

Computational Chemistry: Many Roads to Travel

Donald G. Truhlar

Travel brochures often take a dualistic approach in explaining the attractiveness of some particular corner of the world: You want culture? You want beaches? Come to our place; we have both. In explaining to the general computational scientist the attractions of computational chemistry, we can take a similar dual tack: whether you are seeking basic research on fundamental issues or seeking critical applied areas such as industrial competitiveness, energy, the global environment, and human health—come to our place; we have both.

Although computational chemistry addresses a broad range of questions and uses a broad range of techniques, its foundation rests solidly on a single partial differential equation: the molecular Schrödinger equation. Even in simulations where all the work goes into solving Newton's, Lagrange's, or Hamilton's classical equations of motion, the quantum-mechanical description of atomic and molecular forces and motions lurks in the background as the true description of reality. Any classical method must be validated against it. This can be done either by actual numerical comparisons for similar systems, or by a general argument that the classical equations provide a valid asymptotic approximation to quantum mechanics for the question at hand. This long validation process has already started and is progressing.

The quantum-mechanical description of molecules and other chemical materials is almost always carried out in two steps, a partition usually called the Born-Oppenheimer separation of electronic and nuclear motions. The

solution to step one, the electronic motion, is the electronic energy as a function of nuclear coordinates. This solution, plus the internuclear coulombic energy, provides a potential-energy surface or force field for step two, the nuclear motion. Sometimes both steps require very sophisticated techniques; at other times the difficult part lies in only one step, and the other can be relegated to a few equations that could fit on the back of a floppy disk wrapper.

These steps are so independent of one another that they are sometimes done in different years on different continents by different people who have never met. But a problem isn't fully solved until both steps are carried out. For example,

understanding high-temperature superconductivity is typically thought of as an electronic-structure problem, but electrical resistance is intimately connected to electron scattering by nuclei, and we can't fully understand conductivity without understanding nuclear motions. Similarly, protein folding seems

to be a problem in nuclear motion, but that nuclear motion is driven in part by hydrogen bonding forces whose precise strength is critical. The precise strength of hydrogen bonds in crystals, though, depends on how the electronic orbitals relax in the field of many surrounding dipoles whose orientation is a consequence of the nuclear coordinates. Still, in most work, computations of the two steps are done separately, and often the error due to this separation of effort is negligible compared to all the other approximations that must be made.

The electronic-structure problem determines the energy of a molecule for a given nuclear geometry. Many of the challenges

in electronic structure involve complex optimization. Finding the "optimum" structure of an isolated molecule means finding the lowest-energy nuclear geometry—difficult because of the many dimensions (for medium and large molecules) and the complicated energy surface, which typically has many local minima. Finding the transition state for a chemical reaction means finding a particular saddle point on that surface. The inputs to such calculations are electronic energies, gradients, and Hessians (matrices of second derivatives). Finding these involves optimizing the coefficients in a linear combination of symmetry-adapted many-electron trial functions. To make this manageable we optimize the space they span. To reduce the size of this space we optimize the single-electron basis functions out of which the many-electron ones are made.

Sometimes we use approximations or alternative formulations. One such alternative, the density-functional method, has developed rapidly of late. Instead of explicitly solving for the dynamic coulomb interactions in each possible many-electron wave function, we try to represent the effect of electron correlation by an effective potential dependent on local properties, such as the average electron density and its gradient. Another class of approximate methods is semiempirical molecular orbital theory, in which the integrals in a highly approximate (and thus relatively inexpensive) electronic structure calculation are replaced by empirical parameters. In yet another, usually called molecular mechanics, the potential surface is directly approximated as a sum of transferable terms corresponding, for example, to bond-angle restoring forces, pairwise interactions, and so forth. Determining the optimum parameters of such models again involves a large-scale optimization effort.

The difficulty in electronic structure calculations is adequate representation of electron correlation effects. Many techniques for these calculations have an operation count that scales as N^3 , where

*The foundation of
computational chemistry
rests solidly on one partial
differential equation.*

IEEE CS&E invites
article submissions on
computational
chemistry. Contact the
editor-in-chief's office
(see masthead, p. 3).

$n \approx 5-7$, depending on the method (typically n is higher for more accurate methods), and N is the number of atoms or basis functions. A goal of much present work is finding practical and accurate methods with small n , even with $n = 1$. Successful $n = 1$ methods will be here soon for some problems, but it is not clear yet if generally accurate methods can be devised with low n . Formulating new electron-correlation approaches with better scaling will be an active field of research in the coming years.

One area where we may see especially important results is molecular design, particularly of new catalysts, drugs, and nonlinear optical materials. Many years ago, *ab initio* techniques (literally "from the beginning," or calculated from basic principles) began to compete with experimentation for determining the thermodynamic and spectroscopic properties of small molecules. For example, heats of formation and vibrational spectra of unstable species are now much easier to calculate than to measure, and we can expect such success to extend to larger molecules, more excited states, and more subtle weak interactions.

A particularly promising area for computational chemistry is the rational design of new drugs. Synthesis and biological testing of trial drugs is expensive and inefficient, but computed descriptors can predict the properties and activity of compounds that have never been made. So far this research is still focused on the electronic-structure problem. I envision tremendous strides being made as accurate techniques validate both semiempirical and low-level *ab initio* schemes for structural calculations on larger and more relevant biomolecules. Furthermore, the molecules and the specificity of their interactions with receptors are now being simulated in such a way as to include the effects of finite-temperature environments, and the techniques for this can also be expected to blossom through more and better research.

Computer-aided drug design is just one corner of the emerging world of molecular and materials design. Other design targets include catalysts; polymers, conducting and structural materials (including steels and ceramic and zeolitic nanostructures), and biomimetics. In all these areas semiempirical models will probably be the workhorses, but high-quality results will be more likely if small prototypes of the models are prevalidated against more

computationally demanding and systematically improvable *ab initio* techniques.

The dynamics side of computational chemistry builds on the structural advances. Dynamics is crucial in atmospheric modeling, where key reaction steps in the complex atmospheric stew are often more accessible to theoretical modeling than to isolation in the laboratory. Combustion chemistry, another area with a complex mixture of reaction intermediates, provides a classic success story of the power of computational chemistry. The US Department of Energy has spearheaded an amazingly successful computational attack on understanding the reactive intermediates and mechanisms of very complex combustion systems. Sooty flames yield to computer analysis better than to lasers.

In addition to progress on applied problems, fundamental issues in dynamics are also yielding to computational attack. I expect to see increasing attention paid to such issues arising in the dynamics of condensed-phase processes and electronically excited reagents.

Computational dynamics is also giving us unprecedented insights into the details of reactive events at the molecular level. For over 60 years the unifying concept of chemical reactivity has been the paradigm of a barrier crossing, with atoms traversing the highest-energy point on the lowest-energy path—the saddle point—from the reactants arrangement to the products arrangement. Modern work has successively refined this model. First, we recognize that entropic factors as well as energetic ones determine the molecular geometry of the dynamical bottlenecks. Second, we recognize that tunneling through (rather than passing over) barriers is the dominant mechanism in many cases, and we have developed practical ways to compute tunneling probabilities for realistic systems. Third, accurate solutions of the Schrödinger equations have enabled us to "see" and characterize the individual short-lived quantum states associated with such reactive passages, with computations again preceding experiment.

Accurately solving the Schrödinger

equation returns us to the theme of this magazine. Cross-fertilization from partial differential equation solution techniques developed in engineering fields and applied mathematics may play a powerful role in future solution methods for the Schrödinger equation. For example, a recent application of the preconditioned GMRes method, widely used in other fields for iterative solutions of sparse equations, gave very encouraging

results for solving the dense equations of quantum-mechanical reactive scattering.

Molecular dynamics simulations of large systems over extended times at finite temperature are extremely computationally demanding, and hence this is another area where new, general

techniques could prove to be very useful—for example, the search for more efficient, but yet stable and reversible, integrators. Also ripe for improvement are procedures for sampling; the Metropolis algorithm has been so successful for sampling mechanical and thermal properties for both equilibrium systems and quasi-equilibrium treatments of dynamics that few alternatives have received serious study. Recently, though, much progress has been made in designing new statistical-mechanical ensembles and novel biased trial moves for complex fluids, and also in combining Monte Carlo with other techniques. Perhaps a breakthrough is possible here, though it is dangerous to speculate on ideas that haven't been thought up yet.

Large-scale optimization, solutions of partial differential equations, sampling of rugged landscapes—these are very general problems, broader than computational chemistry, yet too often they are attacked by independent, noncommunicating threads of activity in separate fields. I hope, by reading *IEEE Computational Science & Engineering*, to learn more about what my colleagues in other fields are doing on these problems and try to apply it in chemistry. I hope my fellow chemists will also explain their techniques in these pages. Just maybe that will be useful to nonchemists as well. ♦

**Sooty flames
yield to computer
analysis better than
to lasers.**

Donald G. Truhlar is area editor for computational chemistry; e-mail truhlar@t1.chem.umn.edu.

