Modeling—A Tool for Experimentalists

Romas Kazlauskas

What does a molecule look like and what can it do? Chemists have answered such questions with experiments and, increasingly, with theoretical methods. Molecular modeling is a theoretical method that comprises a broad range of computer methods which allow chemists to display molecules, predict their structures, make short movies of their motions, predict how they bind to each other and react with each other (*1*). As this method becomes more routine and more reliable, experimentalists are using it more frequently to guide and improve experiments and to construct solutions to questions that are impossible to examine experimentally.

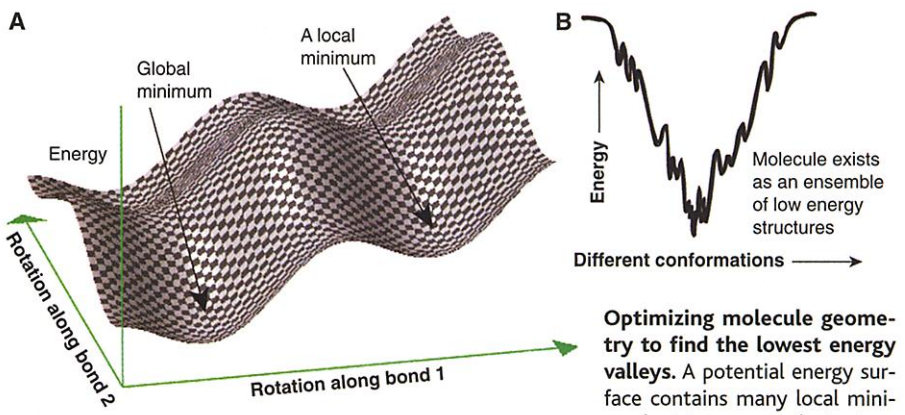
The simplest application of modeling is molecular visualization, which is the use of computers to display molecular structures measured experimentally. Not only are they easier to build and store than plastic models, computer models can both simplify and highlight molecular features. For example, a schematic tracing of a protein chain simplifies a complex protein structure, whereas a space-filling representation of the binding site emphasizes its shape and a stick representation of a bound molecule emphasizes its chemical structure. Each type of image serves a different purpose.

All modeling involves approximations, even the most advanced theoretical methods. The key to effective modeling is to include enough detail in the model to accurately describe the phenomenon in question, but to omit details that waste computer time or add needless complexity that doesn't add to the information obtained. With a good model, the simplest questions can be answered by displaying the structure of the molecules involved. More complex problems, however, require the simulation of a process, such as molecular motion. Quantitative predictions of binding or reactivity usually cannot be obtained from a single structure and may require examination of hundreds or thousands of low-energy structures. Further, chemical reactions involve bond making or breaking, which can only be modeled using quantum mechanics.

Modeling starts with a crude sketch of the molecule that is then optimized geometrically. In this process, the computer attempts to assign an optimal orientation to each atom. To accomplish this, an energy cost is assigned to each nonoptimal orientation, such as a compressed bond angle or an interaction of two atoms bumping against each other. The total energy of the molecule, or its strain, is the sum of all these costs. The computer program iteratively adjusts the bond lengths and angles to lower the total energy. When structure adjustments cannot lower the total energy any further, the structure has reached an energy minimum and the geometry optimization of the model is complete.

Structure predictions usually use molecular mechanics, where equations from classical physics and empirically derived parameters describe molecular bonding. This hybrid approach to molecular structure quickly yields accurate three-dimensional (3D) structures. The parameters for each molecular fragment include atom size, optimal bond lengths, and optimal 3D orientation. This approach assumes that the molecular building blocks, known as atom types, in one molecule interact the same way when they are present in a different molecule. For organic molecules, structures predicted by molecular mechanics agree well with experimental measurements. Determinations of protein structure by x-ray crystallography or by nuclear magnetic resonance (NMR), for example, use molecular modeling to optimize the fit of structures to the experimental data (*2*).

One challenge in geometry optimization is finding the most stable structure, the global minimum. Molecules adopt many stable structures or local minima, which differ by rotations about single bonds (see figure below). Geometry optimization finds minima by making small changes in theoretical structures to yield the closest stable structure only, which is not necessarily the most stable structure. Researchers use additional methods to make larger changes in structure that can cross an energy barrier and find the global minimum. For small molecules, a systematic search of the possible rotations about single bonds reveals the global minimum. For larger molecules or biomolecules, the number of bond rotations required for a systematic search is enormous, rendering such a search impractical. For these molecules, researchers use either a random search or molecular dynamics.



Optimizing molecule geometry to find the lowest energy valleys. A potential energy surface contains many local minima (stable structures), the lowest of which is the global minimum (the most stable structure). The minima differ by rotations about single bonds. The potential energy surface shown here illustrates two degrees of freedom, but real molecules may contain degrees of freedom on the order of tens to thousands.

Molecular dynamics involves the creation of a very short movie of molecular motions. The researcher chooses the degree of motion by setting the temperature. The atoms move as a function of both molecular mechanics forces and of their motion in the previous step. These movies show a chaotic jiggling of atoms and conformational changes of the molecule. Cooperative motions are included because the forces of molecular mechanics allow one group to avoid collision with another. By checking the energy of new conformations created by molecular dynamics, researchers can find additional minima, including the global minimum.

Even with these tools, some structures are too complex to find

The author is in the Department of Chemistry, McGill University, Montréal, Québec H3A 2K6, Canada. E-mail: romas.kazlauskas@mcgill.ca

the global minimum. For example, conformational searching cannot predict a protein's 3D structure from its amino acid sequences because the number of possibilities overwhelms today's computers (the protein folding problem).

Another example where one searches for the best structure is docking, which analyzes how, say, a drug candidate binds to a protein target. Lam *et al.*, for example, used docking to test possible orientations of drug candidates during their development of an HIV protease inhibitor (*3*).

Modeling aids the interpretation of experimental results by explaining why one enzyme is more selective than another. The x-ray structure of a nonselective lipase, for example, shows a wide substrate-binding pocket, but that of a selective lipase depicts a pocket twisted to one side (see figure at right) (*4*). Modeling shows which substrate would fit snugly into the selective lipase but not the nonselective one. In other cases, deciding whether a substrate "fits" requires a more careful look at the structural details. For example, Schulz *et al.* correlated the observed reactivity of a substrate with the length of a key hydrogen bond between the substrate and the enzyme (*5*).

Using modeling alone in protein engineering to predict mutations that would alter enzyme function is difficult because the mutation must not disrupt enzyme stability or catalysis. But experimental methods, like random mutagenesis and screening, suffer from the need to test enormous numbers of mutants. An alternative is to combine modeling and experimentation to identify which mutations cause desired changes but retain catalytic activity and stability. Li *et al.* used this approach to extend the substrate range of a P450 enzyme to shorter substrates (*6*). Whereas the wild type showed no activity on fatty acid analogs with chain lengths shorter than C_{10}, a variant with five mutations catalyzed efficient hydroxylation of a shorter fatty acid analog, C_8. Horsman *et al.* used this approach to dramatically increase the enantioselectivity of an esterase (*7*) from 12:1 to 60:1 by changing a single amino acid.

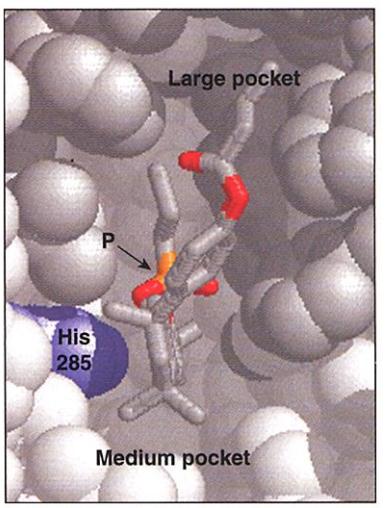
To make quantitative predictions with the use of modeling, one usually needs to include many low-energy structures in prediction, not just one. Molecules do not exist as a single structure but rather as an ensemble of rapidly equilibrating structures, all of which contribute to an aggregate behavior. Another way to think about the importance considering of many structures instead of one is to recall that rates of reaction and selectivities depend on free energies. Free energies consist of enthalpy (heat content) and entropy (disorder). Therefore, one must include multiple structures in order to include entropy (*8*).

To predict the different binding of enantiomers to several hosts, Aerts generated 5000 different complexes using molecular dynamics and optimized each complex geometrically to generate two sets of compounds (*9*). He compared all the lowest energy compounds from one set with the all the lowest energy compounds from another to correctly predict which interaction is stronger on average. Aerts reasoned that each enantiomer would be solvated by the same amount such that the solvent contribution would cancel out, and could therefore be ignored.

In other cases, however, one must include solvent as a consideration. The strands in a DNA double helix modeled without solvent would separate in a molecular dynamics simulation, but when water molecules are included in the model as a solvent, the helices remain together. Similarly, modeling an α helix without water incorrectly predicts that it would remain stable upon heating, but modeling with water predicts the experimentally observed unfolding.

An exhaustive search of all possible conformations is not practical for many large biomolecules, quantitative predictions are difficult to make. Hæffner *et al.* (*10*) simplified the problem by focusing on a part of a structure (the reactive region of a substrate-enzyme complex) and restricting its energy calculation to a few residues. In this way, they obtained models that agreed well with experimental data.

Molecular mechanics cannot model all chemical reactivity accurately because it fails to take into account the wave nature of matter. Quantum mechanics can be used to help measure this, but the complexity of these calculations limits this approach to tens of atoms, or at most several hundred. To model a reaction, researchers use yet another combined approach: quantum mechanics to model the reacting portion and molecular mechanics to model the rest of the molecules and the solvent. With this approach, Eksterowicz and Houk (*11*) explained the selectivity of numerous important catalysts in organic synthesis. Monard and Merz (*12*) modeled the structure and the distribution of electrical charges in the active site of carbonic anhydrase; an approach using only molecular mechanics did not yield a stable model. Zheng *et al.* (*13*) used modeling to distinguish between two mechanisms where the experimental evidence was ambiguous.



Snug as a substrate in a lipase. The substrate-binding site of the lipase (shown in this space-filling representation) contains a large pocket that tilts to the right. This tilt matches the orientation of a large group in the substrate (represented with a stick illustration). Image created with the use of the program RasMol (www.umass.edu/microbio/rasmol/). Reprinted in part with permission from the *Journal of Organic Chemistry* **66**, 3041 (2001). Copyright 2001, American Chemical Society.

The most productive strategies in chemistry are a combination of experiment and modeling. Experiments are too slow and too expensive for exhaustive optimization. Modeling is not yet accurate enough or fast enough to consistently predict behavior; it is, however, increasingly able to predict or explain experimental results. Because of these improvements, coupled with easier modeling methods due to faster computers and better software, experimentalists increasingly use modeling to focus, guide, and interpret experiments. Modeling is becoming a tool for experimentalists, not just theoreticians.

References

1. A. R. Leach, *Molecular Modeling: Principles and Applications* (Longman, Essex, UK, 1996). For an online course and links to programs and other modeling resources, see (*14*).
2. G. Rhodes, *Crystallography Made Crystal Clear* (Academic Press, New York, ed. 2, 2000), chap. 7, pp. 146–151.
3. P. Y. S. Lam *et al.*, *Science* **263**, 380 (1994).
4. D. G. Gascoyne *et al.*, *J. Org. Chem.* **66**, 3041 (2001).
5. T. Schulz *et al.*, *Protein Sci.* **9**, 1053 (2000).
6. Q. S. Li *et al.*, *Biochim. Biophys. Acta* **1545**, 114 (2001).
7. G. Horsman *et al.*, in preparation.
8. P. A. Kollman *et al.*, *Acc. Chem. Res.* **33**, 889 (2000).
9. J. Aerts, *J. Comput. Chem.* **16**, 914 (1995).
10. F. Hæffner *et al.*, *Biophys. J.* **74**, 1251 (1998).
11. J. E. Eksterowicz *et al.*, *Chem. Rev.* **93**, 2439 (1993).
12. G. Monard *et al.*, *Acc. Chem. Res.* **32**, 904 (1999).
13. Y.-J. Zheng *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 432 (2001).
14. www.ch.ic.ac.uk/local/organic/mod