

MOLECULAR-MODELING METHODS AND USE FOR PRODUCT AND PROCESS DESIGN

Phillip R. Westmoreland
Department of Chemical Engineering
University of Massachusetts Amherst, Amherst, MA 01003, USA

Athanassios Z. Panagiotopoulos
Department of Chemical Engineering
Princeton University, Princeton, NJ 08544, USA

Abstract

Significant progress has been made in recent years on development of molecularly based methods for accurate prediction of chemical and physical properties relevant for product and process design. A number of ideal-gas properties are routinely predictable. However, despite progress in computing hardware and simulation methodologies, prediction of condensed-phase properties of interest to industrial applications is not currently practiced on a routine basis. Present status of theory, molecular models, and industrial practice will be described.

Molecular simulation refers to computational statistical mechanics based on force fields. Most existing force fields have been optimized to the configurational properties of isolated molecules and thermodynamic or structural properties of liquids near room temperature. Recently developed force fields have become available that also reproduce phase coexistence properties and critical parameters for selected systems, but they are not yet generally applicable to many systems of interest. Simulation methodologies for rapid determination of intermolecular potential parameters from experimental data are discussed. Two key unresolved questions remain, namely how to incorporate polarizability and other non-additive interactions, and the logistics of large-scale efforts to obtain parameters for broad classes of components.

Computational quantum chemistry does not require predetermined force fields. Instead, it predicts energy and related properties from the nuclear (geometric) structure and the electronic orbital structure using the Schrödinger equation. Solution time and storage requirements increase rapidly with the number of atoms, but high accuracy can be achieved. At present, most results are for the zero-K, isolated-molecule (ideal-gas) case, but methods are advancing rapidly, including for solvation.

Keywords

Force fields, intermolecular potentials, thermodynamic properties, phase coexistence, Monte Carlo, molecular dynamics, quantum chemistry, kinetics

Introduction

Desirable physical and chemical properties of substances are the reasons we produce them. We need a fuel; then we want it to be combustible (have a reasonable ignition temperature and exothermic heat of combustion), to be capable of being extracted (fluid properties, known P - V - T

behavior) or manufactured (physical or chemical separation of unwanted components, VLE behavior in refining, catalytic processing), and to be suitable for storage or distribution (P - V - T behavior again, perhaps adsorption). Perhaps we need a particular thermoplastic.

It must have the right monomer; there must be suitable synthesis chemistry for the monomer; the monomer must be sufficiently pure; there must be a suitable polymerization process; and polymerization and processing of the polymer must give a molecular structure and morphology that yields the desired glass transition temperature, decomposition temperature. Perhaps we need a pharmaceutical HIV inhibitor. It must dock into the receptor site; it must have satisfactory ADMET behavior (absorption, distribution, metabolism, excretion, toxicology); and it must be produced by stepwise chemical synthesis, by a natural or modified biological organism, or a combination of methods.

Modeling is increasingly a way to obtain useful property predictions. Accurate, precise measurement is the reference for establishing any property. However, not all properties needed for product and process design can be measured accurately or at all. Accurate correlations are next best. To be used with confidence, such correlations must be based on a sound fundamental foundation and on sufficient reference data. In the face of insufficient data or unacceptable precision, we would like to be able to predict properties theoretically.

This paper reviews the practical prediction of useful properties from interatomic and intermolecular energies of interaction. These energies may be correlated (e.g., van der Waals) or predicted from quantum mechanics. Most of the properties are then generated using statistical mechanics through equations or computational simulations.

Part I: Force Fields and Molecular Simulations

Despite progress in computing hardware and simulation methodologies, prediction of properties of interest to industrial applications for condensed phases is not currently practiced on a routine basis. The lack of appropriate intermolecular potential functions is often quoted as *the* most important barrier for application of atomistic simulation methodologies to problems of industrial interest. A legitimate question is why this is still the case, given that many general force fields are now available that have been optimized to structural, bonding and thermodynamic properties of a large number chemical components. The answer to this question is that many important properties are not currently considered during potential model development. In particular, most generally available force fields have not been tested for their ability to reproduce phase coexistence properties of pure fluids and mixtures over broad temperature ranges. Most existing force fields also have not been extensively tested for density or concentration conditions far removed from the pure liquid state at ambient temperature and pressure. Unfortunately, one cannot expect good performance of these force fields at conditions that have not been considered in their development. This, in turn, severely limits the reliability and predictive ability of molecular modeling methods.

This section focuses on computational methodologies that have been found useful for development of intermolecular potential models that can reproduce coexistence properties over a broad temperature range. More extensive earlier reviews of these methods are available (Panagiotopoulos, 2000a; de Pablo and Escobedo, 2002).

Gibbs Ensemble Monte Carlo

The Gibbs Ensemble Monte Carlo simulation methodology (Panagiotopoulos, 1987; Panagiotopoulos *et al.*, 1988; Smit *et al.*, 1989) enables direct simulations of phase equilibria in fluids. Gibbs ensemble simulations are performed in two separate microscopic regions, each within periodic boundary conditions. The thermodynamic requirements for phase coexistence are that each region should be in internal equilibrium, and that temperature, pressure and the chemical potentials of all components should be the same in the two regions. System temperature in Monte Carlo simulations is specified in advance. The remaining three conditions are satisfied by performing three types of Monte Carlo moves: displacements of particles within each region (to satisfy internal equilibrium), fluctuations in the volume of the two regions (to satisfy equality of pressures), and transfers of particles between regions (to satisfy equality of chemical potentials of all components). The method has been frequently used, in combination with configurational bias sampling discussed in the following paragraph, to determine phase diagrams of a large number of fluids. Reviews of the Gibbs method and its applications are available (Panagiotopoulos 1995, 2000a).

Configurational Bias Sampling and Expanded Ensembles

The most common bottleneck in achieving convergence in methods that rely on particle transfers is the prohibitively low acceptance of transfer attempts. For dense fluid phases, especially for complex, orientation-dependent intermolecular potentials, configurations with "holes" in which an extra particle can be accommodated are highly improbable, and the converse step of removing a particle involves a large cost in energy.

Configurational-bias sampling techniques significantly improve sampling efficiency for Gibbs or grand canonical Monte Carlo simulations. The methods have been reviewed in detail by Frenkel (1995), Frenkel and Smit (1996), and Siepmann (1999). The methods trace their ancestry to biased sampling for lattice polymer configurations proposed by Rosenbluth and Rosenbluth (1955). Development of configurational-bias methods for canonical and grand canonical simulations and for continuous-space models took place in the early 1990's (Frenkel *et al.*, 1992; de Pablo *et al.*, 1992; Siepmann and Frenkel, 1992; Laso *et al.*, 1992).

Configurational-bias methods are based on segment-by-segment insertions or removals of a multisegment molecule. Several trial directions are attempted for every

segment insertion, and a favorable growth direction is preferentially selected for the segment addition. This way, the acceptance probability of insertions is greatly enhanced.

For each segment growth or removal step, a correction factor (often called “Rosenbluth weight”) is calculated. The product of the Rosenbluth weights of all steps is incorporated in the overall acceptance criterion for particle insertions and removals in order to correct for the bias introduced by the non-random growth along preferential directions.

Another approach for handling multisegment molecules is based on the concept of expanded ensembles (Lyubartsev *et al.*, 1992; Wilding and Müller, 1994; Vorontsov-Velyaminov *et al.*, 1996; Escobedo and de Pablo, 1996). Expanded ensembles for chain molecules construct a series of intermediate states for the molecule of interest, from a non-interacting (phantom) chain to the actual chain with all segments and interactions in place. These intermediate states can be semi-penetrable chains of the full length or shortened versions of the actual chain. Estimates of the free energy of the intermediate states are required to ensure roughly uniform sampling, as for thermodynamic and Hamiltonian scaling methods mentioned in the previous section. The advantage of expanded ensembles over configurational-bias methods is that arbitrarily complex long molecules can be sampled adequately, if sufficient computational effort is invested in constructing good approximations of the free energies of intermediate states.

Histogram Reweighting Methods

Early in the history of development of simulation methods it was realized that a single calculation can, in principle, be used to obtain information on the properties of a system for a range of state conditions (McDonald and Singer, 1976; Wood, 1968; Card and Valleau, 1970). However, the practical application of this concept was severely limited by the performance of computers available at the time. In more recent years, several groups have confirmed the usefulness of this concept, first in the context of simulations of spin systems and later for continuous-space fluids (Wilding, 1995; Panagiotopoulos *et al.*, 1998; Wilding, 1997; Potoff and Panagiotopoulos, 1998).

Simulation data are subject to statistical uncertainties, which are particularly pronounced near the extremes of particle numbers and energies visited during a run. When data from multiple runs are combined, the question arises of how to determine the optimal amount by which to shift the raw data in order to obtain a global free energy function. Ferrenberg and Swendsen (1988) provided a solution to this problem that minimizes the differences between predicted and observed histograms. All thermodynamic quantities for the system over the range of densities and energies covered by the histograms can be obtained from the combined histograms.

In the absence of phase transitions or at temperatures near a critical point, the values of all observable quantities are independent of initial conditions, since free energy barriers for transitions between states are small or nonexistent. However, at lower temperatures, free-energy barriers for nucleation of new phases become increasingly larger. The states sampled at a given temperature and chemical potential depend on initial conditions, a phenomenon known as hysteresis. The histogram reweighting method can be applied to systems with large free-energy barriers for transitions between states, if care is taken to link all states of interest *via* reversible paths.

Thermodynamic and Hamiltonian scaling

Two methods related to histogram reweighting are thermodynamic and Hamiltonian scaling Monte Carlo. Thermodynamic-scaling techniques proposed by Valleau (1991) are based on calculations in the constant pressure (*NPT*), rather than the grand canonical (μVT) ensemble and provide information for the free energy over a range of volumes, rather a range of particle numbers. Thermodynamic scaling techniques can also be designed to cover a range of Hamiltonians (potential models) in the Gibbs (Kiyohara *et al.*, 1996) or Grand Canonical ensemble (Errington and Panagiotopoulos, 1998).

In their Hamiltonian-scaling form, the methods are particularly useful for optimizing parameters in intermolecular potential models to reproduce experimental data such as the coexisting densities and vapor pressures. Thermodynamic and Hamiltonian scaling methods require estimates for the free energy of the system as a function of conditions, so that the system can be forced to sample the range of states of interest with roughly uniform probability.

Gibbs-Duhem Integration

Most methods for determination of phase equilibria by simulation rely on particle insertions to equilibrate or determine the chemical potentials of the components. Methods that rely on insertions experience severe difficulties for dense or highly structured phases. If a point on the coexistence curve is known (e.g. from Gibbs ensemble simulations), the remarkable method of Kofke (1993a, 1993b) enables the calculation of a complete phase diagram from a series of constant-pressure simulations that do not involve any transfers of particles. The method enables calculations of solid-fluid coexistence (Agrawal and Kofke, 1995), for which other methods described in this paper are not applicable. The method and its applications have been recently reviewed (Kofke, 1999).

For one-component systems, the method is based on integration of the Clausius-Clapeyron equation over temperature, which involves only “mechanical” quantities that can be simply determined in the course of a standard Monte Carlo or molecular dynamics simulation. From the known point on the coexistence curve, a change in

temperature is chosen, and the saturation temperature at the new temperature is predicted. Two independent simulations for the corresponding phases are performed at the new temperature, with gradual changes of the pressure as the simulations proceed to take into account the enthalpies and densities at the new temperature as they are being calculated.

Density-of-States Methods

In recent years, several methods that work directly with the density-of-states (or free energy) of a system have been proposed. These methods have the potential advantage that they use more information from a simulation than simply the run averages and can overcome free energy barriers that would prevent efficient sampling; *e.g.*, in the presence of phase transitions or for systems with glassy dynamics. Two broad classes of such methods are based on the Wang-Landau technique (Wang and Landau, 2001) and the “transition matrix” approach (Fitzgerald *et al.*, 1999). In the Wang-Landau method, one accumulates a histogram of the density of states “on the fly” in a temperature-independent way. The method has been further developed and its convergence characteristics for continuum fluids improved by Yan and de Pablo (2003) and Shell *et al.* (2003). The transition matrix approach is based on keeping track of probabilities of transition between any two states, thus using information from rejected as well as accepted moves. The method has been used by Errington (2003) to study vapor-liquid equilibria in simple fluids to a high accuracy. An extensive review of density-of-states methods is in preparation (Shell *et al.*, 2004).

General Force Fields

Many force fields for atomistic simulations have been developed over the past decades. No attempt is made in the present paper to provide an exhaustive review of all such efforts. Instead, the main approaches taken by previous researchers are briefly illustrated.

Most current force fields describe bonded interactions by stretching, bending and torsional potentials. The bonded interaction parameters are obtained by fitting to *ab initio* data, usually obtained at the Hartree-Fock level with large basis sets. Partial charges are used for describing electrostatic interactions. In recent years, the Ewald summation technique for handling the long-range character of the electrostatic interactions has become popular.

In one approach for describing non-bonded interactions, pioneered in the early work by Jorgensen *et al.* (1984), non-bonded potential parameters are obtained by fitting experimentally observed liquid densities and enthalpies and structural properties near room temperature. The latest versions of these models are “all-atom” force fields explicitly incorporating hydrogens. Examples include the OPLS (Jorgensen *et al.*, 1996), CHARMM

(Mackerell *et al.*, 1995), AMBER (Cornell *et al.*, 1995) and COMPASS (Sun, 1998) force fields. The primary focus of these models is on biological and organic molecules in aqueous solution. For simple hydrocarbons, there is good, but not quantitative, agreement between calculated and experimental coexistence curves (Chen *et al.*, 1998).

Another approach to obtaining force field parameters is to minimize the deviations between predicted and observed bond lengths, angles, vibrational frequencies, conformational energies, and minimum-energy conformations of clusters. A current example of this approach is the MMFF force field developed by Halgren (1996a-d). Unfortunately, even for simple hydrocarbon systems, this approach can result in completely unrealistic thermodynamic properties (Chen *et al.*, 1998).

The force fields described in this section share several common features. In particular, non-bonded dispersion interactions are described by Lennard-Jones 12-6 potentials. The only exception is the COMPASS force field, which uses a 9-6 potential. Only two-body interactions are included; three- and higher-order interactions are effectively incorporated in the values of the potential model parameters. The errors introduced by this approximation are most pronounced if calculations at densities and temperatures significantly different from the ones used in the parameterization. For example, the value of the dipole moment for the water molecule in these models is significantly higher than the gas-phase value. This is done in order to compensate for polarizability interactions that lead to a higher value of the dipole moment in liquid water, but it leads to incorrect values for the second virial coefficient and energy at low densities.

Force Fields for Phase Coexistence Properties

In the recent years, availability of the methods described in the previous section has had an impact on force field model development by enabling determination of parameters that reproduce accurately phase coexistence properties over broad temperature ranges. The performance of models with respect to coexistence properties is of paramount importance for design of separation processes. However, there are many other reasons why such models are highly desirable. In particular, at low temperatures (near the triple point), the chemical potential of the liquid essentially determines the vapor pressure, so that models that do well for phase coexistence also describe accurately the free energy of the dense liquid. Near the critical point, phase coexistence involves phases with densities intermediate between those of a gas and a highly subcritical liquid. The requirements for good performance over a broad range of densities and temperatures are much more severe than the thermodynamic properties at a single temperature and density used in development of many currently available general force fields.

Hydrocarbons

Hydrocarbon molecules are ubiquitous in industrial processes and form the building blocks of biological systems. They are non-polar and consist of a small number of groups, thus making them the logical starting point for potential model development.

Three accurate united-atom potential sets for *n*-alkanes have appeared recently. The TRAPPE (Martin and Siepmann, 1998) and NERD models (Nath *et al.*, 1998) use the Lennard-Jones (12,6) potential to describe non-bonded interactions among methyl and methylene groups, while the model of Errington and Panagiotopoulos (1999) uses the exponential-6 functional form. All three reproduce the experimental phase diagrams and critical points. The exponential-6 model is slightly better with respect to representation of the vapor pressures. Deviations from experimental data for the exponential-6 united atom model are comparable to those for a recently developed all-atom TRAPPE model (Chen and Siepmann, 1999).

United-atom potentials for branched alkanes have been developed by Cui *et al.*, (1997), Martin and Siepmann (1999) and for α -olefins by Spyriouni *et al.* (1999).

Polar fluids

There have been several recent studies of the phase behavior of polar compounds such as *n*-alkanols (van Leeuwen, 1996), hydrogen sulfide (Kristof and Liszi, 1999) and carbon disulfide (Kristof *et al.*, 1996). However, no transferable force fields that can be used to obtain the phase behavior of polar fluids with reasonable accuracy are currently available. An example of the challenges in predictions of the behavior of polar fluids is provided by Visco and Kofke (1998), who studied two *ab initio* and one empirical potential model for hydrogen fluoride. The *ab initio* models did not model accurately the saturated liquid densities, and none of the models predicted correctly the vapor pressure and heat of vaporization as functions of temperature.

Water

Because of the importance of water in biological and chemical systems, a large number of models have been proposed for atomistic simulations. Rigid fixed point charge models for water are often used in simulations of biological systems because of their simplicity and reasonable predictions of the structure and thermodynamics of liquid water at ambient conditions. Such models include the TIP4P (Jorgensen *et al.*, 1983), SPC (Berendsen *et al.*, 1981) and SPC/E (Berendsen *et al.*, 1987) models. All represent water as a single Lennard-Jones (12,6) sphere within which are embedded positive and negative charges. The "Exp-6" model (Errington and Panagiotopoulos, 1998) is also a fixed point charge model,

but utilizes the exponential-6 functional form for the repulsion and dispersion interactions and was parameterized to the phase coexistence properties. The SPC/E model predicts a critical temperature near the experimental value, but underestimates the vapor pressure (and thus the chemical potential) by more than a factor of two. The SPC model has the correct vapor pressure, but severely underestimates the critical temperature. The TIP4P model has an even lower T_c and overestimates the vapor pressure. The Exp-6 model does very well for both T_c and P_{VP} , and has the closest ρ_c to the experimental value. However, the Exp-6 model does not do as well as the other models for structure of the liquid, in particular with respect to the oxygen-oxygen pair correlation function. It also has a dielectric constant lower than that the experimental value at room temperature.

Clearly, none of these simple fixed point charge models for water can adequately reproduce thermodynamic and structural properties over a broad temperature range. As also suggested earlier, the inclusion of only two-body interactions also results in a higher "effective" dipole moment relative to the isolated (gas phase) molecule.

Several polarizable models for water are available in the literature, but none seem to be better than the simple fixed point charged models with respect to the coexistence properties and critical parameters (Kiyohara *et al.*, 1998). Some of the newer "fluctuating-charge" models developed from *ab initio* calculations (Stern *et al.*, 1999) have not yet been tested with respect to their phase coexistence behavior. Chen *et al.* (2000) performed an extensive search for optimized parameters for polarizable water models. Two optimized force fields were proposed. The SPC-pol-1 force field yields good saturated vapor and liquid densities, heats of vaporization and dielectric constants but does not represent well the liquid structure. The TIP4P-pol force field yields excellent liquid structures but does not perform well for the coexistence properties near the critical point. Chen *et al.* concluded that "we have to continue our quest for improved water force fields."

Mixtures

A key question for all force fields for atomistic simulations is their ability to predict properties of mixtures without use of additional adjustable parameters for the unlike-pair interactions. Extensive phase coexistence data are available for many mixtures, but they have not generally been used to validate proposed models. This is primarily because of the difficulties associated with simulations of phase coexistence for mixtures prior to the development of the computational methods discussed in the previous section.

For mixtures of non-polar components such as hydrocarbons or inert gases, there is considerable evidence that the newer force fields optimized to the phase equilibrium properties can be used for reliable predictions

of mixture behavior. For example, for a mixture of ethane with *n*-heptane (Errington and Panagiotopoulos, 1999), excellent agreement is obtained with experimental data without use of any mixture parameters.

An important unresolved question is that of combining rules for unlike-pair parameters of the repulsion and dispersion terms. The Lorentz-Berthelot combining rules (arithmetic mean for the diameters and geometric mean for potential well depths) have been used almost universally for unlike-pair parameters in predictive simulations. However, some recent results (Potoff *et al.*, 1999), suggest that different combining rules may be more appropriate for groups that differ significantly in size or polarity. Fig. 8 presents calculations for the pressure-composition diagram of propane with CO₂. The Kong (1973) combining rules, which result in less attractive unlike-pair interactions than the Lorentz-Berthelot rules, give better agreement with experimental data for this and similar mixtures. Potoff *et al.* (1999) also find that for mixtures with large differences in polar character between the components, current models do not result in quantitative agreement with experiments.

Discussion and Future Directions

A number of computational methods that have been developed over the past few years enable the routine calculation of free energies and phase coexistence properties for model systems. These methods are now being used to obtain optimized intermolecular potentials for many substances. The resulting force fields have not yet attained the level of generality of currently available academic and commercial force fields, and they have not yet been incorporated in standard simulation codes. However, these developments are likely to occur in the near future.

Despite relatively rapid progress, there are many unresolved questions in the area. The first set of questions involves the functional form of the potential models. It is clear that for strongly interacting molecules, such as water, no simple two-body effective potential can quantitatively reproduce both structure and thermodynamics and additional interactions are necessary. The most important of these is likely to be polarizability. Recent methodological developments (Martin *et al.*, 1998; Chen *et al.*, 2000) permit incorporation of polarizability in Monte Carlo calculations with a penalty of a factor of ten in CPU time relative to non-polarizable model calculations. This is still a rather large cost, so the quest for other methods to incorporate polarizability will likely continue.

An interesting new concept that may have some impact on potential model development is the use of finely discretized lattice models suggested by Panagiotopoulos and Kumar (1999) and Panagiotopoulos (2000b). The basic idea is that lattice models can be made to approximate their continuum analogs by increasing the ratio, ζ , of particle diameter to lattice spacing. The

surprising result is that relatively low values of ζ are required for quantitative agreement of the calculated thermodynamic properties between lattice and off-lattice models. For Lennard-Jones and exponential-6 fluids, $\zeta=10$ results in agreement between lattice and continuum calculations for the phase envelope and critical parameters to within narrow statistical uncertainties. Even the structure of the fluids, which would have been expected to show some lattice artifacts, is identical between the two systems. The advantage of the finely discretized models is numerical speed. For ionic systems, the speed advantage is a factor of up to 100, while for the Lennard-Jones and exponential-6 interactions have a factor between 10 and 20. A recent paper (Kumar, 2000) illustrates how this concept can be used to obtain quantitatively accurate lattice models for hydrocarbons.

Finally, a barrier to rapid force field development and testing with respect to free energies and phase coexistence predictions has been the lack of availability of portable, high quality codes appropriate for this purpose. Multiple groups (primarily academic) have been active in this area, but there has been relatively limited coordination and exchanges of data and codes.

There are currently very few industrial contributors to potential model development for simulations of fluid phases (e.g. Nath, 2003). However, the Fluid Properties Simulation Challenges (<http://www.comsef.aiche.org>) are based on the idea that industry and government can encourage development of predictive force fields and methods by assessing and rewarding success.

Part II: Computational Quantum Chemistry

Ideal-gas thermochemistry has been the most useful property prediction from computational quantum chemistry ("computational chemistry"). The methods are also rapidly becoming valuable for product and process design by generating reaction thermodynamics and kinetics, solution properties, and the potentials for molecular simulations.

Calling these methods *ab initio* is common but somewhat misleading because it implies that "first-principles" calculations are performed without reference to any data except fundamental physical constants. The two groups of Parr, Craig, and Ross (1950) had separately and independently begun some of the first calculations - and thus they described each set of calculations as being *ab initio*! These types of calculations are indeed based on fundamental energy equations. However, the results are affected by the functional forms chosen for the atomic and molecular wavefunctions, by how completely the electronic configurations are included, and by the statistical mechanics used to convert quantum-mechanical, zero-kelvin properties into thermal properties.

This section will briefly review quantum-chemistry theories and methods that are in common use. Methods of obtaining property values and relations will then be described.

Theory and Methods of Quantum Chemistry

Hehre et al. (1986) present an excellent summary of the basic theories. In computational quantum chemistry, molecular-scale structures, rovibrational properties, and electronic states are based on energies from the electronic Schrödinger equation, usually summarized as:

$$E\Psi = H\Psi \quad (1)$$

The quantum-mechanical Hamiltonian operator H expresses energy functionality in terms of kinetic energy and potential energy. In contrast to the classical Hamiltonian function, this H is an operator on the quantized eigenfunction Ψ , the system wavefunction.

Equation 1 is a simplification of the time-dependent Schrödinger equation, which includes both nuclear and electronic kinetic energy. Nucleus-nucleus, electron-electron, and nucleus-electron Coulombic interactions are included in both forms of the equation. Nuclear motion is decoupled from the analysis by the Born-Oppenheimer approximation, which recognizes that electronic velocities are much faster than the nuclear motions. Thus, energy may be calculated very accurately for specific nuclear arrangements (molecular structures) where the nuclei are not moving relative to each other; *i.e.*, at zero kelvins.

Wavefunctions and Basis Sets

A valuable and useful approximation of the system wavefunction Ψ is to compose it of wavefunctions for the atomic and molecular orbitals. The atomic orbitals are represented by functional forms or “basis sets.” Molecular orbitals ϕ are represented by linear combination of atomic orbitals (LCAO).

Basis sets are commonly based on the atomic orbitals for H atom: 1s, 2s, 2p, 3s, 3p, 3d, and so on. Rather than directly using analytical functions solved for purely hydrogenic orbitals, Gaussian-form functions are commonly used to simulate atomic-orbital wavefunctions. For a given orbital, size may be varied by interpolating smaller (contracted) and larger (diffuse) sizes. Likewise, shape may be varied naturally by combining a base atomic orbital with another. Thus, 6-31G(d,p) refers to a basis set that uses linear combinations of six gaussian “primitives” to represent each core, nonvalence molecular orbital; of three gaussian primitives for the contracted representation and one gaussian primitive for the diffuse representation of each valence orbital; of d -orbital character mixed into p -orbital functions; and of p -orbital character mixed into s -orbital functions.

The system or molecular wavefunction is then constructed from the molecular-orbital wavefunctions. Orbital spin is included functionally with each molecular orbital, pairing the electrons (restricted methods such as RHF) or considering them individually (unrestricted methods). The Hartree-Fock system wavefunction is

constructed as a determinant of the molecular orbitals, the simplest mathematical formula which allows antisymmetric behavior (*i.e.*, interchanging two orbitals, represented by two rows, must change the sign of Ψ).

Higher-accuracy basis sets have also been developed, such as correlation-consistent basis sets [Wilson et al., 1999, and references within]. For atoms with high numbers of electrons like transition metals, effective core potentials have been developed that include relativistic effects from the electrons closest to the nucleus and reduce the number of non-valence gaussians to be evaluated.

Solution Methods: Level of Theory

A wide range of accuracy and computational expense exists among different solution methods. With a specified basis set, the Schrödinger energy may be calculated and minimized by varying undetermined constants of the system wavefunction. For solution methods that obey the variational principle, this minimum is guaranteed not to be less than the true energy. Thus, any lower value from a different level of theory would be a more accurate energy.

Undetermined constants being optimized are in the interpolations of contracted and diffuse representations, in the interpolations of mixed atomic-orbital types, and particularly in the linear combinations of atomic orbitals that form molecular orbitals. Larger basis sets will have more adjustable constants, should give better answers, and will require a more costly calculation.

Hartree-Fock is the simplest *ab initio* method. From the variational principle, Roothaan (1951) and Hall (1951) derived optimization equations in terms of undetermined constants based on single-electron orbital energies interacting with a charge field from the other electrons. Constants are determined by iterative solution; thus the method is also known as the Self-Consistent Field method. A central limitation is that no electron interacts with any other electron, but only with a field, so electron correlation effects like van der Waals interactions are not included. CPU usage goes up approximately as N^3 , where N is strictly the number of basis-set functions but is roughly proportional to the number of non-hydrogen atoms.

Møller-Plesset perturbation theory incorporates electron correlation through second- (MP2) and fourth-order (MP4) modifications to the Hamiltonian operator. MP2 and MP4 CPU usages go up approximately as N^5 and N^7 . This rate of increase limits calculations to rather small molecules even with rapidly increasing computer power.

Configuration Interaction and Coupled Cluster methods include the effects of individual electrons occupying energy levels higher than in the ground state, promoted singly, doubly, or triply. CCSD(T), coupled cluster with singles, doubles, and extrapolated triples, is a high but useful level of *ab initio* calculation.

Electronic density-functional theory (DFT) was developed for material physics, but it has become popular for chemical properties in the last 10 years. It scales roughly like Hartree-Fock (N^3) but gives more accurate

structures by including electron correlation. Unfortunately, it does not obey the variational principle.

Rather than solving in terms of the wavefunction Ψ , solution is for the electron density $\rho = \langle \Psi^* | \Psi \rangle$. Kohn and co-workers (Hohenberg and Kohn, 1964; Kohn and Sham, 1965) divided electronic energy into four parts: the conventional electron kinetic energy, potential energy of nuclear-nuclear and nuclear-electron interaction, potential energy of electron-electron repulsion, and “exchange energy” E^{xc} , which incorporates all other electron-electron interactions.

The Kohn-Sham equations are variational, but because exchange energy is represented by approximations, DFT implementations are not. Happily, excellent and quite general approximations have been developed. E^{xc} is broken into “exchange” and “correlation” functionals, E^x and E^c . The former was greatly advanced by Becke’s 3-parameter functional (Becke, 1993) that mixes the classical local-density approximation (sufficient for much solid-state physics) with effects of the gradient of electron density (necessary for chemical accuracy) from Hartree-Fock theory and other methods. It is often coupled with the gradient-corrected correlation functional of Lee, Yang, and Parr (Lee *et al.*, 1988) to give the B3LYP method.

The most chemically accurate results presently come from extrapolative methods like G3 and CBS-QB3. Curtiss, Pople, and co-workers developed the successive G1, G2, and G3 methods (Curtiss *et al.*, 1998) to extrapolate to the most accurate energy from different levels of basis sets and different levels of theory (different “model chemistries”). Complete Basis Set methods extrapolate basis-set effects directly to an infinite-set limit (Nyden and Petersson, 1981; Montgomery *et al.*, 2000). CBS-Q and CBS-QB3 are based on MP2 and B3LYP optimizations, respectively. These two methods and G3 are somewhat expensive, but they give $\Delta_f H^\circ$ within about 1 kcal/mol and are affordable for molecular sizes of about 12 non-hydrogen, first- and second-row atoms. In our experience, CBS-QB3 converges most reliably.

Semiempirical methods, such as AM1 and PM3, are the lowest-level quantum chemical methods. They make approximations in the Schrödinger equation based on parametric fits to data and so are high-level correlations. For simple, ground-state organic molecules or as estimates for higher-level calculations or very large systems, they can be quite acceptable. They do not perform well on transition states or for hydrogen bonding.

Quantum-chemical modeling of solvent effects has reached a practical, commercial stage. Solvent molecules may be represented individually or in continuum solvation models that take various forms (Cramer and Truhlar, 1999; Truhlar, 2001), all representing the solvent effect as a dielectric field. COSMO, the COnductor-like Screening MOdel (Klamt and Schüürmann, 1993), has been coded into a wide range of quantum chemistry codes. COSMO treats the molecule as ideally screened, and in the COSMO-RS extension (Klamt, 1995), a surface around

the molecule is evaluated pointwise for its effective screening charge density. Different levels of quantum chemistry can be used for this evaluation. Developed originally at Bayer, additional companies are using this method for prediction of mixture properties, as described below.

Software is available from a number of companies and other organizations. As examples, Accelrys, Tripos, and the OCTA project emphasize molecular simulations, while quantum chemistry codes are available from Gaussian, Wavefunction (Spartan), Schrödinger (Jaguar), Accelrys, Fujitsu (CACHe), Hypercube (Hyperchem), and DOE’s Pacific Northwest Laboratories (NWChem).

Promising Developments

Multi-scale modeling combines methods suited to different scales of time and size (Maroudas, 2000). It is challenging, but it is also becoming a practical reality. QM/MM (Singh and Kollman, 1986; Field *et al.*, 1990), the Embedded-Atom method of Daw and Baskes (1983) and the ONIOM method of Morokuma (Svensson *et al.*, 1996) make high-level calculations for a central set of atoms and couple them to a lower-level treatment of the larger system. In a different approach applied to materials physics, Abraham *et al.* (1998) have overlapped physical regions in order to couple scales from quantum-mechanical to finite-element continuum modeling.

Ab Initio Molecular Dynamics has proven quite useful for modeling condensed matter. Car and Parrinello (1985) introduced the approach of computing interatomic forces by DFT during a molecular dynamics calculation. In recent years, the VASP codes of Kresse and Hafner (1993) have won widespread acceptance for modeling metals, ionic solids and heterogeneous catalysis.

Reactive molecular simulations using fitted quantum-mechanical reactive force fields are also promising. Covalent-bond reaction chemistry is always influenced by internal energy and state of matter, while electronic structure (discussed below) is usually computed for isolated molecules at zero kelvins. In 1990, Brenner introduced hydrocarbon force fields that gave good structures but also allowed bond making and breaking (Brenner, 1996). In papers by Nyden *et al.* (1992), Rappe *et al.* (1997), van Duin *et al.* (2001), and their subsequent work, this approach has advanced in both its force fields and codes. Computationally, reactive molecular dynamics with many molecular motions but rare reaction events makes such calculations very lengthy, but techniques similar to Parallel Replica Dynamics (Voter, 1998) are promising. Quantitative success has been demonstrated in polymer pyrolysis (Nyden *et al.*, 2004).

Property-Prediction Methods

Strictly speaking, the only result from quantum chemical calculations is energy for a set of atoms with specified nuclear positions -- a “structure.” This structure

may be optimized to locate the absolute minimum-energy structure (ground state), the structure of a classical transition state (maximum with respect to a reaction coordinate, minimum with respect to all other degrees of freedom), or structures of excited electronic states.

The Hessian (matrix of energy second derivatives) gives force constants, which may be represented as classical harmonic vibrational frequencies ω by Newton's Second Law and Hooke's Law:

$$F = m \frac{\partial^2 x}{\partial t^2} = -kx \Rightarrow \omega = \sqrt{\frac{k}{m}} \quad (2)$$

for displacement x and Hookian force constant k . These frequencies are generally higher than spectroscopically measured, requiring an approximate scaling factor ranging from 0.89 for HF to 0.995 for BLYP DFT calculations (Scott and Radom, 1996; National Institute of Standards and Technology, 2003).

Ideal-gas specific heat $C_p^\circ(T)$ and entropy $S^\circ(T)$ may be calculated with the optimized molecular structure and its frequencies. Statistical mechanics gives equations for translational, rotational, vibrational, and electronic contributions (McQuarrie, 1973; Ochterski, 2000). The Rigid-Rotor Harmonic Oscillator or RRHO model assumes purely harmonic rovibrational modes, and relatively low levels of quantum-chemical theory are often sufficient to give good numbers. Corrections for anharmonicity are reasonably straightforward for hindered internal rotations (Benson, 1976) but more difficult for other types.

Heat of reaction at zero K is calculated by using the computed energies and zero-point energies, weighted stoichiometrically as is normally done with heats of formation. The reference states for energy of *ab initio* calculations are the dissociated atoms. Thus, to obtain heats of formation, atomic energies are used relative to the elements in their standard states. For these enthalpies to be accurate, rather high levels of theory are required, such as CBS-Q or -QB3, G3, or CCSD(T)/cc-pVTZ.

Transport properties and phase equilibrium are more the province of molecular simulations, but the necessary force fields now usually are based in large measure on quantum chemistry. Dipole moments are easily obtained, based on partitioning partial charges by local electron density on each atom. COSMO-based methods are increasingly quantitative for prediction of properties like vapor pressure and vapor-liquid equilibrium of mixtures.

Quantitative Structure-Property Relations or QSPRs are widely used, coming from a long tradition of fundamentally based, purely statistical correlations of engineering properties. For comparison, group additivity is fruitful for ideal-gas thermochemistry (Benson, 1976) and UNIFAC thermodynamic activities (Fredenslund *et al.*, 1975, 1977) because the groups are valid molecular descriptors of the properties. QSPRs may use more subtle molecular descriptors as well, such as *ab initio* bond

lengths, angles, dihedrals, and dipole moments. They have become very useful in the pharmaceutical and chemicals industries.

Part III: Examples from industrial practice

Molecular simulations and computational chemistry have proven useful predictors of properties in industry, including for product and process development. They are proving especially valuable in early development and design activities. In final process design, precision is quite important. In the early stages, accurate, less-precise values and trends from fundamentally based predictions are critical for success.

Numerous examples have been described by Westmoreland *et al.* (2002), who present analyses based on site reports from 91 companies and other organizations in Europe, Japan, and the United States and on activities at about 500 additional organizations around the world.

Ideal-gas thermochemistry has been the most widely used property prediction from computational chemistry. Rondan (1996) reported that these approaches had become a cost-efficient solution to Dow Chemical's need for heats of reaction for safety analysis and process design. As an example, he described computation of the heat of formation for tetrahydropyrimidine, a closed-shell, cyclic $C_4N_2H_8$ species. A G2 calculation on a Cray C90 required 150 CPU hours and 20 Gb storage. It cost \$8,000 in computer charges and about \$20,000 total, compared to an estimated \$70,000 for a bomb calorimetry determination of the same uncertainty. The results were then used to generate a Benson group for extension to other similar compounds. By 2000, Dow estimated that the cost-effectiveness for better calculations had improved to \$2,000 vs. \$150,000 (Westmoreland *et al.*, 2002).

Development of homogeneous catalysts for olefin polymerization has been another major application. These catalysts are large organometallic molecules interacting with the olefins by a transition metal. Access to the active site is governed by the organic structure that surrounds it (Fig. 1), which leads to control of polymer structure, morphology, and physical properties. Computational chemistry, including solvation effects, has been used to test possible leads and develop commercial catalysts at numerous companies. Some companies that have been engaged in such activities are Asahi Chemical, BASF, BP / Amoco, Degussa, Dow, DSM, DuPont, Enichem, the Institut Français du Pétrole, Mitsubishi, Phillips Petroleum, and Totalfina Elf (Westmoreland *et al.*, 2002).

The biggest success story in financial terms has been drug discovery, which has made great use of molecular simulation. In the 1970s when computational chemistry began to be practically useful, and continuing on into the 1980s, the dream was to design drugs by complete calculation. Experience proved that, more realistically, calculation could be a valuable aid to experiments: interpreting protein structures from crystallography, especially the binding or sites for disease agent or drug

“docking”; testing binding effectiveness at these sites; and developing libraries of chemical compounds for combinatorial chemistry. Among the drugs cited as having been developed with particular aid from molecular modeling are Agenerase, an HIV protease inhibitor from Vertex; Cozaar, an antihypertensive from Merck and DuPont Pharmaceuticals (now Bristol-Myers Squibb), and Aricept, an acetylcholinesterase inhibitor from Eisai for Alzheimer’s treatment (Westmoreland *et al.*, 2002). An important, unresolved technical need is to be able to compute protein folding, modeling a functional three-dimensional protein from its molecular strands.

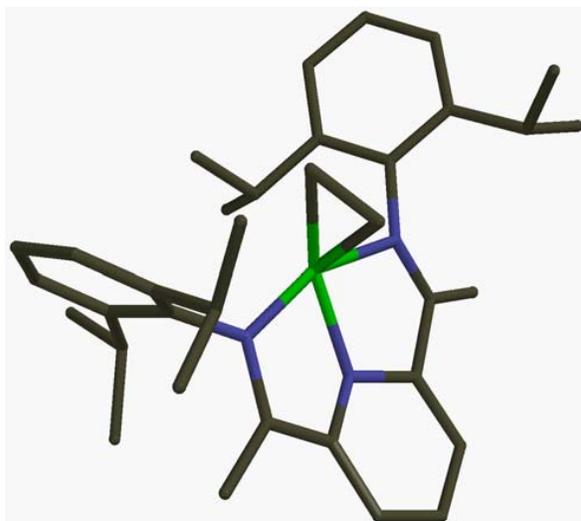


Figure 1. Ethylene bound atop a homogeneous catalyst of iron (triangle) liganded with di(methyl imide xylene) aniline.

Computed properties for phase equilibrium are beginning to be evaluated and used by the chemical process industries. Kolár *et al.* (2002) described how Mitsubishi Chemical Corporation is using COSMO-RS and UNIFAC to choose solvents for pharmaceutical product and process development. The work is relevant in their company for tasks extending into reactions, extraction, and product formulation.

Still other properties have proven amenable to molecular modeling and modeling-based QSPRs. A few are adsorption properties for gas separations, band gaps in electronic and photonic materials, colors of dyes and pigments, polymer structure-morphology relations, gelling agents for personal care products, and mouth-feel of cereals and other foods (Westmoreland *et al.*, 2002).

This power should not be overstated. Even as high accuracy for molecular systems begins to be available, problem size and cost are issues. Accuracy is always important, but the needs for precision vary; cost and even feasibility may be determined by the requested precision. Also, many things cannot yet be computed. Large systems are obvious examples, although systems of the size of tens of millions of atoms can now be treated at low levels of

theory. The Human Genome Project provides another example. While the molecular structure of the genome may be known, its translation to generating specific proteins is only beginning to be worked out, much less the translation to physiological function.

Education is another issue, both for the potential users of codes and for the potential users of results. Selection of the best tool (code, method, basis set, etc.) is still a matter of experience supplemented by some recommendations from the literature and fellow users. Reliable convergence and ease of use are good, but the right answer to the wrong question may be dangerously useless.

Finally, these tools are aids to understanding, not wisdom itself. Numbers with many significant digits and visually appealing pictures should not be mistaken for the ability to solve any and all problems in property prediction. These tools are aptly likened to analytical instruments, which are invaluable in answering questions but are only aids for developing the questions. Accurate properties are necessary for all product and process design, and their value still requires scientific and engineering expertise.

Acknowledgements

Research in AZP’s group related to the present manuscript was supported by the US Department of Energy, Office of Basic Energy Sciences, under grant DE-FG02-01ER1512. PRW’s research in this area is supported by the US Department of Energy, Office of Basic Energy Sciences, under grant DE-FG02-91ER14192.

References

- Abraham, F.F., J.Q. Broughton, N. Bernstein, and E. Kaxiras (1998). Spanning the length scales in dynamic simulation, *Comput. Phys.*, **12**, 538.
- Agrawal, R. and D. A. Kofke, (1995). Solid-Fluid Coexistence for Inverse-Power Potentials. *Phys. Rev. Lett.*, **74**, 122-125.
- Becke, A. D. (1993). Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.*, **98**, 5648–52.
- Benson, S. W. (1976). *Thermochemical Kinetics*, 2nd Ed. Wiley-Interscience, New York.
- Berendsen, H. J. C., J. P. M. Postma, W. F. van Gusteren and J. Hermans (1981). In *Intermolecular Forces*, Pullmann, B. (ed.); Reidel: Dordrecht, 331.
- Brenner, D. W. (1996). Chemical Dynamics and Bond-Order Potentials. *Mat. Res. Soc. Bull.* **21**, 36.
- Car, R., and M. Parrinello (1985). Unified Approach for Molecular Dynamics and Density-Functional Theory. *Phys. Rev. Lett.*, **55**, 2471–2474.
- Card, D. N. and J. P. Valleau (1970). Monte Carlo study of the thermodynamics of electrolyte solutions. *J. Chem. Phys.*, **52**, 6232-40.
- Chen, B., M. G. Martin and I. J. Siepmann (1998). Thermodynamic Properties of the Williams, OPLS-AA and MMFF94 All-atom Force Fields for Normal Alkanes. *J. Phys. Chem. B*, **102**, 2578-86.

- Chen, B., and J. I. Siepmann (1999). Transferable potentials for phase equilibria. 3. Explicit-hydrogen description of normal alkanes. *J. Phys. Chem. B*, **103**, 5370-9.
- Chen, B., J. Xing and J. I. Siepmann (2000). Development of Polarizable Water Force Fields for Phase Equilibrium Calculations. *J. Phys. Chem.*, **104**, 2391-40.
- Cornell, W. D., P. Cieplak, C. L. Bayly, I. R. Gould, K. M. Merz Jr., D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman (1995). A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.*, **117**, 5179-5197.
- Cramer, C. J., and D. G. Truhlar (1999). Implicit Solvation Models: Equilibria, Structure, Spectra, and Dynamics. *Chem. Rev.*, **99**, 2161-2200.
- Cui, S. T., P. T. Cummings and H. D. Cochran (1997). Configurational bias Gibbs ensemble Monte Carlo simulation of vapor-liquid equilibria of linear and short-branched alkanes. *Fluid Phase Equilibria*, **141**, 45-61.
- Curtiss, L. A., K. Raghavachari, P. C. Redfern, V. Rassolov, and J. A. Pople (1998). Gaussian-3 theory for molecular energies of first- and second-row compound. *J. Chem. Phys.*, **109**, 7764-76.
- Daw, M. S., and M. I. Baskes (1983). Semiempirical, quantum mechanical calculation of hydrogen embrittlement in metals. *Phys. Rev. Lett.*, **50**, 1285-1288.
- De Pablo, J. J., M. Laso and U. W. Suter (1992). Simulation of polyethylene above and below the melting point, *J. Chem. Phys.*, **96**, 2395-2403.
- De Pablo, J. J., and F. A. Escobedo (2002). Molecular Simulations in Chemical Engineering: Present and Future. *AIChE J.*, **48**, 2716-2721.
- Errington, J. R. (2002). Direct calculation of liquid-vapor phase equilibria from transition matrix Monte Carlo simulation. *J. Chem. Phys.*, **118**, 9915-25.
- Errington, J. R., and A. Z. Panagiotopoulos (1998). Phase equilibria of the modified Buckingham exponential-6 potential from Hamiltonian scaling grand canonical Monte Carlo. *J. Chem. Phys.*, **109**, 1093-1100.
- Errington, J. R., and A. Z. Panagiotopoulos (1999). A New Intermolecular Potential Model for the n-Alkane Homologous Series. *J. Phys. Chem. B*, **103**, 6314-6322.
- Escobedo, F., and J. J. de Pablo (1996). Expanded grand canonical and Gibbs ensemble Monte Carlo simulation of polymers. *J. Chem. Phys.*, **105**, 4391-4394.
- Ferrenberg, A.M., and R. H. Swendsen (1988). New Monte Carlo technique for studying phase transitions. *Phys. Rev. Lett.*, **61**, 2635-2638.
- Field, M. J., P. A. Bash, and M. Karplus (1990). A Combined Quantum Mechanical and Molecular Mechanical Potential for Molecular Dynamics Simulations. *J. Comp. Chem.*, **11**, 700-733.
- Fitzgerald, M., R. R. Picard, and R. N. Silver (1999). Canonical transition probabilities for adaptive Metropolis simulation. *Europhys. J.*, **46**, 282-287.
- Fredenslund, A., R. L. Jones, and J. M. Prausnitz (1975). Group-Contribution Estimation of Activity Coefficients in Nonideal Liquid Mixtures. *AIChE J.*, **21**, 1086-1099.
- Fredenslund, A., J. Gmehling, and P. Rasmussen (1977). *Vapor-Liquid Equilibrium Using UNIFAC*, Elsevier Scientific Publishing Company, Amsterdam.
- Frenkel, D., G. C. A. M. Mooij, and B. Smit (1992). Novel Scheme to Study Structural and Thermal Properties of Continuously Deformable Molecules. *J. Phys.: Condens. Matter*, **4**, 3053-3076.
- Frenkel, D. (1995). Chapter in Baus, M., L. F. Rull and J.-P. Ryckaert (editors), *Observation, Prediction and Simulation of Phase Transitions in Complex Fluids*, NATO ASI Ser. C, vol. 460, pp. 357-419.
- Frenkel, D., and B. Smit (1996). *Understanding Molecular Simulation*, Academic Press, London.
- Halgren, T.A. (1996a). Merck Molecular Force Field. I. Basis, Form, Scope, Parametrization, and Performance of MMFF94. *J. Comp. Chem.*, **17**, 491-519.
- Halgren, T. A. (1996b). Merck Molecular Force Field. II. MMFF94 van der Waals and Electrostatic Parameters for Intermolecular Interactions. *J. Comp. Chem.*, **17**, 520-552.
- Halgren, T. A. (1996c). Merck Molecular Force Field. III. Molecular Geometries and Vibrational Frequencies for MMFF94. *J. Comp. Chem.*, **17**, 553-586.
- Halgren, T. A. (1996d). Merck Molecular Force Field. IV. Conformational Energies and Geometries for MMFF94. *J. Comp. Chem.*, **17**, 587-615.
- Hall, G. G. (1951). The molecular orbital theory of chemical valency. VIII - A method of calculating ionization potentials. *Proc. Royal Soc. (London)*, **A205**, 541-552.
- Hehre, W. J., L. Radom, P. V. Schleyer, and J. Pople (1986). *Ab initio molecular orbital theory*, Wiley-Interscience, New York.
- Hohenberg, P., and W. Kohn (1964). Inhomogeneous Electron Gas. *Phys. Rev.*, **136**, B864-B871.
- Jorgensen, W.L., J. D. Madura, and C. J. Swenson (1984). Optimized Intermolecular Potential Functions for Liquid Hydrocarbons. *J. Am. Chem. Soc.*, **106**, 6638-6646.
- Jorgensen, W.L., D. S. Maxwell, and J. Tirado-Rives (1996). Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.*, **118**, 11225-36.
- Jorgensen, W.L., J. Chandrasekhar, and J. D. Madura (1983). Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, **79**, 926-935.
- Jorgensen, W.L., J. D. Madura, and C. J. Swenson (1984). Optimized Intermolecular Potential Functions for Liquid Hydrocarbons. *J. Am. Chem. Soc.*, **106**, 6638-6646.
- Kiyohara, K., T. Spyrouni, K. E. Gubbins, and A. Z. Panagiotopoulos (1996). Thermodynamic-Scaling Gibbs Ensemble Monte Carlo: A new method for determination of phase coexistence properties of fluids. *Molec. Phys.*, **89**, 965-974.
- Kiyohara, K., K. E. Gubbins, and A. Z. Panagiotopoulos, (1998). Phase Coexistence Properties of Polarizable Water Models. *Molec. Phys.*, **94**, 803-808.
- Klamt, A. (1995). Conductor-like Screening Model for Real Solvents: A New Approach to the Quantitative Calculation of Solvation Phenomena. *J. Phys. Chem.*, **99**, 2224-2235.
- Klamt, A., and G. Schüürmann (1993). COSMO: A New Approach to Dielectric Screening in Solvents with Explicit Expressions for the Screening Energy and its Gradient. *J. Chem. Soc., Perkin Trans.*, **2**, 799-805.
- Kofke, D. A. (1993a). Gibbs-Duhem integration: A new method for direct evaluation of phase coexistence by molecular simulation. *Molec. Phys.*, **78**, 1331-1336.
- Kofke, D. A. (1993b). Direct evaluation of phase coexistence by molecular simulation via integration along the saturation curve. *J. Chem. Phys.*, **98**, 4149-4162.
- Kofke, D. (1999). Semigrand Canonical Monte Carlo Simulations: Integration along Coexistence Lines. *Adv. Chem. Phys.*, **105**, 405-441.
- Kohn, W., and L. J. Sham (1965). Quantum Density Oscillations in an Inhomogeneous Electron Gas. *Phys. Rev.*, **137**, A1697-A1705.

- Kolár, P. J.-W. Shen, A. Tsuboi, and T. Ishikawa (2002). Solvent selection for pharmaceuticals. *Fluid Phase Equilib.*, **194-197**, 771-782.
- Kong, C. L. (1973). Combining rules for intermolecular potential parameters. II. Rules for the Lennard-Jones (12-6) potential and the Morse potential. *J. Chem. Phys.*, **59**, 2464-2467.
- Kranias, S., D. Pattou, B. Levy, and A. Boutin (2003). An optimized potential for phase equilibria calculation for ketone and aldehyde molecular fluids. *Phys. Chem. Chem. Phys.*, **5**, 4175-4179.
- Kresse, G., and J. Hafner (1993). Ab initio molecular dynamics for open-shell transition metals. *Phys. Rev. B*, **48**, 13115-13118; also see <http://cms.mpi.univie.ac.at/vasp/>.
- Kristof, T., and J. Liszi (1997). Effective intermolecular potential for fluid hydrogen sulfide. *J. Phys. Chem. B*, **101**, 5480-5483.
- Kristof, T., Liszi J. and I. Szalai (1996). Vapour-liquid equilibria for a model of liquid carbon disulphide. *Molec. Phys.* **89**, 931-942.
- Kumar, S. K. (2000). Quantitatively accurate lattice models for hydrocarbon polymers, preprint.
- Laso, M., J. J. de Pablo, and U. W. Suter (1992). Simulation of phase equilibria for chain molecules. *J. Chem. Phys.*, **97**, 2817-2819.
- Lee, C., W. Yang, and R. G. Parr (1988). Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B*, **37**, 785-789.
- Lyubartsev, A., A. Martsinovski, and S. Shevkunov (1992). New approach to Monte Carlo calculation of the free energy: Method of expanded ensembles. *J. Chem. Phys.*, **96**, 1776-1783.
- Mackerell, A. D., J. Wiorkiewiczkuczera, and M. Karplus (1995). An all-atom empirical energy function for the simulation of nucleic acids. *J. Am. Chem. Soc.*, **117**, 11946-11975.
- Maroudas, D. (2000). Multi-scale modeling of hard materials: Challenges and opportunities for chemical engineering, *AIChE J.*, **46**, 878-882.
- Martin, M. G., and J. I. Siepmann (1998). Transferable Potentials for Phase Equilibria. 1. United-Atom Description of n-Alkanes. *J. Phys. Chem. B*, **102**, 2569-2577.
- Martin, M. G., B. Chen, and J. I. Siepmann (1998). A novel Monte Carlo algorithm for polarizable force fields: Application to a fluctuating charge model for water. *J. Chem. Phys.*, **108**, 3383-3385.
- Martin, M. G., and J. I. Siepmann (1999). Novel Configurational-Bias Monte Carlo Method for Branched Molecules. Transferable Potentials for Phase Equilibria. 2. United-Atom Description of Branched Alkanes. *J. Phys. Chem B*, **103**, 4508-4517.
- McDonald, I.R., and K. Singer (1967). Machine Calculation of Thermodynamic Properties of a Simple Fluid at Supercritical Temperatures. *J. Chem. Phys.*, **47**, 4766-4772.
- McQuarrie, D. A. (1973). *Statistical Thermodynamics*, Harper and Row, New York.
- Montgomery, J. A. Jr., M. J. Frisch, J. W. Ochterski, and G. A. Petersson (2000). A complete basis set model chemistry. VII. Use of the minimum population localization method. *J. Chem. Phys.*, **112**, 6532-6542.
- Nath, S. K. (2003). Molecular simulation of vapor-liquid phase equilibria of hydrogen sulfide and its mixtures with alkanes. *J. Phys. Chem. B*, **107**, 9498-9504.
- Nath, S. K., F. A. Escobedo, and J. J. de Pablo (1998). On the simulation of vapor-liquid equilibria for alkanes. *J. Chem. Phys.*, **108**, 9905-11.
- National Institute of Standards and Technology (2004). Vibrational scale factors in *Computational Chemistry Comparison and Benchmark Database, Release 9*, NIST Standard Reference Database 101, <http://srdata.nist.gov/cccbdb/vsf.asp>.
- Nyden, M. R., G. P. Forney, and J. E. Brown (1992). Molecular Modeling of Polymer Flammability: Application to the Design of Flame-Resistant Polyethylene. *Macromolecules*, **25**, 1658-1666.
- Nyden, M. R., and G. A. Petersson (1981). Complete Basis Set Correlation Energies. I. The Asymptotic Convergence of Pair Natural Orbital Expansions. *J. Chem. Phys.*, **75**, 1843-1862.
- Nyden, M. R., S. I. Stoliarov, P. R. Westmoreland, Z. X. Guo, and C. Jee (2004). Applications of reactive molecular dynamics to the study of the thermal decomposition of polymers and nanoscale structures. *Mat. Sci. Eng.*, **A465**, 114-121.
- Ochterski, J. W. (2000). Thermochemistry in Gaussian. http://www.gaussian.com/g_whitepap/thermo.htm.
- Panagiotopoulos, A. Z. (1987). Direct determination of phase coexistence properties of fluids by Monte Carlo simulation in a new statistical ensemble. *Molec. Phys.*, **61**, 813-826.
- Panagiotopoulos, A. Z., N. Quirke, M. Stapleton, and D. J. Tildesley (1988). Phase equilibria by simulation in the Gibbs ensemble. Alternative derivation, generalization and extension to mixture and membrane equilibria. *Molec. Phys.*, **63**, 527-545.
- Panagiotopoulos, A. Z. (1995) chapter in Baus, M., L. F. Rull and J.-P. Ryckaert (editors) *Observation, Prediction and Simulation of Phase Transitions in Complex Fluids*, NATO ASI Ser. C, vol. 460, pp. 463-501.
- Panagiotopoulos, A. Z., V. Wong, and M. A. Floriano (1998). Phase Equilibria of Lattice Polymers from Histogram Reweighting Monte Carlo Simulations. *Macromolecules*, **31**, 912-918.
- Panagiotopoulos, A. Z., and S. K. Kumar (1999). Large Lattice Discretization Effects on the Phase Coexistence of Ionic Fluids. *Phys. Rev. Lett.*, **83**, 2981-2984.
- Panagiotopoulos, A. Z. (2000a). Monte Carlo methods for phase equilibria of fluids. *J. Phys.: Condens. Matter*, **12**, R25-R52.
- Panagiotopoulos, A. Z. (2000b). On the equivalence of continuum and lattice models for fluids. *J. Chem. Phys.*, **112**, 7132-7137.
- Parr, R. G., D. P. Craig, and I. G. Ross (1950). Molecular orbital calculations of the lower excited electronic levels of benzene, configuration interaction included. *J. Chem. Phys.* **18**, 1561-1563.
- Potoff, J. J., and A. Z. Panagiotopoulos (1998). Critical point and phase behavior of the pure fluid and a Lennard-Jones mixture. *J. Chem. Phys.*, **109**, 10914-10920.
- Puibasset, J., and R.J.M. Pellenq (2003). Grand canonical Monte Carlo simulation study of water structure on hydrophilic mesoporous and plane silica substrates. *J. Chem. Phys.*, **119**, 9226-9232.
- Rappe, A. K., M. A. Pietsch, D. C. Wiser, J. R. Hart, L. M. Bormann-Rochotte, and W. M. Skiff (1997). RFF, conceptual development of a full periodic table force field for studying reaction potential surfaces. *Mol. Eng.*, **7**, 385-400.
- Rondan, N. G. (1996). Application of Computational Thermochemistry to Industrial Process Design, AIChE Annual Meeting, Chicago.

- Roothaan, C. C. J. (1951). New Developments in Molecular Orbital Theory. *Rev. Mod. Phys.*, **23**, 69-89.
- Rosenbluth, M. N., and A. W. Rosenbluth, (1955). Monte Carlo calculations of the average extension of molecular chains. *J. Chem. Phys.*, **23**, 356-359.
- Scott, A.P., and L. Radom (1996). Harmonic Vibrational Frequencies: An Evaluation of Hartree-Fock, Moller-Plesset, Quadratic Configuration Interaction, Density Functional Theory and Semiempirical Scale Factors. *J. Phys. Chem.*, **100**, 16502-16513.
- Shell, M.S., P. G. Debenedetti, and A. Z. Panagiotopoulos (2003). An improved Monte Carlo method for direct calculation of the density of states. *J. Chem. Phys.*, **119**, 8526-36.
- Shell, M.S., P. G. Debenedetti and A. Z. Panagiotopoulos (2004). Simulation methods for free energy calculations of fluids; Review chapter in preparation.
- Siepmann, J. I., and D. Frenkel, (1992). Configurational bias Monte Carlo: a new sampling method for flexible chains. *Molec. Phys.*, **75**, 59-70.
- Siepmann, J. I. (1999). In D. M. Ferguson, J. I. Siepmann, and D. G. Truhlar, Eds. (1999), "Monte Carlo Methods in Chemical Physics", *Adv. Chem. Phys.*, 105, Wiley, New York, pp. 443-460.
- Singh, U.C., and P.A. Kollman (1986). A Combined Ab Initio Quantum Mechanical and Molecular Mechanical Method for Carrying out Simulations on Complex Molecular Systems: Applications to the $\text{CH}_3\text{Cl}^+\text{Cl}^-$ Exchange Reaction and Gas-Phase Protonation of Polyethers. *J. Comp. Chem.*, **7**, 718-730.
- Smit, B., Ph. De Smedt, and D. Frenkel (1989). Computer simulations in the Gibbs ensemble. *Molec. Phys.*, **68**, 931-950.
- Spyrioumi T., I. G. Economou, and D. N. Theodorou (1999). Molecular simulation of alpha-olefins using a new united-atom potential model: Vapor-liquid equilibria of pure compounds and mixtures. *J. Am. Chem. Soc.*, **121**, 3407-3413.
- Stern, H. A., G. A. Kaminski, J. L. Banks, R. Zhou, B. J. Berne, and R. A. Friesner (1999). Fluctuating Charge, Polarizable Dipole, and Combined Models: Parameterization from ab Initio Quantum Chemistry. *J. Phys. Chem. B*, **103**, 4730-7.
- Sun, H. (1998). COMPASS: An ab Initio Force-Field Optimized for Condensed-Phase Applications-Overview with Details on Alkane and Benzene Compounds. *J. Phys. Chem. B*, **102**, 7338-64.
- Svensson, M., S. Humbel, R. D. J. Froese, T. Matsubara, S. Sieber, and K. Morokuma (1996). ONIOM: A Multilayered Integrated MO + MM Method for Geometry Optimizations and Single Point Energy Predictions. A Test for Diels-Alder Reactions and $\text{Pt}(\text{t-Bu})_3_2 + \text{H}_2$ Oxidative Addition. *J. Phys. Chem.*, **100**, 19357-19363.
- Truhlar, D. G. (2001). Molecular-scale modeling of reactions and solvation. *Foundations of Molecular Modeling and Simulation*, AIChE Symp. Series **97**, 71-83.
- Valleau, J. P. (1991). Density-scaling: a new Monte Carlo technique in statistical mechanics. *J. Comp. Phys.*, **96**, 193-216.
- Van Duin, A. C. T., S. Dasgupta, F. Lorant, and W. A. Goddard III (2001). ReaxFF: A Reactive Force Field for Hydrocarbons. *J. Phys. Chem. A*, **105**, 9396-9409.
- Van Leeuwen, M. (1996). Prediction of the vapour-liquid coexistence curve of alkanols by molecular simulation. *Molec. Phys.*, **87**, 87-101.
- Visco, D. P. Jr., and D. A. Kofke (1998). Vapor-liquid equilibria and heat effects of hydrogen fluoride from molecular simulation. *J. Chem. Phys.*, **109**, 4015-4027.
- Vlot, M. J., J. Huinink, and J. P. van der Eerden (1999). Free energy calculations on systems of rigid molecules: An application to the TIP4P model of H_2O . *J. Chem. Phys.*, **110**, 55-61.
- Vorontsov-Velyaminov P., A. Broukhno, and T. Kuznetsova (1996). Free Energy Calculation by Expanded Ensemble Method for Lattice and Continuous Polymers. *J. Phys. Chem.*, **100**, 1153-1158.
- Voter, A. (1998). Parallel replica method for dynamics of infrequent events. *Phys. Rev. B*, **57**(22), R13985-R13987.
- Wang, F., and D. P. Landau (2001). Efficient, Multiple-Range Random Walk Algorithm to Calculate the Density of States. *Phys. Rev. Lett.*, **86**, 2050-2053.
- Westmoreland, P. R., P. A. Kollmann, A. M. Chaka, P. T. Cummings, K. Morokuma, M. Neurock, E. B. Stechel, and P. Vashishta (2002). *Applying Molecular and Materials Modeling*. Kluwer Academic Publishers, Dordrecht. See also online version at <http://www.wtec.org/loyola/molmodel/>.
- Wierzchowski S.J., and D. A. Kofke (2003). Hydrogen fluoride phase behavior and molecular structure: Ab initio derived potential models. *J. Chem. Phys.*, **119**, 6092-6099.
- Wilding, N. B., and M. Müller (1994). Accurate measurements of the chemical potential of polymeric systems by Monte-Carlo simulation. *J. Chem. Phys.*, **101**, 4324-4330.
- Wilding, N. B. (1995). Critical point and coexistence-curve properties of the Lennard-Jones fluid: A finite-size scaling study. *Phys. Rev. E*, **52**, 602-611.
- Wilding, N. B. (1997). Critical end point behavior in a binary fluid mixture. *Phys. Rev. E*, **55**, 6624-6631.
- Wilson, A. K., D. E. Woon, K. A. Peterson, and T. H. Dunning, Jr. (1999). Gaussian basis sets for use in correlated molecular calculations. IX. Correlation consistent sets for the atoms gallium through krypton. *J. Chem. Phys.*, **110**, 7667-7676.
- Wood, W. W. (1968). Monte Carlo Calculations for Hard Disks in the Isothermal-Isobaric Ensemble. *J. Chem. Phys.*, **48**, 415-434.
- Yan, Q., and J. J. de Pablo (2003). Fast Calculation of the Density of States of a Fluid by Monte Carlo Simulations, *Phys. Rev. Lett.*, **90**, 035701.