

Chemistry Nobel honors computer simulation of biomolecules

Experiments alone aren't always enough to understand why proteins and other large molecules do what they do.

“For the development of multiscale models for complex chemical systems,” as the Royal Swedish Academy of Sciences put it, the 2013 Nobel Prize in Chemistry has been awarded in equal shares to Martin Karplus of Harvard University and the University of Strasbourg in France, Michael Levitt of Stanford University, and Arieh Warshel of the University of Southern California. Throughout fruitful and intertwined careers, the three laureates laid the groundwork for the computational study of biomolecular structures, dynamics, and chemical reactions.

This year's chemistry prize is the first in 15 years to be awarded for theoretical work. The 1998 prize went to Walter Kohn and John Pople for their contributions to the methods of quantum chemistry (see *PHYSICS TODAY*, December 1998, page 20). Kohn and Pople's work dealt with ways to calculate the electronic structures of atoms, molecules, and condensed-matter systems without solving the full Schrödinger equation. In some respects, the work honored by this year's prize builds on those methods. But in others, it deals with ways to understand molecules without explicitly simulating their electrons at all.

A consistent force field

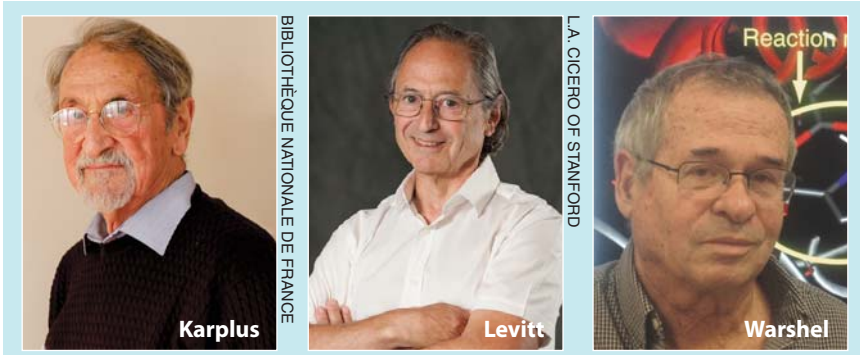
In the fall of 1967, Levitt and Warshel were working together in Shneior Lifson's group at the Weizmann Institute of Science in Rehovot, Israel. Warshel was a PhD student, and Levitt had just arrived on an exchange fellowship after finishing his bachelor's degree at King's College London. Their project was to write a computer program to calculate the energy of a molecular structure by

modeling it as a network of balls and springs.

Spring constants for chemical bonds could be derived from experimental data—vibrational spectra, say—and could be applied to similar bonds in other molecules. Bond angles, too, oscillate harmonically about their equilibrium values in a way that could be empirically parameterized. Torsion angles, describing the three-dimensional arrangement of atoms, and interactions of nonbonded atoms also contribute to the molecular energy. The ingredients of the calculation are summarized in figure 1.

Such ball-and-spring models, now known as “force fields” or “molecular mechanics,” were not new. But they were limited in their applicability. A force-field model that was devised to explain the vibrational spectra of a family of similar molecules would be inapplicable to other molecules, and it would be unable to predict equilibrium molecular structures or other properties. Lifson's idea, to be implemented with Warshel and Levitt, was to use a single model with a single set of parameters to calculate all relevant properties for any molecule. Thus was born the idea of a consistent force field, or CFF.

The key to finding such a general set of parameters was to get the nonbonded-atom interactions just right. Simply including a Lennard-Jones potential to describe the van der Waals forces that arise between all neutral atoms and molecules yielded poor agreement with experiment. In fact, the atoms in a polyatomic molecule aren't quite neutral, because the electrons in a covalent bond between two nonidentical



Laboratory Cryostats

3 K — 800 K



Superconductivity
Optical Spectroscopy
Cryogen Free



Sample in Vapor
Top Loading
Fast Sample
Change



Micro Raman, Micro-PL
Ultra Low Vibration
3 - 5 nm



Cryogenic
Probe Station
Helium Free
Fast Cool-Down

Advanced Research Systems

Tel: +1 610 967 2120

Fax: +1 610 967 2395

www.arscryo.com * ars@arscryo.com

atoms are never quite symmetrically distributed. Even in a carbon–hydrogen bond, generally regarded as nonpolar, the electrons are slightly skewed toward the carbon atom. Accounting for that asymmetry by assigning a small partial charge to each atom improved the model immensely.¹

Warshel and Levitt made two versions of the CFF program. The first, which calculated the energy of a molecular structure and the first and second derivatives of that energy with respect to atomic positions, could predict the structures and vibrational frequencies of moderately sized organic molecules. The second version omitted the second-derivative calculations, so it was fast enough to be applied to larger molecules such as proteins. Levitt used that version to perform protein structure refinements: Starting from a poorly resolved structure derived from x-ray diffraction measurements, he found a nearby structure that the program predicted to be an energy minimum.²

Retinal

As flexible as the CFF program was, it had its limitations. It was unable to describe processes in which bonds break or form or in which charges move from one atom to another. That meant, of course, that it couldn't model chemical reactions. But it was also unable to describe the internal motions of conjugated organic molecules, whose long chains of alternating single and double bonds cause some electrons to become delocalized over a large area of the molecule. Because the deformation of one bond could cause those so-called π electrons to move to a distant bond, conjugated molecules aren't accurately described as systems of masses and independent springs.

In 1969 Karplus was interested in just such a conjugated molecule called retinal, part of the pigment rhodopsin that's present in the rod cells of the retina. Retinal initiates the vision process by absorbing light and changing its shape. Its chemical formula and bond connectivity were known, but its 3D structure was not, and there did not yet exist a quantitative model to compare the energies of a conjugated molecule's different conformations.

Karplus and his postdoc Barry Honig developed a simple model for retinal.³ By treating the π electrons with a rudimentary quantum method and the rest of the bonds with an empirical potential-energy function, they analyzed the changes in energy associated

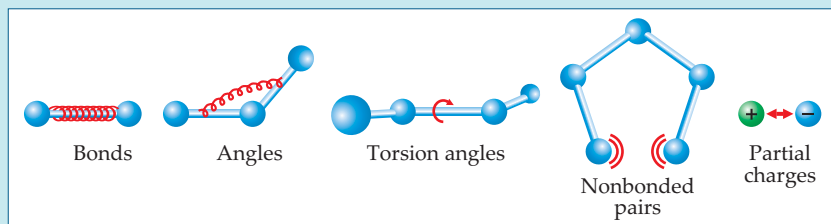


Figure 1. A molecular mechanics model, such as the consistent force field program or CHARMM, evaluates the energy of a molecule using simple empirical potential functions to account for bond stretching, bond angle bending, bond twisting, van der Waals interactions between nonbonded atoms, and electrostatic interactions between partially charged atoms.

with changes in the molecule's torsion angles. The energy-minimizing 3D structure they predicted was later shown to be correct.

The following year Warshel joined Karplus's group as a postdoc, and he brought the CFF program with him. He extended it to model conjugated molecules by treating the π electrons with a version of the Pariser-Parr-Pople method, a semiempirical quantum method developed independently by Pople and by the team of Robert Parr and Rudolph Pariser. Warshel and Karplus were thus able to calculate the energies associated not only with changes in torsion angles but also with changes in bond lengths and bond angles and with vibrational and electronic excitations.⁴

Proteins in space

In 1972 Levitt and Warshel began to collaborate again: first at Weizmann, where Levitt was a postdoc and Warshel was a new faculty member, and then also at Cambridge University, where Levitt joined the faculty and Warshel went to visit. They began to think about protein folding, which was and still is a major challenge in molecular biology. How can a linear chain of amino acids fold into a specific, functional 3D shape on a reasonable time scale? As Cyrus Levinthal had recently noted in what's now known as the Levinthal paradox, even a small protein has so many degrees of freedom that if it were to fold by randomly exploring its configuration space, the process would take millions of years.

At the time, the way to visualize and manipulate protein structures was to build cumbersome physical models out of balls and sticks. Inspired by space exploration, which was much in the news, Levitt and Warshel began to speculate that protein models would be much easier to manipulate on a spacecraft, unencumbered by gravity. "Then we realized that we could simulate that," Levitt recalls.

Tracking the positions of every atom in a folding protein seemed like an unnecessary strain on their computing resources. The relative positions of the atoms in an amino acid don't change much; the protein conformation is largely determined by the torsion angles about the carbon–nitrogen bonds between successive amino acids. So Levitt and Warshel took the simplifying step of replacing each amino-acid side chain with a "dummy atom" that interacts with other dummy atoms via a modified van der Waals potential.

To test their method, they looked at the folding of bovine pancreatic trypsin inhibitor (BPTI), chosen not for its biological interest but for its small size (just 58 amino acids and some 900 atoms) and its relatively well-studied x-ray crystal structure. They found that the BPTI sequence of amino acids, starting from a nearly straight chain, would often fold into a structure very similar to the true BPTI structure.⁵ One such folding sequence is shown in figure 2.

The dummy-atom approximation, later refined by Levitt and others, forms the basis for coarse-grained protein models that are used today.

Exploring enzymes

Warshel and Levitt's 1970s collaborations also looked at enzyme catalysis. It had been a longstanding interest of Warshel's: What is it about certain proteins that allows them to accelerate chemical reactions by up to 20 orders of magnitude? One of Levitt's first protein-structure refinements was on lysozyme, an enzyme that's part of the immune system. Lysozyme helps to kill bacteria by breaking up the polymers of sugar and amino acid, called peptidoglycans, that make up their cell walls. David Chilton Phillips, who'd published the x-ray structure of lysozyme in 1965, also proposed a mechanism: Lysozyme acts like a medieval rack, he said, physically stretching peptidoglycan bonds until they break. But Levitt's work with CFF showed that the enzyme was nowhere

near rigid enough to apply the necessary bond-breaking forces. "Imagine a rack made of foam rubber," he says. "It doesn't work."

How could a simulation help to reveal the right mechanism? Quantum mechanics was clearly needed to model the reaction itself, but even an approximate quantum calculation on the entire enzyme would take far too long. Warshel and Karplus's combination of quantum and molecular mechanical methods turned out to be key: Quantum methods could treat the core reaction region, while the rest of the system, where no bonds break or form, could be treated with molecular mechanics.

Warshel set out to implement the idea: He used CFF to model the periphery of the system and a semiempirical quantum approach on the core, including the mechanical forces that the periphery exerts on the core. "I wrote a very nice program that I used only once," he says. "I couldn't get the bonds to break" in the simulated peptidoglycan.

The first program's failure offered the necessary insight: The interaction between core and periphery was not, in fact, merely mechanical. There were electrostatic forces in play too, exerted not just by the outer reaches of the enzyme but also by the surrounding water molecules. Breaking a chemical bond often imbues the formerly bonded atoms with large partial charges, which need to be stabilized by opposite surrounding charges; otherwise, the energy barrier to reaction is too high. Once Warshel had included the electrostatic effects in his program, the simulated bonds broke easily.

The method, now called QM/MM, that was laid out in the ensuing 1976 paper⁶ has become a powerful tool for studying enzyme catalysis—now that computers are fast enough to treat the reaction region with more accurate quantum methods, such as Kohn's density functional theory. Warshel and Levitt's paper has racked up more than 2000 citations, almost all of them since 1995.

CHARMM

Meanwhile at Harvard, Karplus was also becoming interested in atomic simulations of proteins. His student Bruce Gelin, just out of the US Army, wrote a program based on CFF for calculating energies of macromolecular configurations.⁷ That program would eventually develop into CHARMM—for Chemistry at Harvard Macromolecular Mechanics—one of the most widely dis-

tributed molecular mechanics programs today.⁸

As CHARMM evolved under Karplus's direction, new features were added. Whereas CFF could calculate the forces on atoms in a particular molecular configuration, Gelin and Andrew McCammon added the ability to track the trajectories of atoms under those forces—that is, to simulate macromolecular dynamics. Pioneered by Aneesur Rahman and others for simple condensed-matter systems such as water,⁹ molecular dynamics simulations aim to realistically reproduce the motions of interacting atoms or molecules. The goal might be to simulate a time-dependent process, or it could be simply to explore the system's accessible configuration space at a particular temperature.

In their initial foray into protein molecular dynamics, McCammon, Gelin, and Karplus simulated an isolated BPTI molecule for 9.2 ps, broken down into femtosecond time steps.¹⁰ The calculation revealed that the protein's atoms moved around much more than was evident from the static x-ray crystal structure. Molecular dynamics simulations continue to explore the nature of those fluctuations in various systems: Is a system fluctuating about a single energy minimum, or is it hopping among several local minima? To what extent are the fluctuations involved in the protein's function—for example, an enzyme's ability to catalyze a reaction? (See the article by Karplus, *PHYSICS TODAY*, October 1987, page 68.)

In recent decades, it's become possible to simulate larger systems for longer times. Instead of picoseconds, simulations model processes that last for nanoseconds or microseconds. The computational simplifications that were essential in the 1970s—such as QM/MM, which was incorporated¹¹ into CHARMM in 1987—are still useful today. "It's possible to calculate the electronic structure of an entire protein," says Adrian Mulholland of the University of Bristol in the UK, "but that's not necessarily the best use of your computing time."

The search for still more efficient computational methods continues. As Karplus points out, in a simulation of the enzyme ATPase in solvent, 95% of the atoms belong to water molecules.

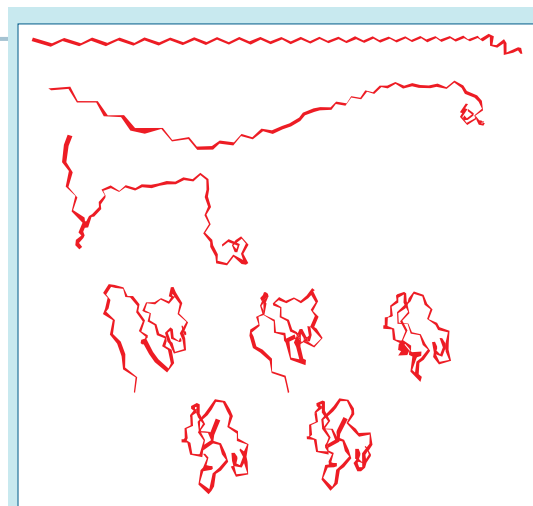


Figure 2. Simulated folding of bovine pancreatic trypsin inhibitor. Each of the protein's 58 amino-acid side chains was approximated as a dummy atom (not shown) that dangled off the main backbone. (Adapted from ref. 5.)

"So most of the calculation is spent simulating water," he says. "People are looking for ways to do that faster."

And Levitt and his colleagues just recently found a way to calculate the *R*-factor, which quantifies the agreement between a model protein structure and x-ray diffraction data, thousands of times faster than before. "Algorithms are funny things," he says. "There's no real theory that tells you when your algorithm is best. I suspect there's a lot of room for improvement."

Full careers

Martin Karplus was born in March 1930 in Vienna; he and his family left for the US in 1938. In 1953 he earned his PhD from Caltech, where he worked under Linus Pauling, and spent the next two years at Oxford University as a postdoc under Charles Coulson. In 1955 he joined the faculty at the University of Illinois, where he worked on the theory of nuclear magnetic resonance coupling constants and developed the equation that bears his name. Five years later he moved to Columbia University, and in 1966 he moved to Harvard University, where he remains today. He is also director of the Biophysical Chemistry Laboratory at the University of Strasbourg.

Michael Levitt was born in May 1947 in Pretoria, South Africa, and earned his PhD from the MRC Laboratory of Molecular Biology at Cambridge in 1971. He spent two years as a postdoc at the Weizmann Institute of Science, then returned to the MRC as a staff scientist. From 1977 to 1979 he worked with Francis Crick at the Salk Institute for Biological Studies in La Jolla, California. After

eight more years at Weizmann, he went to Stanford in 1987.

Arieh Warshel was born in November 1940 in kibbutz Sde Nahum in what is now Israel. Working in Lifson's group, he earned his PhD from Weizmann in 1969. After two years as a post-doc in Karplus's group, he returned to Weizmann to join the faculty. He has been at the University of Southern California since 1976.

Johanna Miller

References

1. S. Lifson, A. Warshel, *J. Chem. Phys.* **49**, 5116 (1968).
2. M. Levitt, S. Lifson, *J. Mol. Biol.* **46**, 269 (1969).
3. B. Honig, M. Karplus, *Nature* **229**, 558 (1971).
4. A. Warshel, M. Karplus, *J. Am. Chem. Soc.* **94**, 5612 (1972); A. Warshel, M. Karplus, *J. Am. Chem. Soc.* **96**, 5677 (1974).
5. M. Levitt, A. Warshel, *Nature* **253**, 694 (1975).
6. A. Warshel, M. Levitt, *J. Mol. Biol.* **103**, 227 (1976).
7. B. R. Gelin, M. Karplus, *Proc. Natl. Acad. Sci. USA* **72**, 2002 (1975).
8. B. R. Brooks et al., *J. Comput. Chem.* **4**, 187 (1983); B. R. Brooks et al., *J. Comput. Chem.* **30**, 1545 (2009).
9. A. Rahman, *Phys. Rev.* **136**, A405 (1964); F. H. Stillinger, A. Rahman, *J. Chem. Phys.* **60**, 1545 (1974).
10. J. A. McCammon, B. R. Gelin, M. Karplus, *Nature* **267**, 585 (1977).
11. P. A. Bash, M. J. Field, M. Karplus, *J. Am. Chem. Soc.* **109**, 8092 (1987); M. J. Field, P. A. Bash, M. Karplus, *J. Comput. Chem.* **11**, 700 (1990).

"Melting" ice yields hints of a second liquid water phase

New experimental findings bolster a controversial theory of supercooled water.

Water can be peculiar at times. At temperatures near freezing, it expands as it's cooled. It can become less viscous as it's compressed. And, according to a theory proposed two decades ago by researchers at Boston University, it can exhibit yet another striking behavior: It can form two liquid phases.¹ (See the article by Pablo Debenedetti and Gene Stanley, *PHYSICS TODAY*, June 2003, page 40.)

The two-liquid hypothesis was inspired by molecular-dynamics simulations that showed supercooled water forming low- and high-density liquids (LDL and HDL) at temperatures below about 220 K. In LDL, water appears to form a hydrogen-bond network similar to that of common ice, with each molecule sitting inside a tetrahedral cage formed by its four nearest neighbors. In HDL, up to two extra molecules squeeze into the interstices of each tetrahedral cage.

The hypothesis has its detractors. David Chandler (University of California, Berkeley) and coworkers, for instance, argue that only one of the liquids can be a true phase corresponding to a metastable equilibrium; the other is likely a nonequilibrium transient state that forms as water freezes to crystalline ice.² That view is backed by numerical work suggesting that HDL and LDL don't form stable interfaces, which are hallmarks of a two-phase system.³

Now a collaboration led by Thomas Loerting (University of Innsbruck, Austria) and Roland Böhmer (Technical University of Dortmund, Germany) has weighed in on the matter with new experimental evidence.⁴ The researchers used a pair of independent techniques

to probe water's phase changes in the deeply supercooled regime. Although their findings don't close the book on the water controversy, they strengthen the case for two liquids.

A pair of glasses

The plausibility of the two-liquid theory stems in part from the well-known existence of two amorphous ices: low- and high-density amorphous ice (LDA and HDA), shown in figure 1. LDA can be made by rapidly cooling water droplets to liquid-nitrogen temperatures. HDA is typically made by chilling crystalline ice to liquid nitrogen temperatures and then compressing it to around 10 kbar.

So-called glassy states, amorphous ices have the mechanical properties of a solid but the disordered molecular configurations of a liquid. When heated above their respective glass-transition temperatures, they melt—or, more precisely, relax—to form ultraviscous supercooled liquids that, though metastable, can persist for hours. With respective densities of about 0.92 g/cm³ and 1.15 g/cm³, those liquids are thought by many to be the LDL and HDL posited by the two-liquid theory.

But are the liquids distinct phases? One way to test is to probe their phase-transition behavior. In a pure substance, a phase transition such as the glass transition has a single thermodynamic degree of freedom; fixing the pressure fixes the transition temperature. (Technically, the glass transition is a kinetic, not an equilibrium, transition, so its temperature at a fixed pressure may vary by a few K depending on the sample preparation.) If LDA and HDA



Figure 1. Amorphous ice can exist in high-density (**top**) and low-density (**bottom**) states, depending on the pressure at which it forms. The two samples shown here, inside 7-mm-diameter rings, have identical mass. (Images courtesy of Osamu Mishima.)

were to exhibit vastly different glass-transition temperatures at a single pressure, it would serve as a clue that the ices—and the low- and high-density liquids they form when they relax—are also distinct.

Hot and cold

Observing both glass transitions at common pressure turns out to be technically complicated. HDA is fairly unstable at low pressures, and LDA is unstable at high pressures. So although LDA's glass transition is straightforward to detect at say, ambient pressure, HDA under the same conditions tends to transform to LDA before becoming warm enough to relax.

One way to improve a glass's stability is to anneal it—to maintain it for an