AIChE

Perspective

From Discovery to Data: What Must Happen for Molecular Simulation to Become a Mainstream Chemical Engineering Tool

Edward J. Maginn
Dept. of Chemical and Biomolecular Engineering, University of Notre Dame, Notre Dame, IN 46556

DOI 10.1002/aic.11932
Published online May 7, 2009 in Wiley InterScience (www.interscience.wiley.com).

Keywords: Monte Carlo, molecular dynamics, molecular simulation, molecular modeling, computational chemistry

Introduction

Thirty years ago, molecular modeling and simulation was practically unknown within the chemical engineering community. The field has grown since that time such that now nearly every academic department has one or more faculty members who carry out some sort of molecular-based modeling activity, AIChE meetings host topical conferences and dozens of sessions focused on molecular modeling and simulation, and various aspects of the field have been the subject of past *Perspectives* articles.[1–4] Molecular modeling and simulation is pervasive, but has it become a mainstream chemical engineering tool in the way that process simulations or finite element modeling are? In other words, if a mainstream engineering tool is defined as a method that is "useful, practical, and accessible" to a wide range of researchers, then does molecular modeling and simulation fit this definition? The answer depends on how you define molecular modeling and simulation.

What Are We Talking About?

In the broadest sense, molecular modeling and simulation can be defined as *the use of computational methods to describe the behavior of matter at the atomistic or molecular level*. There is a clear distinction between this and the familiar continuum-based modeling, in which atomic-level phenomena are neglected. A hierarchy of time- and length-scales exist between the continuum and atomic, however, and in fact the development of "multiscale" modeling tools that bridge the gap between atomic and continuum length scales is an active research area.[2] Figure 1 shows the relationship between some of the more common molecular modeling and simulation techniques, and the characteristic length scale associated with each method.

Correspondence concerning this article should be addressed to E.J. Maginn at ed@nd.edu.

Within this broad range of techniques, a pragmatic distinction can be made between *computational quantum chemistry* methods (on the far left end of Figure 1), *molecular simulation* methods such as molecular dynamics (MD) and Monte Carlo (MC), and *coarse-grained modeling* methods (the two categories on the right side of Figure 1). Computational quantum chemistry is based on quantum physics, and is applied to the electronic structure of atoms or molecules. Molecular simulation is based on classical physics and treats interactions between atoms with empirical potential functions (force fields). As a result, longer length and time scales can be accessed with molecular simulation, albeit with some loss in chemical realism. Coarse-grained methods enable even longer time and length scales to be probed by replacing some atoms with a mean field and/or by lumping collections of atoms into single effective sites.

Of all the techniques listed in Figure 1, only computational quantum chemistry has achieved a level of acceptance that makes it a mainstream chemical engineering tool. Molecular simulation methods such as MD and MC are not yet mainstream tools, even though these techniques are used by many chemical engineers and are capable of determining a much wider range of properties of interest to chemical engineers than quantum methods. Similarly, coarse-grained methods have wide applicability but have not been embraced as a mainstream tool. The reasons for this state of affairs—and ways to overcome the barriers preventing the broader use of MD and MC in chemical engineering—is the focus of this Perspective. Many of the conclusions also apply to coarse-grained methods, but these techniques will not be considered explicitly here.

The Emergence of Molecular Simulation in Research

There is no question that molecular simulation is a vibrant and growing field. Driven by advances in computing speed,

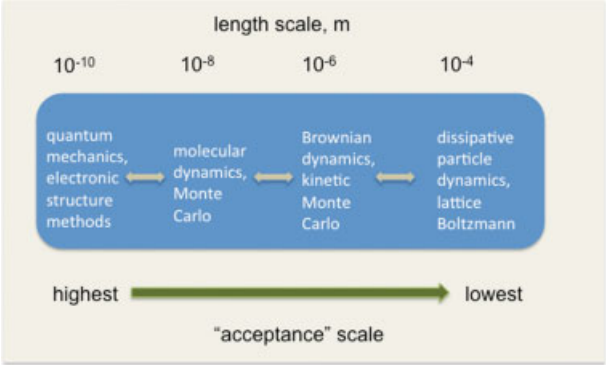

Figure 1. Different molecular modeling and simulation methods, along with characteristic length scales associated with each technique.

parallelization, and methodological advances, molecular simulations are able to address increasingly relevant problems for a fraction of the computational cost required just a few years earlier. As a result, the growth in these methods has been nothing short of fantastic. Figure 2 shows that the portion of the total science and engineering literature that used MD as a function of time has grown rapidly.[5] In 2008, there were over 1.4 million publications in science and engineering, and over 7,000 of these used MD. Since its invention in the 1950s, MD has grown to encompass about 0.5% of the entire science and engineering literature, and now more than 70,000 article using MD have been published. The number of articles using MC is even higher, although some of these are not "molecular-based" MC studies. It is clear that molecular simulation is playing an increasingly important role in science and engineering research. What are all these articles about?

Molecular simulation activities can be grouped into two categories: *discovery-driven* research and *data-driven* research[1]. Distinctions within these two broad classes are shown in Table 1.

The **discovery mode** consists of studies in which new phenomena are predicted that have not yet been observed experimentally, or explanations are sought for known phenomena that are not understood. It also includes efforts aimed at development and validation of new simulation methods or force fields. Some recent examples of new phenomena discovery include calculations predicting a negative thermal expansion coefficient for metal organic framework materials,[6] simulations that explained the vaporization process for ionic liquids,[7] the discovery of new porous materials,[8] and molecular simulations that predicted dramatically enhanced transport diffusivities of gases although smooth carbon nanotube pores.[9] It is more common for molecular simulations to be used to explain the origin of observed phenomena. There are many examples of this type of use, including studies directed at explaining the nature of hydrophobicity,[10] shear-induced crystallization,[11] processes behind amyloid formation[12] and criticality in ionic systems.[13] Finally, examples of methodological improvements and force field development include efforts to develop better Monte Carlo sampling methods,[14–16] improved nonequilibrium molecular dynamics methods,[17] and transferable force fields for thermodynamic and transport property calculations.[18,19] By their nature, discovery mode simulations are dominated by "academic" users (researchers at universities and national labs), and generally do not require the use of standardized methods or software. Often, flexible "prototyping" software serves discovery-driven researchers better than standardized programs, whose source code is often difficult to modify. Force fields must capture the essential physics of the system, but the trends and general phenomena sought in discovery studies do not necessarily require that the force fields yield accurate properties or that they be transferable from one molecule or state point to another.

Data-driven simulations consist of calculations where accurate property predictions are made with little or no input from experiment. Simulations can be used to interpolate between experimental data, extrapolate outside the range where data are available, or predict properties for compounds for which little or no data are available. Examples include predictions of the viscosity of alkanes,[19,20] binary phase diagrams,[21] and gas absorption isotherms.[22] Molecular simulations are also used to develop and validate molecular theories such as the SAFT equation of state,[23] models for adsorption and transport in porous materials,[24] and other equations of state and local composition models.[25,26] In some cases, molecular simulations can be used directly in conjunction with perturbation theory to estimate properties for process simulators.[27] Simulations can even be used as a check of experiments, especially for cases where interpretation is difficult or the experiment itself is hard to conduct. This is becoming an increasingly important area for simulations, with the advent of nanoscale experimental methods that indirectly address molecular-scale phenomena.[28,29] Data-driven simulations are best carried out with widely accepted and validated codes. Accurate force fields are required, and non-specialists need to be able to calculate many properties on a time scale that makes molecular simulation a better alternative than direct experimental property measurement. An oft-quoted statistic from an International Technology Research Institute report[30] illustrates

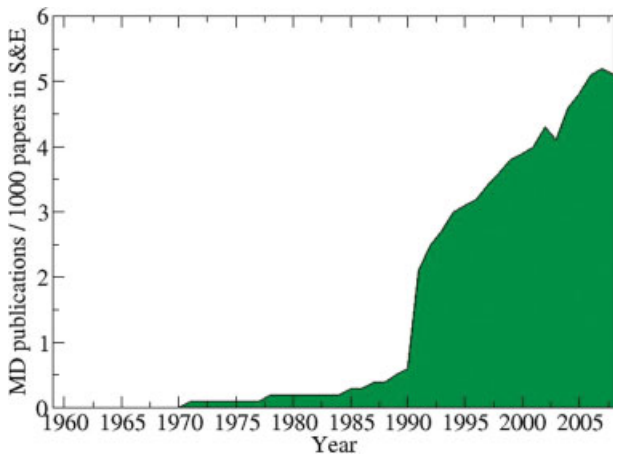

Figure 2. Fraction of articles in science and engineering (S&E) that use molecular dynamics.

Obtained from a search of Web of Science records for S&E journals.[5] Similar growth is observed for related methods such as Monte Carlo. As of 2008, over 0.5% of the more than 1.4 million articles published each year in Science and Engineering use MD. In specialized journals, the fraction is much higher. For example, in 2008 3% of all articles appearing in Physical Review Letters used MD.

Table 1. Examples of Discovery-Driven and Data-Driven Molecular Simulation Activities

	Type of activity	Examples	Tools and Users
Discovery	1. Search for new phenomena	• Simulations of new materials (nanotubes, metal organic frameworks, ionic liquids) • Nanoscale phenomena	• General or custom software • Standard or tailored force fields • New methods often required • Mainly academic users
Discovery	2. Explanation of phenomena	• Nature of hydrophobicity • Phase transitions • Anomalous behavior of water • Protein folding	• General or custom software • Standard or tailored force fields • May be associated with experimental projects • Academic or industrial users
Discovery	3. Development and validation of methods and force fields	• New free energy methods • Histogram reweighting • Accelerated dynamics • General force fields that can reproduce PVT and transport properties	• Rapid, flexible prototyping software • Automated force field development tools • Several methods required if validating against many properties • Mainly academic users
Data	1. Prediction of properties	• High pressure, temperature properties of materials • Activity coefficients, Henry's Law constants, isotherms • Enthalpies of mixing • Heat capacities	• Robust methods and software • Accurate and transferable force fields • Validated methods • Easy to use software • Database of force fields, properties • Industrial and some academic users
Data	2. Test theories and models	• Ideal Adsorbed Solution • Stefan-Maxwell • Regular Solution • SAFT • Critical scaling	• Specialized codes (wide range of phenomena) • Standard force fields • Mainly academic, some industrial users
Data	3. Validation of difficult experiments	• Crystal structure refinement • STM, AFM studies • Complex mixtures • Extreme conditions	• Accurate force fields • Specialized codes (wide range of phenomena) • Industrial and academic users

how data-driven quantum chemical calculations have achieved this practical level of importance to industry. In 1996, measuring a heat of formation for a single molecule cost $70,000, but could be computed with reasonable accuracy for $20,000. By 2000, the same experiment cost $100,000 but an even more accurate calculation cost $2,000. Today, it is hard to imagine a company that needs heats of formation not using computation almost exclusively.

The Industrial Fluid Properties Simulation Collective (IFPSC)[31] was established to help make data-driven molecular simulation a reality for the chemical industry. Their vision was that, "*A robust, accurate, and easy-to-use set of modeling tools will be widely available for the prediction of physical properties of fluids and obtaining insight into the connections between molecular structure and properties.*" Molecular simulations will become a mainstream chemical engineering tool only when data-driven molecular simulations can be conducted in a routine manner by non-experts. Essentially, this means that mainstream status will occur when the IFPSC vision is realized, but this vision remains elusive. The remainder of this Perspective describes the barriers and challenges that stand in the way of this vision.

Making Molecular Simulations a Mainstream Chemical Engineering Tool

The ordering of methods in Figure 1 is from smallest to largest characteristic length scale, but this ordering also holds for the relative "acceptance" of the methods in chemical engineering, where acceptance refers to the ease with which non-experts can and do use the methods with confidence to obtain meaningful results. For computational quantum chemistry, acceptance is high for several reasons. At its heart, computational quantum chemistry has a single, well-defined objective: find the energy of one or more molecules by solving the time-independent Schrödinger equation. From this one can understand bonding, structure, chemisorption thermodynamics, gas phase thermochemistry, and reaction mechanisms and rates. It is the most "mature" of the simulation techniques with a robust support infrastructure. The majority of practitioners use a handful of validated codes with refined interfaces. Benchmark calculations can be run and compared to those compiled by NIST at the Computational Chemistry Comparison and Benchmark DataBase. Web-based interfaces such as WebMo have been developed for setup and submission of jobs for a wide range of different packages. Basis sets can be downloaded (or uploaded) from the Basis Set Exchange service, supported by the Department of Energy Office of Science. A hierarchy of approximations is used, with generally accepted guidelines for how to increase the accuracy of a given calculation. As a result of all of these factors, computational quantum chemistry is used—and has widespread acceptance—by researchers in both academia and industry.[32] While experts pioneer new methods and applications, non-experts can easily use this tool in their daily research; it is common to see experimentalists supplement their work with quantum calculations.

The same cannot be said for molecular simulation, which is much more "diffuse" than quantum chemistry. As discussed previously, MD and MC are used to examine a vast range of topics, with most of the academic emphasis being on discovery-driven research. An enormous number of methods and techniques have been invented to enable these calculations to be carried out. Not surprisingly, there are few "standard" procedures or approaches that are agreed upon. The accuracy of a simulation hinges on the quality of the force field used, but there are no clear guidelines for how to systematically improve the quality of a force field. While there are several supported open source and commercial MD codes, many researchers still develop and write their own MD software because it facilitates discovery research. With MC, the number of choices for "standard" codes is much smaller, so most researchers develop their own methods specific to the problem they are studying. Setting up an MD or MC calculation typically requires a fair amount of patience and expertise; aside from a few examples, user interfaces for cutting edge software are primitive or nonexistent. Easy-to-use commercial software is available, but in general these closed proprietary codes lack the capabilities and power of modern open source codes. There are no formalized benchmark standards to compare calculations against. For all these reasons, it is much less likely for a non-specialist to conduct an MD or MC simulations to learn something about a system than for a non-expert to run a quantum calculation.

The major exception to this has been in the pharmaceutical industry, where molecular simulation has been enthusiastically adopted for both discovery-driven and data-driven applications. There are at least two reasons for this. First, the pharmaceutical industry spends significantly more on R&D than does the chemical industry and they have chosen to invest in molecular simulation. As an example, in 2007 both Pfizer ($8.1 billion), and Johnson and Johnson ($7.2 billion) *outspent the 22 major U.S. chemical companies combined spending of $5.38 billion*.[33] Second, the problems addressed in biological simulations tend to be more amenable to standard approaches. Temperatures and pressures are narrowly focused at physiological conditions, the dominant solvent is water and force fields parameterized to handle sugars, amino acids, nucleic acids and lipids, enable a huge range of biological problems to be addressed. In contrast, problems in the chemical industry are often concerned with wide ranges of temperatures and pressures, a vast number of different solvents, and a much broader range of molecule types and topologies.[34] In other words, while the problems in biological simulation are extremely complex, the set of inputs to a simulation and scope of properties to be computed are much more defined and limited than those present for a calculation in the chemical industry.

Molecular simulations can become a mainstream chemical engineering tool for data prediction as well as a new form of engineering modeling tool by following some of the examples that have worked for computational quantum chemistry and molecular simulations in the pharmaceutical industry. Below are some requirements that must be met for this to occur.

Narrow the scope of the problems addressed

Having a broad range of problems to address is desirable for discovery-driven molecular simulations, but for data-driven applications, the scope of what is calculated must be defined so that tools can be developed and optimized for those properties. Industry needs to indicate what the most important properties are, so that reliable tools for predicting these properties can be developed. Figure 3 shows one such ranking developed by Gupta and Olson at Dow Chemical Co.[34] In their estimation, simulations should focus on properties that are both hard to measure experimentally and are important in process design, such as those in the upper right quadrant of Figure 3. Those in the upper left quadrant are also important but less relevant, mainly because engineering models have been developed which do an acceptable job correlating these properties.

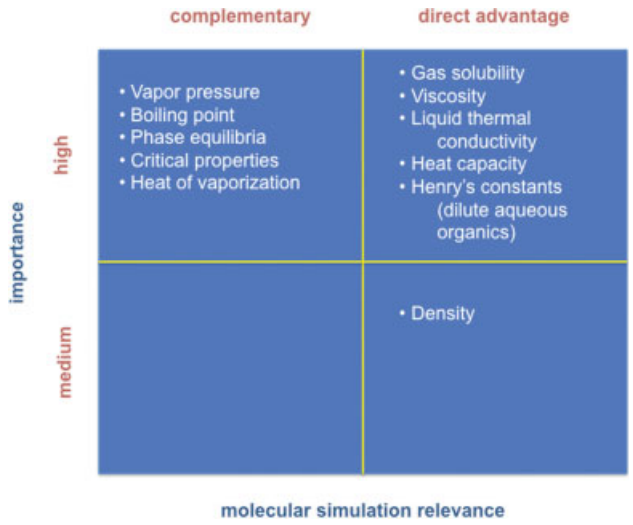

Figure 3. Physical properties for which molecular simulation can be used as a predictive data-driven research tool, according to researchers at Dow Chemical.

The abscissa ranks properties according to their relevance as compared to the difficulty of making the experimental measurement. The ordinate ranks properties according to their importance in chemical process design and development. The top right quadrant lists properties where molecular simulations are most likely to have an impact in chemical process design. (After Gupta and Olson, Ref. 34).

Simplify calculation setup

Given reasonable computer resources, a modern parallelized code, and a clear understanding of the properties to be computed, the amount of time a user might have to wait for the simulations necessary to compute the properties in the upper half of Figure 3 is actually quite short—perhaps on the order of a few hours to a few days. Yet if one were to try to compute all of the key properties in Figure 3, the actual time to completion would be measured in *months* and not *days*. This is because setting up a particular calculation can often take longer than the simulation itself, and each different property might require the use of a different method, a different piece of software, a different simulation and a different set of analysis tools. The human "setup" and "analysis" times for a simulation are a major barrier to the more widespread adoption of molecular simulations. Once a clearly delineated set of

properties to compute is identified, a set of tools to make such calculations easier to initiate should be developed. Web-based tools like WebMo are one model for how this can be done. Another is to use "front ends" developed with scripting languages such as Python, which can be linked to compiled "back end" simulation engines.

Provide integrated software and analysis tools

To calculate all the important properties in Figure 3, different simulation techniques must be used. MC is intrinsically the best method for computing phase equilibria, while MD is needed for transport properties. Either MD or MC can be used to compute volumetric properties and heat capacities. The fact that different methods are needed for different properties adds several complications to the use of these techniques.

First, there are several distinct MC methods that can be used, each having its own strengths and weaknesses. Second, MC is not an especially easy tool to make general, because individual moves that are useful for certain types of molecules are either unnecessary or ineffective for other types. Third, a user needs the expertise to select the proper method, ensemble and collection of moves for a given problem. This means general MC software capable of handling a range of molecules and force fields with smart systems that can select ensembles and moves are needed. As of now, there is only one open source MC code available—MCCCS Towhee—that comes close to meeting these requirements,[35] although in reality it is still far from what is needed in terms of performance, parallelization and ease of use.

On the other hand, MD is much more generally applicable than MC because its method of propagating trajectories forward in time is independent of the kind of molecules involved. There are many more general MD codes available for use, including LAMMPS, NAMD, GROMACS, CHARMM and DL_POLY. The techniques available in these codes are sufficient for computing transport properties as well as volumetric properties. By using methods such as thermodynamic integration, some thermodynamic properties can be computed with MD.

The bottom line is that no single or integrated package of software exists that can compute all of the properties required for chemically oriented data-driven research. The *individual* tools do exist, however, so ultimately efforts are needed to integrate these disparate methods into a useful form. Tools that translate input files, output files, and force field files between different software packages, as well as analysis programs that take the output of simulations and compute the desired properties would help. Such tools are being developed but on a case-by-case basis in different research groups. Coordination of these activities would pay big dividends for data-driven simulations.

Develop and archive force fields

If calculations are to be accurate for data-driven applications, then high-quality force fields must be available. A huge number of force fields have been developed over the years, but functional forms vary and the parameters are generally scattered throughout the literature. Extensive parameterizations have been made for biological molecules and several codes incorporate these force fields in their distribution. Even so, developing (or even selecting) a force field for a given molecule requires expertise and patience.There are at least two major problems associated with force fields that must be overcome. First, there are huge gaps in the availability of force fields for many molecules at the temperatures and pressures required for the chemical industry. Parameterizations must be conducted and verified against experimental data for a wide range of compounds of interest. Industry should identify classes of compounds and state points where the need is greatest, and funding to support open force field development for these systems should be provided. Force field development is often viewed as a thankless, necessary evil in the academic community, but it is an important task that would be taken on by researchers if funding were provided.

Second, once force fields are developed, that information needs to be captured and disseminated. A large amount of force field development work has been done already, but the results are scattered throughout the literature. Many force fields outside the biological area have been developed in an *ad hoc* manner for a few select compounds, with little regard to transferability or generality. Exceptions exist, such as the free TraPPE[18] and commercial COMPASS[36] force fields. Still, much more needs to be done before generally applicable and broad-based force fields are available. Like the different basis sets used in computational quantum chemistry, agreed upon functional forms and sets of parameters should be developed and then stored in an online database, similar to the basis set exchange database. Authors who develop new force fields must be encouraged to upload their force fields to these databases, while those who use a given parameter set could upload their results to provide feedback on the utility of the force field. By having force fields uploaded to a single verifiable site, problems that regularly plague conventionally published force fields such as typos, omitted parameters and poorly specified details of use would be eliminated. We have established a simple force field database at Notre Dame (http://www.nd.edu/~lmmr) in which Lennard-Jones parameters for a wide range of atom types are available for upload or download. Likewise, Ilja Siepmann's group has posted a searchable web-based database of TraPPE parameters at http://www.chem.umn.edu/groups/siepmann/trappe/intro.php. This basic concept could be built upon and expanded by agencies such as NSF, DOE or NIST. It is impossible to understate the importance of having a central repository for force field parameters. With such a tool, the research community could see and objectively evaluate the performance of force fields side-by-side, thereby driving innovation and improvement.

Provide validation databases

Validation refers to two different things. First, researchers need a sense of how accurate a data-driven calculation is, which is determined by comparison against experimental data. Second, one needs confidence that a particular method or piece of software performs as intended.

In the first case, a database should be established where calculated properties can be uploaded and stored, along with related experimental data. This could be associated with the force field database mentioned previously. In addition to the actual properties computed, complete details of how the simu-

lations were carried out, including input and force field files, would be archived. A database similar to NIST's popular Chemistry WebBook or Computational Chemistry Comparison and Benchmark DataBase are good models for this. To encourage submission of computed properties to the database, an approach similar to that taken for experimental data by the journals *Fluid Phase Equilibria*, *Journal of Chemical Engineering Data*, *International Journal of Thermophysics*, *Journal of Chemical Thermodynamics,* and *Thermochimica ACTA* could be used. Experimental data submitted to these journals are compared against data in NIST's Thermodynamics Research Center's database. Authors must provide their data in a format that enables comparison and archiving. Having similar capabilities for simulated data would enable direct comparison of the accuracy of different approaches, force fields and techniques. It would help identify weaknesses in methods and force fields as well as outright errors, thereby leading to systematic improvements. Moreover, such a validation procedure would contribute greatly to the credibility of simulations, by archiving an objective collection of molecular simulation successes (and failures).

The second type of validation that is required is one in which researchers can check whether a particular method or program is performing a calculation in the correct manner. Perhaps the best example of the way in which this is currently done is the extensive use and reliance on accepted Lennard-Jones fluid properties when developing new simulation techniques. Benchmarks for more complex systems such as water, aromatics, alkanes, and salts are needed. Forces and energies of molecular configurations could be tabulated and used to test codes. Benchmarks are also needed for analysis tools and methods that assign uncertainties. This latter area is one that is especially important, as there are widely varying methods used to assign uncertainty, not all of which are statistically sound. Again, NIST provides validation services for quantum calculations, and should consider doing the same for molecular simulations. Much of the validation that presently takes place occurs behind the scenes by informal discussions among researchers. It is time for benchmarks to be formally established and made known to all researchers.

Create a molecular simulation database

One of the more time-consuming aspects of conducting a molecular simulation is generating an initial equilibrated system. To do this from scratch, a representation of a single molecule of each type appearing in the simulation must be generated and then replicated to populate a simulation box at a given density. Finally, a simulation is run to relax the system at the state point of interest. NIST already provides structures of many single molecules on the Chemistry WebBook site. It would be a simple matter to develop an online "molecular database" where researchers could deposit structures of single molecules as well as entire equilibrated configurations for systems they have studied. The protein data bank (PDB) archive is a good model for this kind of database. The PDB works because it allows researchers to upload new structures as well as search for and download archived structures. Interestingly, the establishment of the protein data bank led to the pdb standard for how to describe molecular configurations. Similarly, developing a molecular simulation database would drive standardization of input and output file formats, as well as the development of tools that can translate formats between those used by different software packages.

Summary and Conclusions

Molecular simulation is a vibrant and growing area of research within the chemical engineering community. Its use is pervasive within the academic community. The most common form of use is in discovery-driven research, where new phenomena are predicted and explained. To become a mainstream computational tool, however, non-expert users must be capable of carrying out *data-driven* molecular simulations to compute properties of interest or develop and test new engineering models. Such data-driven calculations are already routinely performed for quantum chemical simulations, and to some extent for molecular dynamics simulations in the biological community. Monte Carlo and molecular dynamics simulations are not yet a mainstream tool in the chemical field, however, because several barriers are preventing their adoption.

The recommendations made here will help overcome these barriers. Many of these suggestions are already being undertaken by researchers around the world, but in an *ad hoc* manner. An agency or multinational body needs to help coordinate these efforts by establishing goals and guidelines, and providing funding to support the development of the tools, software and databases required. In the U.S., NIST is the most logical agency, given their mission of supporting commerce and their expertise in related matters. Other organizations such as the Centre Europeen de Calcul Atomique et Moleculaire (CECAM), or a consortium of companies and universities could also provide the necessary organization and coordination. When harnessed, the user community at large can accomplish much more than a single institution or company, as amply seen in the overwhelming success of open source and free software.

Over the last 50 years, molecular simulation has grown into a powerful tool for discovery-driven research, and now makes up a significant fraction of the total research enterprise. These advances are the result of twin developments in methods and computing power. While these advances continue, the adoption of molecular simulation as a mainstream tool for data-driven research will require the integration of disparate methods and the harnessing of information technology tools such as databases and web-based applications. Ultimately this is an engineering problem, and chemical engineers are poised to play a key role in this challenge in the coming decade.

Literature Cited

1. de Pablo JJ, Escobedo FA. Molecular simulations in chemical engineering: Present and future. *AIChE J*. 2002; 48:2716–2721.
2. de Pablo JJ. Molecular and multiscale modeling in chemical engineering - Current view and future perspectives. *AIChE J*. 2005;51:2372–2376.
3. Larson RG, Zhou Q, Shanbhag S, Park SJ. Advances in modeling of polymer melt rheology. *AIChE J*. 2007;53: 542–548.

4. Monson PA. Molecular thermodynamics of solid-fluid and solid-solid equilibria. *AIChE J*. 2008;54:1122–1128.
5. Brach C. Personal communication. University of Notre Dame Libraries; 2009.
6. Dubbeldam D, Walton KS, Ellis DE, Snurr RQ. Exceptional negative thermal expansion in isoreticular metal-organic frameworks. *Ang Chemie Int Ed*. 2007;46:4496–4499.
7. Strasser D, Goulay F, Kelkar MS, Maginn EJ, Leone SR. Photoelectron spectrum of isolated ion-pairs in ionic liquid vapor. *J Phys Chem A*. 2007;111:3191–3195.
8. Earl DJ, Burton AW, Rea T, Ong K, Deem MW, Hwang SJ, Zones SI. Synthesis and Monte Carlo structure determination of SSZ-77: A new zeolite topology. *J Phys Chem C*. 2008;112:9099–9105.
9. Chen HB, Johnson JK, Sholl DS. Transport diffusion of gases is rapid in flexible carbon nanotubes. *J Phys Chem B*. 2006;110:1971–1975.
10. Giovambattista N, Lopez CF, Rossky PJ, Debenedetti PG. Hydrophobicity of protein surfaces: Separating geometry from chemistry. *Proc Nat Acad Sci*. 2008;105:2274–2279.
11. Duff N, Lacks DJ. Shear-induced crystallization in jammed systems. *Phys Rev E*. 2007;75:031501.
12. Hall CK. Thermodynamic and kinetic origins of Alzheimer's and related diseases: a chemical engineer's perspective. *AIChE J*. 2008;54:1956–1962.
13. Kim YC, Fisher ME, Panagiotopoulos AZ. Universality of ionic criticality: Size- and charge-asymmetric electrolytes. *Phys Rev Lett*. 2005;95:195703.
14. Singh JK, Errington JR. Calculation of phase coexistence properties and surface tensions of n-alkanes with grand canonical transition matrix Monte Carlo simulation and finite size scaling. *J Phys Chem B*. 2006;110:1369–1376.
15. Karayiannis NC, Giannousaki AE, Mavrantzas VG, Theodorou, DN. Atomistic Monte Carlo simulation of strictly monodisperse long polyethylene melts through a generalized chain bridging algorithm. *J Chem Phys*. 2002;117:5465–5479.
16. Panagiotopoulos AZ. Direct determination of phase coexistence properties of fluids by Monte Carlo simulation in a new ensemble. *Mol Phys*. 1987;61:813–826.
17. Hulse RJ, Rowley RL, Wilding NV. Transient nonequilibrium molecular dynamic simulations of thermal conductivity: 1. Simple fluids. *Int J Thermophys*. 2005;26:1–12.
18. Martin MG, Siepmann JI. Transferable potentials for phase equilibria. 1. United atom description of n-alkanes. *J Phys Chem B*. 1998;102:2569–2577.
19. Gordon PA. Development of intermolecular potentials for predicting transport properties of hydrocarbons. *J Chem Phys*. 2006;125:014504.
20. Pan G, McCabe C. Prediction of viscosity for molecular fluids at experimentally accessible shear rates using the transient time correlation function formalism. *J Chem Phys*. 2006;125:194527.
21. Kamath G, Potoff JJ. Monte Carlo predictions for the phase behavior of H2S + n-alkane, H_2S+CO_2, CO_2+CH_4 and H_2S+CO_2+CH_4 mixtures. *Fluid Phase Equil*. 2006; 246:71–78.
22. Shi W, Maginn EJ. Atomistic simulation of the absorption of carbon dioxide and water in the ionic liquid 1-n-hexyl-3-methylimidazolium bis(trifluoromethylsulfonyl)imide ([hmim][Tf2N]. *J Phys Chem B*. 2008;112:2045–2055.
23. Shukla KP, Chapman WG. SAFT equation of state for fluid mixtures of hard chain copolymers. *Mol Phys*. 1997;91:1075–1081.
24. Keskin S, Liu JC, Johnson JK, Sholl DS. Testing the accuracy of correlations for multicomponent mass transport of adsorbed gases in metal-organic frameworks: Diffusion of H-2/CH4 mixtures in CuBTC. *Langmuir*. 2008;24:8254–8261.
25. Hu Y, Liu HL. Participation of molecular simulation in the development of molecular-thermodynamic models. *Fluid Phase Equil*. 2006;241:248–256.
26. Shi W, Maginn EJ. Molecular Simulation and Regular Solution Theory Modeling of Pure and Mixed Gas Absorption in the Ionic Liquid 1-n-Hexyl-3-methylimidazolium Bis(Trifluoromethylsulfonyl)amide ([hmim][Tf2N]). *J Phys Chem B*. 2008;112:16710–16720.
27. Elliott JR, Gerek ZN, Gray N. Combining molecular dynamics and chemical process simulation: The SPEAD model. *Asia-Pacific J Chem Eng*. 2007;2:257–271.
28. Kruger D, Fuchs H, Rousseau R, Marx D, Parrinello M. Pulling monatomic gold wires with single molecules: An ab initio simulation. *Phys Rev Lett*. 2002;89:186402.
29. Jobic H, Theodorou DN. Quasi-elastic neutron scattering and molecular dynamics simulation as complementary techniques for studying diffusion in zeolites. *Microporous Mesoporous Mater*. 2007;102:21–50.
30. Westmoreland PR, Kollman PA, Chaka AM, Cummings PT, Morokuma, K, Neurock M, Stechel EB, Vashishta P. Applications of molecular and materials modeling. Spring-Verlag, New York (2008); first published as a report in 2002.
31. http://fluidproperties.org.
32. Cheng HS, Parekh VS, Mitchell JW, Hayes KS. Quantum mechanical calculations of chemical thermodynamics: Practice and limitations. *AIChE J*. 1997;43:2153–2156.
33. Baum R. Facts and figures. *C&E News*. 2008;83:3.
34. Gupta S, Olson JD. Industry needs in physical properties. *I&EC Res*. 2003;42:6359–6374.
35. http://towhee.sourceforge.net.
36. Sun H. COMPASS: An ab initio forcefield optimized for condensed phase application: Overview with details on alkane and benzene compounds. *J Phys Chem B*. 1998; 102:7338–7364.