculation of the external transient Euler flow field of an F-16 flying at Mach 1.8 was performed. The flow simulation and graphics visualization for 24 iterations required a total processing time of less than one second.

WORLDWIDE BUSINESS CENTERS

Domestic

Corporate Headquarters
608 Second Avenue South
Minneapolis, MN 55402
(612) 333-5889, telex 4991729

Government Marketing/
Government Relations Office
National Place
1331 Pennsylvania Avenue, Suite 1331 N
Washington, DC 20004
(202) 638-6000

Central Region Office
5350 Manhattan Circle
Boulder, CO 80303
(303) 499-3055
- Albuquerque
- Chicago
- Colorado Springs
- Detroit
- Minneapolis/St. Paul
- St. Louis

Eastern Region Office
4041 Powder Mill Road, Suite 600
Calverton, MD 20705
(301) 595-5100
- Atlanta
- Bernardsville
- Boston
- Cincinnati
- Darien
- Huntsville
- New York
- Pittsburgh
- Tampa
- Washington, DC

Petroleum Region Office
5847 San Felipe, Suite 3000
Houston, TX 77057
(713) 975-8998
- Dallas
- Tulsa

Western Region Office
5776 Stoneridge Mall Road, Suite 350
Pleasanton, CA 94566
(415) 463-2800
- Huntington Beach
- Los Angeles
- Seattle
- Sunnyvale

International

Cray Asia/Pacific Inc.
Suite 1003
Sutherland House
3 Chater Road Central
Hong Kong
(5)-214669, telex 89765

Cray Canada Inc.
4211 Yonge Street, Suite 210
Willowdale, Ontario
Canada M2P 2A9
(416) 229-2729

Cray Research France S.A.
7 Rue de Tilsitt
75017 Paris, France
(01)-766-01-55, telex 660568

Cray Research GmbH
Perhamerstrasse 31
8000 Munich 21, West Germany
(089)-56014-0, telex 5213211

Cray Research S.R.L.
Via Vivaio, 15
20122 Milan, Italy
(02)-70-98-27, telex 315403

Cray Research Japan, Limited
Ichibancho Eight One Building
6-4 Ichibancho
Chiyoda-Ku, Tokyo 102, Japan
(03)-239-0711, telex 28893

Cray Research (UK) Ltd.
Cray House
London Road, Bracknell
Berkshire RG12 2SY
United Kingdom
(0344)-485971, telex 848841

CRAY RESEARCH, INC.

Cray Research, Inc.
608 Second Avenue South
Minneapolis, MN 55402
Tel: (612) 333-5889
Telex: 4991729