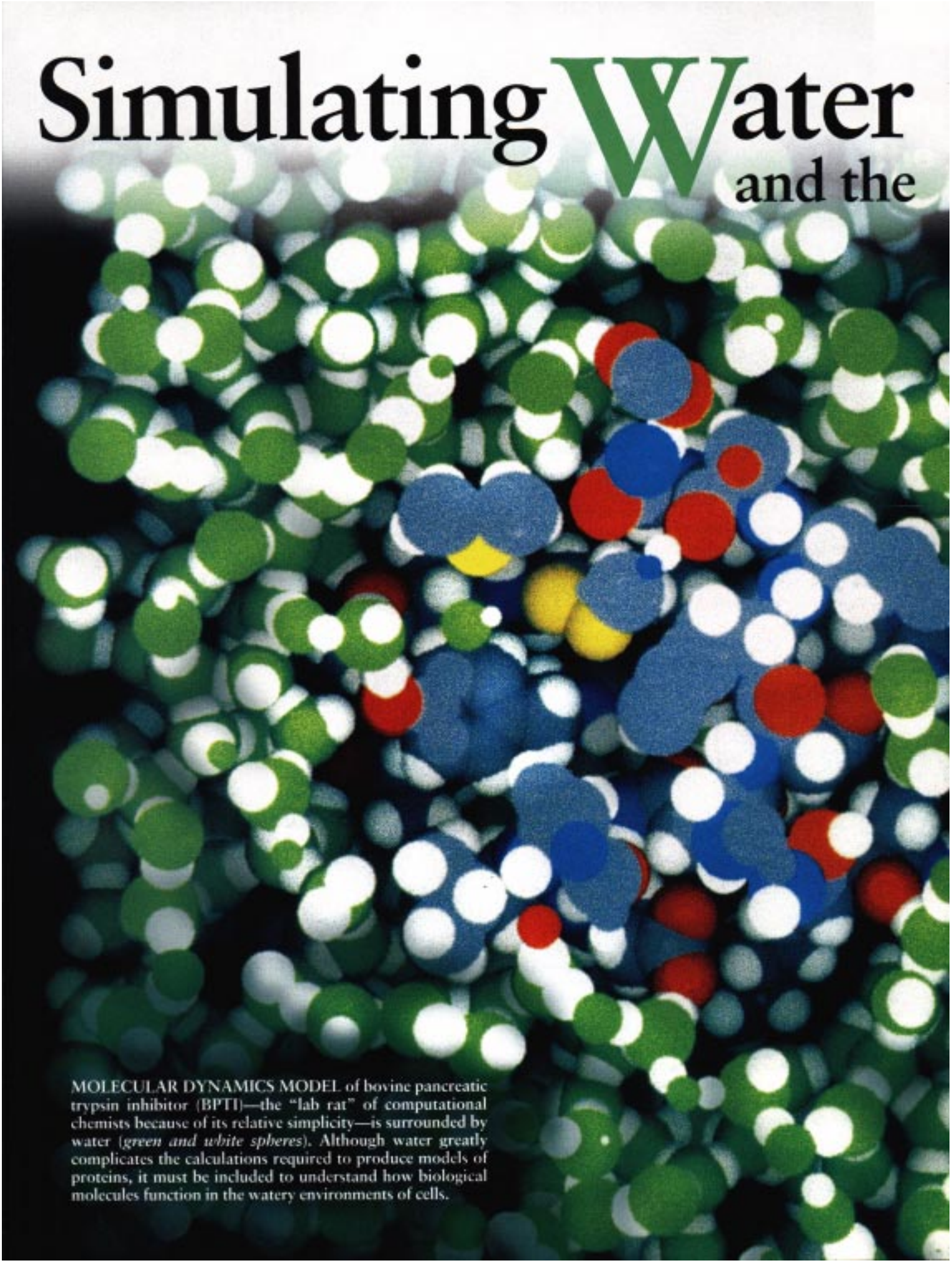
Simulating Water and the

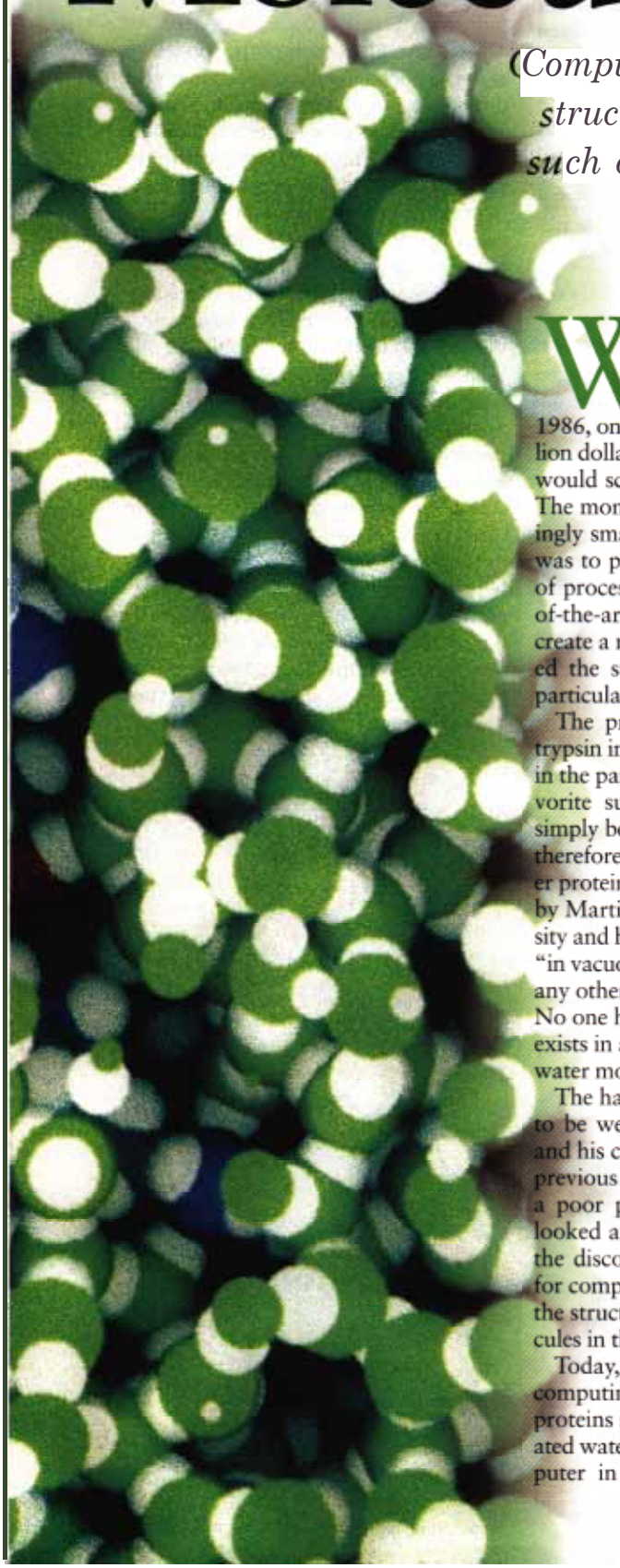


MOLECULAR DYNAMICS MODEL of bovine pancreatic trypsin inhibitor (BPTI)—the "lab rat" of computational chemists because of its relative simplicity—is surrounded by water (*green and white spheres*). Although water greatly complicates the calculations required to produce models of proteins, it must be included to understand how biological molecules function in the watery environments of cells.

Molecules of Life

Computer modeling reveals how water affects the structures and dynamics of biological molecules such as proteins, yielding clues to their functions

by Mark Gerstein and Michael Levitt



Water is cheap, if not free, in most places in the world. But during the summer of 1986, one of us (Levitt) spent half a million dollars on an amount of water that would scarcely wet the point of a pin. The money was not to buy the vanishingly small amount of water. Rather it was to pay for the roughly two weeks of processing time on a gigantic state-of-the-art supercomputer required to create a model of how the water affected the structure and movement of a particular protein.

The protein was bovine pancreatic trypsin inhibitor (BPTI), which is found in the pancreases of cattle. BPTI is a favorite subject of computer modelers simply because it is relatively small and therefore easier to study than most other proteins. It had been modeled before, by Martin Karplus of Harvard University and his colleagues in 1977, but only "in vacuo" (as if in a vacuum)—without any other molecules interacting with it. No one had visualized BPTI as it really exists in a living cell, with thousands of water molecules surrounding it.

The half a million dollars turned out to be well spent. Not only did Levitt and his colleague Ruth Sharon find the previous in vacuo model of BPTI to be a poor predictor of how the protein looked and behaved in the real world, the discovery helped to pave the way for computational chemists to simulate the structures of other biological molecules in their native, watery environs.

Today, given the great advances in computing technology, we can model proteins such as BPTI and their associated water molecules on a desktop computer in a couple of days, spending about 80 cents for electricity. Scientists have now simulated the aqueous ("in water") structures of more than 50 proteins and nucleic acids such as DNA.

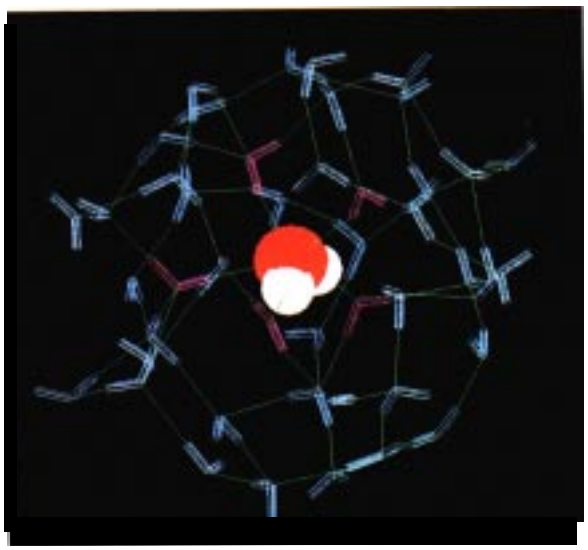
Why is it so important to understand the effects of water on the shapes of biological molecules? Principallv. because a molecule's struture yields clues to how it functions, helping scientists decipher the biochemical inter-actions that add up to life. On a more practical level, understanding the structures of biological molecules in water may one day help researchers design new drugs that act by blocking or enhancing various biochemical pathways.

The Water Within

To understand how water affects the structures of biological molecules, we must first appreciate the distinctive properties of water itself. These properties stem from the unique structure of water and the way this structure allows water to "manage" the electric charges of other molecules.

A single water molecule (H_20) has an essentially tetrahedral geometry, with an oxygen atom at the center of the tetrahedron, hydrogen atoms at two of the four corners and clouds of negative charge at the other two corners. The clouds of negative charge result from the way in which the atomic structures of oxygen and hydrogen combine. In' simplified terms, oxygen has eight negatively charged electrons circling its positively charged nucleus: two in an inner shell and six in an outer shell. The inner shell's maximum capacity is two electrons, so it is full, but the outer shell can hold as many as eight. Hhydrogen has

ALL ILLUSTRATIONS BY MARK GERSTEIN AND MICHAEL LEVITT



HYDROGEN BONDS give water its unique properties. In this model of liquid water, the central molecule *(red and white spheres)* has formed hydrogen bonds (green lines) with five other water molecules *(pink* vs). Its hydrogen atoms *(white)* have bonded with the oxygens of two other waters, and its oxygen atom *(red)* has bonded with a hydrogen from each of three other water molecules. Each molecule of liquid water usually forms four or five hydrogen bonds.

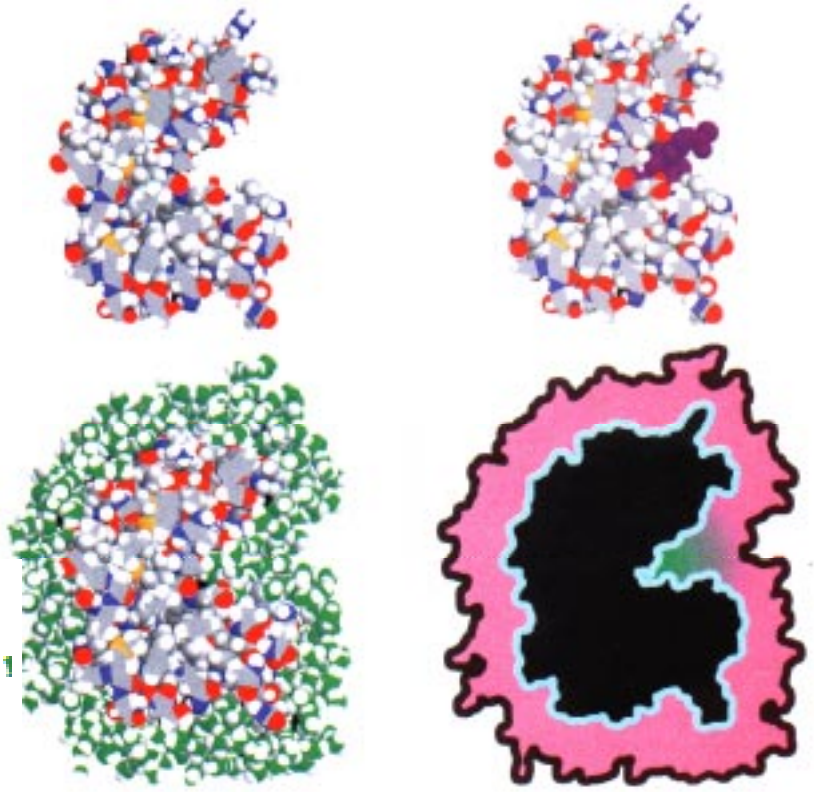
only one electron. When oxygen combines with two hydrogens, it attracts each hydrogen's electron in an attempt to fill its outer shell. Because each hydrogen electron spends more time around the oxygen atom than around its own positively charged nucleus, a water molecule is polar: it has two clouds of slight negative charge around the oxygen atom, and its two hydrogen atoms are left with slightly positive charges. These two types of charges counterbalance one another, however, so that water molecules are electrically neutral.

Chemists usually do not draw the clouds of negative charge around the oxygen atom of a water molecule; instead they conventionally depict water in a V shape ***[see illustration above].*** Each side of the V corresponds to an oxygen-hydrogen bond roughly $1\ 0^{-8}$ centimeter in length. The angle formed between the two sides of the V is close to 105 degrees-slightly less than the 109.5-degree angle formed between any two sides of a perfect tetrahedron.

Because of the polarity of water molecules, interactions between a positively charged hydrogen of one water molecule **and** the negatively charged oxygen of another are favorable. These interactions are called hydrogen bonds. Reflecting water's tetrahedral geometry, each molecule in liquid water often forms four hydrogen bonds: two between its hydrogens and the oxygen atoms of two other water molecules, and two between its oxygen atom and the hydrogens of other water molecules. But the detailed structure of liquid water-unlike ice, which is usually composed of a lattice of water molecules arranged with perfect tetrahedral geometry -- can be quite random and irregular. The actual number of hydrogen bonds per liquid water molecule ranges from three to six, with an average of about 4.5. The necessity of maintaining a tetrahedral, hydrogen-bonded structure gives water an "open," loosely packed structure compared with that of most other liquids, such as oils or liquid nitrogen.

To construct a computer model of water, we need to take into account two different types of forces: intramolecular and intermolecular. The interactions ***within*** a water molecule are modeled in terms of the short-range, springlike forces created by the chemical bonds between each molecule's hydrogens and oxygen. The interactions ***between*** water molecules are modeled in terms of long-range, electrical forces. The intramolecular forces restrain the lengths of the bonds between the oxygen of each water molecule and its hydrogens-and the angle formed between each of these bonds-to certain set values. These forces behave like springs: the more an outside force distorts the bonds, the more the bonds resist the force.

The long-range, intermolecular forces between water molecules behave differently from the intramolecular forces: they decrease in magnitude with increasing distance. Fundamentally, the long-range forces arise from the attraction between opposite charges and the re-



ACTIVE SITE of lysozyme -- a naturally **occurring enzyme that kills bacteria by breaking down sugary molecules in their cell walls-is in the protein's main groove (upper left). The groove is shaped precisely to accommodate the molecules it cleaves *(purple spheres, upper right).* Modeling how water (green *and white spheres, lower* left) interacts with the groove helps scientists to create a map of the active site *(lower right; green shading indicates* easily displaced water molecules). Such maps can be key in designing new drugs to block or enhance a particular enzyme's activity.**

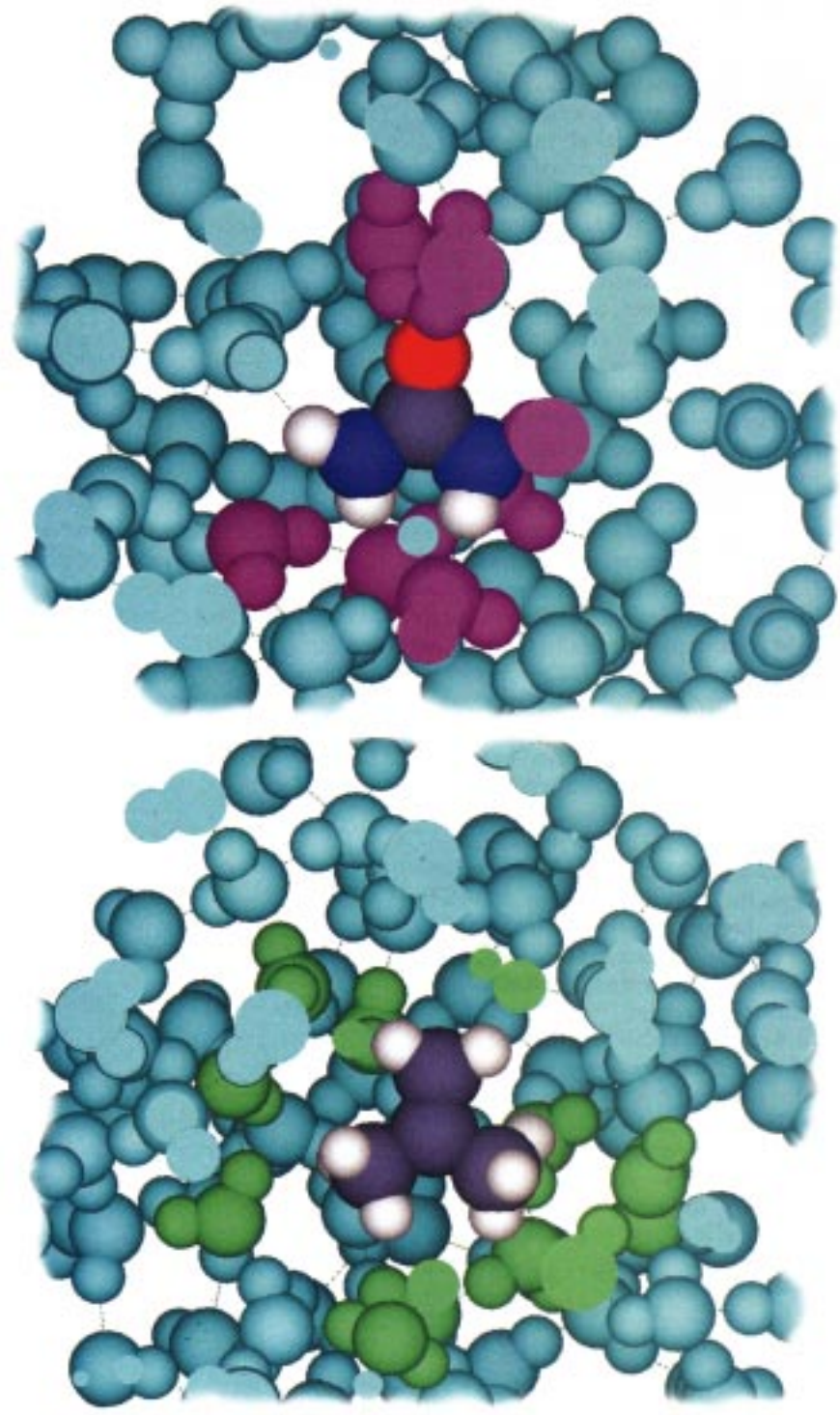
pulsion between similar charges. These forces give rise to hydrogen bonds as well as to weaker attractions called van der Waals forces.

The computer simulation of water molecules was pioneered in the late 1960s by Aneesur Rahman and Frank H. Stillinger of Bell Laboratories. Rahman and Stillinger simulated the motion of 216 water molecules in a rectangular box. (The researchers chose to model 216 water molecules because that number constitutes a box of waters six molecules deep, six molecules high and six molecules wide.) In their five-picosecond simulation-the longest possible using the computing technology of the time-Rahman and Stillinger found that the behavior of water is a direct consequence of the energetic relations among the water molecules. The simulation was able to reproduce quantitatively many of the bulk properties of water, such as its average structure, rate of diffusion and heat of vaporization.

Simulating Life

The importance of water in living processes derives not only from its ability to form hydrogen bonds with other water molecules but also from its capacity to interact with various types of biological molecules. Because of its polar nature, water readily interacts with other polar and charged molecules, such as acids, salts, sugars and the various regions of proteins and DNA. As a result of these interactions, water can dissolve polar molecules, which are consequently described as hydrophilic ("water-loving"). In contrast, water does not interact well with nonpolar molecules such as fats, giving rise to the observation that oil and water do not mix. Nonpolar molecules are therefore termed hydrophobic ("water-fearing").

Biological molecules such as proteins and DNA contain both hydrophilic and hydrophobic parts arranged in long chains. The three-dimensional structures of these molecules are dictated by the way these chains fold into more compact arrangements, so that hydrophilic groups are on the surface where they can interact with water, and hydrophobic groups are buried in the interior, away from water. In 1959 Walter Kauzmann proposed that such a hydrophobic effect was crucial for protein folding, and the role of hydrophobicity in protein folding is still a subject of great interest today [see "The Protein Folding Problem," by Frederic M. Richards; SCIENTIFIC AMERICAN, January 1991].



MOLECULES with nearly identical shapes interact differently with water depending on whether they are polar, or have partial charges on some of their atoms, or lack charges and are therefore nonpolar. Urea, a polar molecule found in urine, forms hydrogen bonds with water molecules (purple spheres, top). In contrast, the nonpolar isobutene does not form such bonds; instead water molecules hydrogen-bond with one another around isobutene, forming a cagelike structure (green *spheres, bottom).*

There are three types of waters that must be considered when building a computer model of a biological molecule in aqueous solution: the "ordered waters" that surround and strongly interact with the molecule, the "bulk waters" beyond and any waters that may be buried within the molecule. A single cell contains billions of water molecules. Almost all the space not occupied by the atoms of biological molecules is filled with water. Human cells are, in fact,

mostly water; the human body is roughly 60 percent water by weight.

How do we model all these waters, together with the individual atoms of a biological molecule? In simple terms, we first describe the basic interactions between all the atoms and then let the system evolve according to the laws of Newtonian physics. Such a simulation requires two basic ingredients: a way to describe the interactions within and among water and biological molecules-the intramolecular and intermolecular forces, in other words-and a procedure for charting their movements through time, called molecular dynamics.

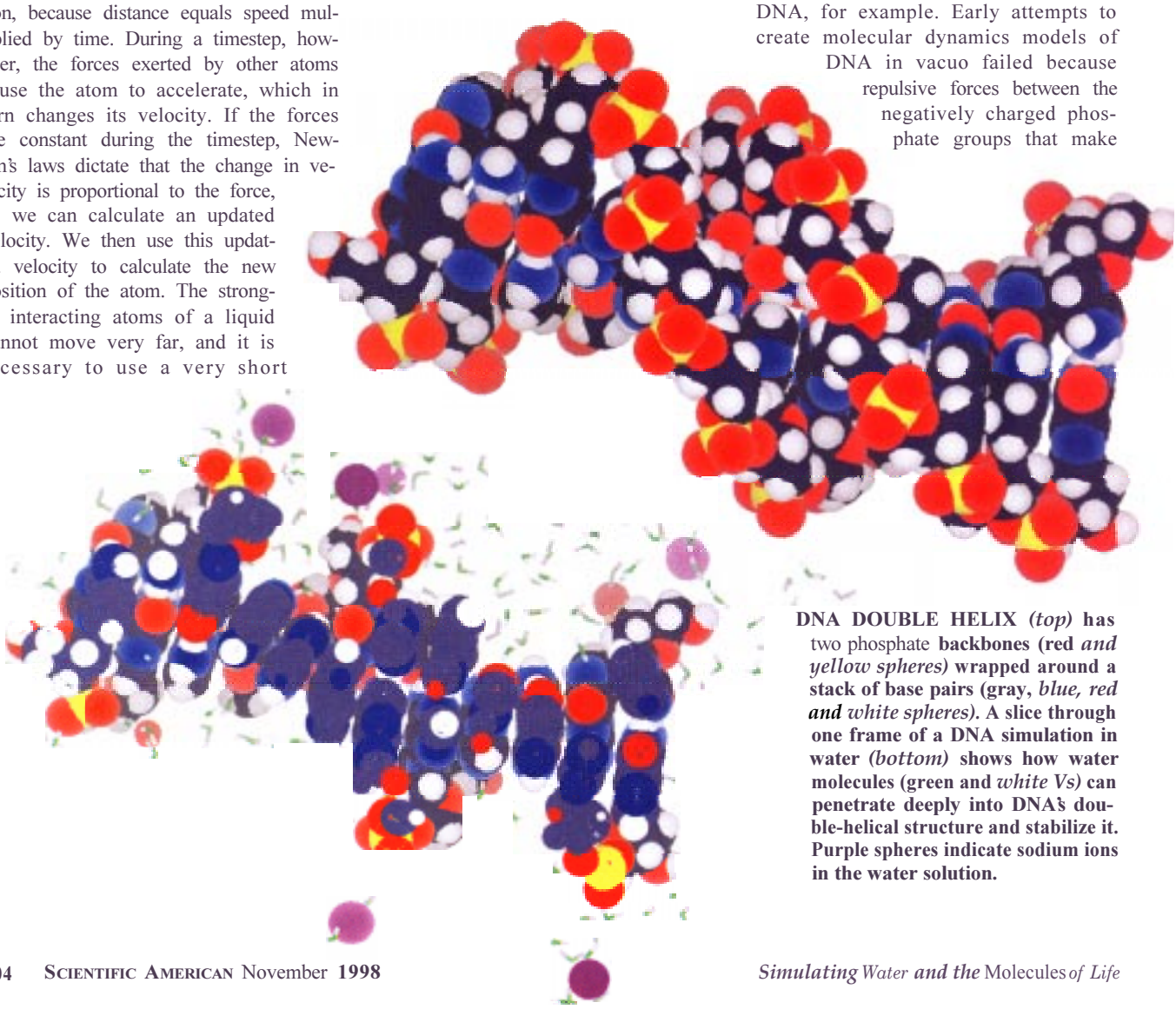
Molecular dynamics produces a sequence of configurations very much like frames in a movie. Each atom moves through time in a series of discrete steps, called timesteps. Essentially, the new position of an atom is its old position plus the distance it traveled during a given timestep. If no forces acted on an atom, the distance it traveled would be a function of its velocity at the previous position, because distance equals speed multiplied by time. During a timestep, however, the forces exerted by other atoms cause the atom to accelerate, which in turn changes its velocity. If the forces are constant during the timestep, Newton's laws dictate that the change in velocity is proportional to the force, so we can calculate an updated velocity. We then use this updated velocity to calculate the new position of the atom. The strongly interacting atoms of a liquid cannot move very far, and it is necessary to use a very short timestep: one femtosecond (10^{-15} second). During this period, a water molecule moves only 1/500 of its diameter.

In a long simulation, calculating each timestep for all the atoms in a biological molecule with its ordered waters yields an enormous amount of data. A small protein in water, for instance, produces half a million sets of Cartesian coordinates in a nanosecond, each describing the positions of about 10,000 atoms. The movie generated by such a simulation is exquisitely detailed. We can see every water molecule rotating, shifting and vibrating over millions of frames.

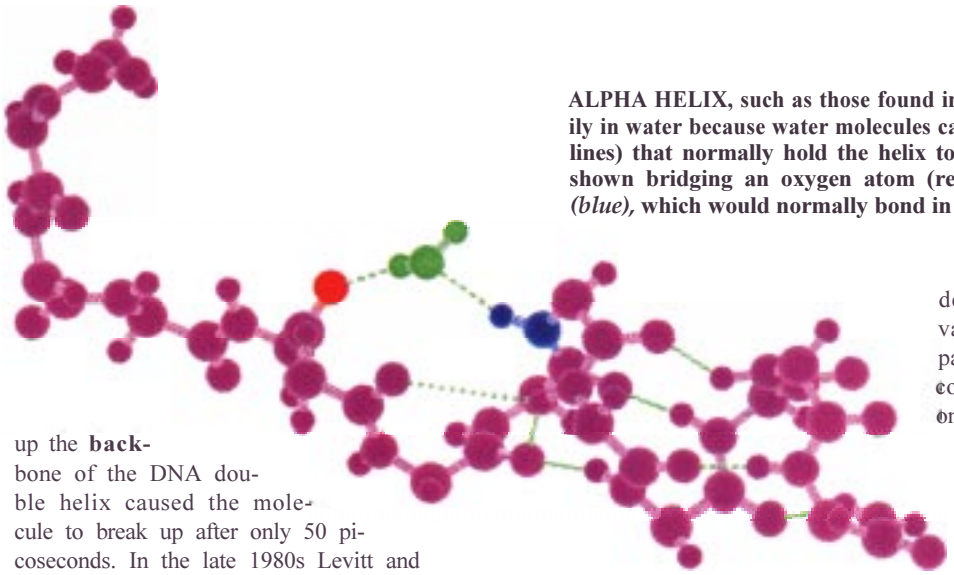
To illustrate how computer simulation can depict the way water affects molecular dynamics, let us consider two simple organic molecules, isobutene and urea, which have similar shapes but very different properties. Isobutene, a fuel produced by oil refineries, is a Y-shaped, nonpolar (and therefore hydrophobic) molecule whose backbone consists of four carbon atoms, two of which are linked by a double bond. Urea is a product of protein metabolism that is excreted in urine. It, too, has a Y-shaped structure: a carbonyl group (C=O) linked to two amino groups (NH_2). Unlike isobutene, urea is a strongly polar molecule that is hydrophilic.

When we carry out molecular dynamics simulations of isobutene and urea, we see that water behaves differently around the two molecules ***[see illustration on preceding*** page]. Water molecules interact directly with urea, forming hydrogen bonds with urea's oxygen and hydrogen atoms as well as with one another. In contrast, water molecules turn away from the hydrophobic isobutene and form hydrogen bonds only among themselves, creating a cage of ordered waters surrounding the molecule.

Visualizing how water molecules interact with such simple molecules helps us understand the behavior of water with more complex biological molecules, such as proteins and nucleic acids. Water is integral to the structure of DNA, for example. Early attempts to create molecular dynamics models of DNA in vacuo failed because repulsive forces between the negatively charged phosphate groups that make



DNA DOUBLE HELIX ***(top)*** **has** two phosphate **backbones (red** ***and yellow spheres)*** **wrapped around a stack of base pairs (gray,** ***blue, red and white spheres).*** **A slice through one frame of a DNA simulation in water** ***(bottom)*** **shows how water molecules (green and** ***white Vs)*** **can penetrate deeply into DNA's double-helical structure and stabilize it. Purple spheres indicate sodium ions in the water solution.**



ALPHA HELIX, such as those found in most proteins, can unfold more easily in water because water molecules can replace the hydrogen bonds (green lines) that normally hold the helix together. A water molecule (green) is shown bridging an oxygen atom (red) and a nitrogen-hydrogen group *(blue),* which would normally bond in the folded helix.

up the **backbone** of the DNA double helix caused the molecule to break up after only 50 picoseconds. In the late 1980s Levitt and Miriam Hirshberg of the National Institute for Medical Research in London succeeded in making a 500-picosecond model of DNA by including water molecules that stabilized the double-helical structure by forming hydrogen bonds with the phosphate groups. Subsequent simulations of DNA in water have revealed that water molecules are able to interact with nearly every part of DNA's double helix, including the base pairs that constitute the genetic code.

In contrast, water is not able to penetrate deeply into the structures of proteins, whose hydrophobic regions are tucked on the inside into a close-fitting core. Accordingly, protein-water simulations have focused on the protein surface, which is much less tightly packed than the protein interior.

The way in which water molecules interact with the surfaces of proteins results in much interesting geometry-particularly in the deep grooves on the surfaces of enzymes, proteins that foster chemical reactions in cells. Hydrogen-bonded water molecules have difficulty fitting into these grooves and are easily displaced by ligands-the molecules with which an enzyme is intended to interact-which might explain why the active sites of enzymes frequently occur in grooves. We often find that the arrangement of water molecules in an empty active site mimics the geometry and structure of the actual ligand, knowledge that is sometimes used in drug design.

Living in the Real World

How closely do simulations of biological molecules in water resemble reality? Unfortunately, we cannot answer this question definitively, because no experimental technique can provide as much detailed information about individual molecules and their interactions as computer modeling can. What we can do is to compare various aggregated and averaged values derived from simulations with experimental results.

One of most important approaches that can be used to verify the structures of biological molecules simulated in water is neutron and x-ray scattering. In a neutron-scattering experiment, we direct a beam of neutrons at a small sample and record how the neutrons are scattered by the molecules that make up the sample. Each space between the molecules acts as a tiny slit, yielding a characteristic diffraction pattern. By analyzing these patterns, we can readily determine the spacing between the various molecules. When we compare neutron-scattering results with computer simulations, we find that, on average, these distances coincide.

To confirm the dynamics of a molecular simulation, we compare the predicted behavior of the simulated biological molecule in water with its observed properties in the laboratory. For example, most proteins contain at least one alpha helix, where the amino acids that make up the protein twist to form a coil. We know from experiments that heat causes these alpha helices to uncurl, yet in early attempts to simulate the behavior of a simple alpha helix in vacuo at elevated temperatures, the helix remained intact. Only by adding water to the simulation were Levitt and Valerie Daggett of the University of Washington able to mimic an alpha helix's actual behavior [see ***illustration above].***

Such computer simulations are yielding more and more information about the shapes of various biological molecules and how they perform their jobs in a living organism. We are, however, constantly running up against the limitations of computing technology and the cost of supercomputer time as we seek to conduct simulations of increasingly complex biological molecules in their watery environments. When scientists publish models of biological molecules in journals, they usually draw their models in bright colors and place them against a plain, black background. We now know that the background in which these molecules exist-water-is just as important as they are. ❑

The Authors

MARK GERSTEIN and MICHAEL LEVITT have collaborated since 1993, when Gerstein became a postdoctoral fellow in Stanford University's department of structural biology, which Levitt still chairs. Levitt obtained his Ph.D. in 1971 from the University of Cambridge. He has held academic positions at the Laboratory of Molecular Biology in Cambridge, England, the Salk Institute for Biological Studies in San Diego and the Weizmann Institute of Science in Rehovot, Israel. A frequent consultant for pharmaceutical companies, Levitt also founded the biotechnology company Molecular Applications Group in Palo Alto, Calif. Gerstein is an assistant professor at Yale University. He completed his Ph.D. at Cambridge in 1993.

Further Reading

WATER: NOW YOU SEE IT, NOW YOU DON'T. Michael Levitt and Britt H. Park in ***Structure,*** Vol. 1, No. 4, pages 223-226; December 15, 1993.

PACKING AT THE PROTEIN-WATER INTERFACE. Mark Gerstein and Cyrus Chothia in ***Proceedings of the National Academy of Sciences USA,*** Vol. 93, No. 19, pages 10167-10172; September 17, 1996.

For electronic archives of molecular structures, visit bioinfo.mbb.yale.edu and hyper.stanford.edu on the World Wide Web.