

**Chemical Theory and Computation Special Feature: Ab initio quantum chemistry:
Methodology and applications**

Richard A. Friesner

PNAS 2005;102;6648-6653; originally published online May 3, 2005;
doi:10.1073/pnas.0408036102

This information is current as of September 2006.

Online Information & Services	High-resolution figures, a citation map, links to PubMed and Google Scholar, etc., can be found at: www.pnas.org/cgi/content/full/102/19/6648
References	This article cites 109 articles, 2 of which you can access for free at: www.pnas.org/cgi/content/full/102/19/6648#BIBL This article has been cited by other articles: www.pnas.org/cgi/content/full/102/19/6648#otherarticles
E-mail Alerts	Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or click here .
Subspecialty Collections	This article, along with others on similar topics, appears in the following collection(s): Chemical Theory and Computation Special Feature www.pnas.org/cgi/collection/chem_theory
Rights & Permissions	To reproduce this article in part (figures, tables) or in entirety, see: www.pnas.org/misc/rightperm.shtml
Reprints	To order reprints, see: www.pnas.org/misc/reprints.shtml

Notes:

Ab initio quantum chemistry: Methodology and applications

Richard A. Friesner*

Department of Chemistry, Columbia University, 3000 Broadway, MC 3110, New York, NY 10025

Edited by Bruce J. Berne, Columbia University, New York, NY, and approved March 11, 2005 (received for review November 4, 2004)

This Perspective provides an overview of state-of-the-art *ab initio* quantum chemical methodology and applications. The methods that are discussed include coupled cluster theory, localized second-order Moller–Plesset perturbation theory, multireference perturbation approaches, and density functional theory. The accuracy of each approach for key chemical properties is summarized, and the computational performance is analyzed, emphasizing significant advances in algorithms and implementation over the past decade. Incorporation of a condensed-phase environment by means of mixed quantum mechanical/molecular mechanics or self-consistent reaction field techniques, is presented. A wide range of illustrative applications, focusing on materials science and biology, are discussed briefly.

coupled cluster | density functional theory | second-order Moller–Plesset perturbation theory | mixed quantum/molecular mechanics

Over the past three decades, *ab initio* quantum chemistry has become an essential tool in the study of atoms and molecules and, increasingly, in modeling complex systems such as those arising in biology and materials science. The underlying core technology is computational solution of the electronic Schrodinger equation; given the positions of a collection of atomic nuclei, and the total number of electrons in the system, calculate the electronic energy, electron density, and other properties by means of a well defined, automated approximation (a “model chemistry”). The ability to obtain “good-enough” solutions to the electronic Schrodinger equation for systems containing tens, or even hundreds, of atoms has revolutionized the ability of theoretical chemistry to address important problems in a wide range of disciplines; the Nobel Prize awarded to John Pople and Walter Kohn in 1998 is a reflection of this observation.

In its exact form, the electronic Schrodinger equation is a many-body problem, whose computational complexity grows exponentially with the number of electrons, and hence, a brute force solution is intractable. Hartree–Fock theory, a mean field approach, produces reasonable results for many properties but is incapable of providing a robust description of reactive chemical events in which electron correlation has a major role. Thus, a key problem has been the development of treatments of electron correlation that exhibit a tractable scaling in computational effort with the size of the system. The considerable progress that has been made along these lines is outlined below.

Given a well defined theoretical framework of approximation, the next requirement is efficient computational implementation. Considerable sophistication is required to achieve acceptable accuracy and efficiency; the leading

quantum chemistry programs are millions of lines of computer code, and mathematical algorithms to reduce formal scaling of computational effort with system size have an increasingly crucial role in meeting the challenge of handling complex system relevant to practical applications. Below, the most important computational advances are described briefly, and their impact on the ability to address critical problems is analyzed.

The treatment of large, condensed-phase systems (e.g., proteins in aqueous solution) entirely by *ab initio* methods is extremely expensive computationally. However, it is often the case that a relatively small region of the system can be modeled at the *ab initio* quantum chemical level, whereas the remainder can be treated more approximately [e.g., by means of molecular mechanics (MM) or continuum solvation models]. The technologies for coupling quantum chemical methods to these alternative types of models [mixed quantum mechanics (QM)/MM and self-consistent reaction field (SCRf) approaches] have become an essential component of the theoretical arsenal, enabling realistic modeling of even the most complex molecular structures. The key issues in these types of calculations, as well as a discussion of recent successful implementations, are presented below.

Last, a remarkable range of applications have appeared in the past decade, impacting nearly every aspect of chemistry, biology, and materials science. Space permits discussion of only a small fraction of these calculations, which are presented below as examples of what is possible at the current state of the art. Continued improvements in the theory and implementation, and reduction in the cost/performance of computing, ensure that dramatic progress will con-

tinue to be made in the years to come, advancing toward the ultimate goal of theory achieving full partnership with experiment as both an explanatory and predictive methodology.

Quantum Chemical Theory

Two highly productive approaches to solution of the electronic Schrodinger equation have arisen over the past 50 years. Wavefunction-based approaches expand the electronic wavefunction as a sum of Slater determinants, the orbitals and coefficients of which are optimized by various numerical procedures. Hartree–Fock theory is the simplest method of this type, involving optimization of a single determinant; however, as mentioned above, its usefulness is limited because of complete neglect of electron correlation. We discuss three types of correlated wavefunction-based approaches: second-order Moller–Plesset perturbation theory (MP2), with an emphasis on the localized version of this methodology (1); methods based on the coupled-cluster ansatz, focusing on the widely used CCSD(T) (coupled cluster with single, double, and triple perturbative excitations) variant (2); and multireference perturbation methods, such as CASPT2 (complete active space with second-order perturbation theory) (3). Each of these approximations has a different computational scaling with the number of electrons and is (arguably) the method of choice for different types of problems.

The second class of theoretical approaches are based on density functional

This paper was submitted directly (Track II) to the PNAS office.

Abbreviations: MM, molecular mechanics; QM, quantum mechanics; SCRf, self-consistent reaction field; MP2, second-order Moller–Plesset perturbation theory; DFT, density functional theory.

*E-mail: rich@chem.columbia.edu.

© 2005 by The National Academy of Sciences of the USA

theory (DFT), which, following the Hohenberg–Kohn theorem (4), mandates expression of the total energy of the system as a functional of the electron density. Because the electron density depends on only three coordinates (as opposed to the $3N$ coordinates of N electrons), the computational effort required to solve the equations of DFT is comparable with that required for Hartree–Fock theory, thus rendering DFT highly attractive from the point of view of computational implementation. However, the correct functional of the energy is unknown and has to be constructed by heuristic approximation. Initial functionals, based principally on behavior of the electron gas (5), were lacking in the accuracy required for chemical applications. Breakthroughs over the past two decades (6–10) have led to the development of functionals capable of remarkable accuracy and breadth of applicability across the periodic table, although it is important to note that there remain limitations as well.

At present, there are two principal classes of functionals that have been extensively deployed and tested in large-scale applications as well as small molecule benchmarks: gradient-corrected (7, 8, 10) (e.g., BLYP; ref. 7), and hybrid (7, 9) (e.g., B3LYP; ref. 7), functionals. Gradient-corrected functionals begin with the local-density approximation but add terms involving the gradient of the electron density. Hybrid functionals also incorporate gradient corrections but add an empirically fitted admixture of exact Hartree–Fock exchange. Although there are differences between the individual functionals in each class, the differences between classes are substantially larger; hence, we focus below on one example from each class (BLYP and B3LYP, as cited above) as representative members.

Coupled cluster methods are the most computationally expensive but also, in principle, the most accurate of the approaches that we consider. By an exponential ansatz for the coupling of correlated electron pairs, highly excited determinants are incorporated into the wavefunction without the combinatorial explosion of effort implied by naïve configuration–interaction approaches; at present, the CCSD(T) approach (in which triple excitations are treated perturbatively) provides the best trade-off of accuracy and efficiency (2). The formal scaling with number of electrons N of CCSD(T) is N^7 ; for technical reasons, it has been difficult to improve on this behavior in practice. Thus, calculations are limited to small- to medium-sized molecules, and parallel supercomputers are required at the upper end of this range.

In return for this high computational cost, the level of accuracy that one can expect from a benchmark CCSD(T) calculation for atomization energies (and other less demanding thermochemical quantities) of small organic molecules is on the order of a few tenths of 1 kcal/mol (1 cal = 4.18 J) (11). Equilibrium bond lengths and bond angles are typically highly accurate (12). Indeed, the level of precision in achieved in geometry is unnecessary for energetic accuracy; hence, less expensive methods are often used to compute structures. Excited-state methods have been developed and have displayed encouraging performance (13).

Turning next to MP2 methods, the computational cost is reduced considerably. Whereas formal scaling is N^5 , numerical methods developed in the past decade (discussed further below) yield substantial reductions in scaling both formally and asymptotically (as well as significant absolute central processing unit reductions for smaller systems), enabling systems containing a few hundred atoms to be treated routinely. The performance of MP2 methods for properties involving making or breaking electron pairs is substantially inferior to that of CCSD(T) and also not as good as the best DFT methods, although recent results have come closer to DFT benchmarks (14). However, for most systems, MP2 does extremely well for nonbonded interactions and internal conformational energetics, typically delivering accuracy within ≈ 0.3 kcal/mol if one extrapolates to the basis set limit (15, 16, 23). A recent interesting exception has been noted for stacking interactions of the benzene dimer, where CCSD(T) corrections are required to avoid errors in the range of 0.5–1.0 kcal/mol (17); more investigation will be required to enumerate other outliers of this type.

Multireference perturbation methods, such as CASPT2 (3), are more difficult to use than either CCSD(T) or MP2 in that they require a careful definition of the active space of electrons to be treated at the multireference level to be effective (electrons not in the active space are treated perturbatively). The computational cost increases exponentially with the size of the active space. Nevertheless, there are some problems where CCSD(T) methods have severe difficulties because of the fact that the electronic wavefunction is no longer dominated by a single (Hartree–Fock) determinant; an obvious example is generating a complete potential energy curve in the process of breaking a bond (18). In this and other cases, CASPT2 provides a reasonable alternative enabling incorporation of multireference character from the start, and it may be the method of choice in many such cases.

Both gradient correct and hybrid DFT methods achieve excellent results for equilibrium geometries (19, 20) across the periodic table. For atomization energies, there is a clear difference in performance; for example, the BLYP functional has an average error of 7.09 kcal/mol for the G2 set of 148 small molecules, whereas the B3LYP functional has an average error of 3.11 kcal/mol for the same data set (21). For transition-metal-containing systems, the experimental data are less reliable, and there are many different types of bonding to consider. However, in typical situations, the performance of B3LYP is clearly respectable, with average errors for bond energies of small transition metal complexes in the range of 3–5 kcal/mol (82). Thus, hybrid DFT provides a quantitative means of investigating reactive chemistry in large systems, albeit with large errors for a small number of outliers.

However, DFT is inferior to basis set limit MP2 in achieving high precision for energy differences in which bonds are not made or broken, including hydrogen bonding and conformational energetics (22, 23). It is not the case that DFT yields bad results for these problems (indeed, it is clearly an improvement over Hartree–Fock); it simply is less accurate (e.g., by ≈ 0.5 –1 kcal/mol for hydrogen bonding energies) than MP2 when averaged over many cases. Furthermore, the ability of DFT to model dispersion (van der Waals) interaction has not yet been established (15); again, MP2 represents a qualitatively superior alternative, which can itself be applied to large systems at a quantitatively (but not qualitatively) larger computational cost. Last, the time-dependent version of DFT (TDDFT) in many cases provides reasonable results for excited states (24–26) and is considerably less expensive than alternatives such as CASPT2- and CCSD(T)-based excited-state methods.

We can summarize the above results by considering which quantum chemistry methodology is the current method of choice for various types of computational chemistry modeling problems. For accurate studies of small molecules in the gas phase, CCSD(T) and related methods are the approach of choice, perhaps combined with less expensive methods to carry out geometry optimization and/or obtain basis set convergence. For investigation of reactive chemistry in medium to large systems, DFT is at present the best approach, in some cases it is possible to check a truncated model of the actual system by using coupled-cluster-based methods (although typically not at a benchmark level). MP2 methods are generally useful for studying nonbonded interactions in both small and large systems and in developing molecular mechanics

force fields (discussed further below); these types of calculations are highly relevant in biology and materials science problems. DFT can be used also in the investigation of conformational and hydrogen bond interactions, provided that the accuracy limitations are understood. Last, there are a number of choices for excited states, none of which are completely satisfactory but several of which have yielded respectable results. This area should benefit substantially from development efforts.

Computational Implementation

Below, we briefly discuss state-of-the-art computational implementation for each of the four quantum chemical models presented above, beginning with the least expensive (DFT). There are two fundamentally different numerical approaches that are used currently for solving the Kohn–Sham equations (27) as is required in large-scale DFT calculations. The first approach, arising primarily from the physics community, represents the electron density orbitals by a plane wave basis set (25). To date, efficient use of such methods has been possible only for gradient-corrected (as opposed to hybrid) DFT functionals, because the matrix elements of the exact exchange operator are difficult to evaluate in a plane wave basis set. The approach is most naturally applied to condensed-phase systems (periodic solids or liquids) in which the imposition of periodic boundary conditions (implicit in the use of plane waves, which are periodic) is appropriate.

Plane-wave methods have been implemented by two different algorithmic approaches. Car–Parinello methods (28–31) mandate performing *ab initio* molecular dynamics, evolving the system by means of an extended Lagrangian formalism. This type of approach is most useful in studying dynamics in liquids and molten solids, as well as in locating stationary points by means of simulated annealing in systems in which it is difficult to make a good initial guess. Conjugate gradient-based optimization approaches, implemented in programs such as CASTEP (30) and VASP (31) are typically used to study solid state and surface science problems.

The alternative to plane-wave methods is the use of localized, atom-centered Gaussian basis sets, as has been used for many years in solving the Hartree–Fock equations. Indeed, most localized basis implementations of DFT were developed by modifying existing Hartree–Fock codes, as was done, for example, in the Gaussian series of programs (32). The use of Gaussian basis functions allows hybrid DFT functionals to be treated in straightforward fashion.

Over the past decade, the following advances have increased the efficiency of Gaussian-based DFT methods substantially:

1. Numerical quadrature methods were developed to perform integrals accurately over the exchange–correlation functional, which cannot be handled analytically (33, 34).
2. The computation of cost of the Coulomb term with system size can be reduced from N^2 to $\approx N$ by the use of multipole expansions (35–37).
3. Two numerical approaches to reducing the formal scaling (as opposed to asymptotic scaling) of the Coulomb and exchange terms have been developed. The first approach, pseudospectral methods (37, 38), were introduced more than two decades ago, and are robustly implemented in the JAGUAR program (39). The second approach, resolution of the identity, is technically different in its details but similar in spirit. Resolution-of-the-identity methods (36) are implemented efficiently in the TURBOMOLE program (40). Both approaches in their current form deal effectively with the Coulomb term; pseudospectral methods handle exact exchange (and, thus, hybrid DFT functionals) with equal effectiveness. Both approaches provide substantial central processing unit reductions for medium to large systems as compared with conventional two-electron integral methodology.

Two fundamental advances have dramatically reduced the computational cost, and scaling with system size, of MP2 calculations as compared with the original formulations. The first advance is the development of localized MP2 (LMP2) theory several decades ago (1). In this approach, the Hartree–Fock orbitals are localized to bonds and lone pairs, and electrons are correlated starting from these orbitals rather than the delocalized canonical Hartree–Fock orbitals. In principle, this methodology should lead to substantial reductions in computational effort, because an electron in a localized orbital can be correlated by excitation into a localized virtual space. In practice, there are technical difficulties in realizing these theoretical gains by using standard transforms of analytical two-electron integrals. However, the combination of localized methods with pseudospectral (23) (and, more recently, resolution of the identity; ref. 41) methods enables, as in the case of DFT, a reduction of the formal scaling, in this case from N^5 to N^3 (asymptotically approaching N^2 and even N at very large distances), and the

concomitant ability to treat large systems with computational effort that is a small multiple of DFT, at least for single point calculations. Efficient LMP2 methods are implemented in the JAGUAR (39) and QCHEM (42) programs, as well as in codes described in ref. 41.

It has proven to be more difficult to reduce the formal, or asymptotic, scaling, of coupled cluster and multireference perturbation methods, although this area is of considerable active research interest (discussed further in *Conclusion and Future Directions*). Details of the technology in both cases have improved substantially, and application of these methods to small to medium-sized molecules in the gas phase is now routine, although requirements for computational power rise steeply as the size of the system increases. Efficient coupled-cluster methods are implemented in the GAUSSIAN (32) and QCHEM (42) programs; the MOLCAS program (43) contains an optimized CASPT2 methodology.

All four fundamental approaches have benefited from the tremendous advances in computational hardware over the past decade. Localized basis function DFT and local MP2 calculations can be performed routinely for systems with hundreds of atoms on personal computers, which offer unparalleled cost/performance savings. Plane-wave methods require tightly coupled parallel machines if large unit cells are to be handled, increasing their actual cost in dollar terms significantly beyond what would be assessed from reported central processing unit times. Large-scale coupled-cluster and multireference calculations also require parallel machines if the upper end of the accessible range of molecules is to be addressed.

Last, a crucial issue in obtaining accurate quantum chemical results is convergence of desired quantities (e.g., atomization energies) with basis set size, particularly for wavefunction-based methods. One approach is to employ composite methods in which each challenging aspect of an *ab initio* calculation (geometry optimization, basis set convergence, inclusion of high-level correlation) can be individually optimized with the most cost-effective technology, and the results combined (typically via a simple linear scheme, often with an additive correction to mimic the effect of high angular momentum functions) to yield a level of accuracy not far from that obtainable from employing the most expensive type of calculation across the board. Typical methods, such as the widely used G2 (21) and G3 (44) theory, incorporate a coupled-cluster component, and are capable of

achieving 1–2 kcal/mol average errors in atomization energies at a cost far smaller than benchmark CCSD (T) calculations with large basis sets. A number of alternative approaches of this type have also been developed, some of which appear to yield impressive cost/performance improvements (45). An alternative approach is to employ extrapolation methods, which enable the effects of very large basis sets to be inferred from a series of smaller basis set calculations (46, 47).

Condensed-Phase Environments: SCRF and Mixed Quantum/Classical Methods

The remarkable advances in *ab initio* quantum chemical methodology and implementation described above should not obscure the fact that such calculations are still very expensive computationally as compared with more approximate models. Evaluation of the energy of a configuration of a 50-atom molecule by using a molecular mechanics potential function requires substantially <1 ms on a modern personal computer, as compared with the several minutes needed even at the DFT level for a single-point electronic structure calculation. The huge difference in multiplicative coefficient persists, regardless of asymptotic scaling, as one goes to larger systems; this fact makes quantum chemical calculations for systems containing thousands or tens of thousands of atoms impractical, particularly if significant numbers of configurations need to be studied. Furthermore, such heroic efforts are also unnecessary in many applications; if the reactive event of interest is localized (as is always the case in biological systems and often the case in materials problems), then treatment of all but a small reactive region via a classical force field is, in general, an excellent approximation, reproducing the structural and electrostatic effects of the environment on the reactive chemistry with a reasonable degree of fidelity.

Mixed QM/MM methods (48) provide a means of combining quantum chemical and molecular mechanics models in a way that preserves the integrity of the overall energy function. The key issue is the interface between the QM and MM regions. If the QM and MM regions are not covalently linked, then the problem is relatively straightforward; introduction of van der Waals parameters onto the QM atoms, and optimization of these parameters to fit fully QM hydrogen bonding energies (49), will yield a model that does a reasonable job of reproducing the fully QM results over the entire phase space. However, when there are covalent connections between regions (for example, when part of a protein active site needs to be treated at the QM level), definition of the

interface requires more sophisticated technology.

Over the past decade, two approaches have emerged which have enabled significant progress to be made in developing accurate QM/MM interfaces. The first approach employs link atoms, which are fictitious atoms that are used to cap the QM and MM regions (50–52). The second uses frozen localized orbitals, derived from model molecules, at the interface between a QM and MM atom (49, 53). The quality of the results depends not only on the functional form of the interface (probably, both methods are capable of achieving good results if appropriately optimized) (54) but also on the details of the parameterization. The most accurate results are achieved by making the parameters explicitly depend on the local chemical environment; for example, in ref. 49, parameters are developed for each amino acid side chain, with cuts between the α -carbon and β -carbon as well as in the peptide backbone. Such specific parametrization is capable of yielding a robust model with average errors on the order of 0.5–1.0 kcal/mol as compared with a fully quantum chemical results, for various tests (conformational energetics, deprotonation energies, and more complex assessments in which the size of the MM region is increased systematically). For the most part, these errors are smaller than the intrinsic errors in the quantum chemical methods and, hence, are unlikely to be the limiting factor in achieving high accuracy.

In principle, the computational cost of a QM/MM single point calculation should be nearly identical to that of a QM calculation of the same size as the QM region because the MM region contributes negligibly. For stationary point optimization, a problem arises because of the fact that a large system, containing thousands of atoms, will ordinarily take many more gradient cycles to optimize than a system containing hundreds of atoms. However, this problem can be solved by adiabatically minimizing the MM region after every QM geometry step (49, 55, 56). This strategy renders a QM/MM optimization that is approximately as expensive as the corresponding optimization of the QM region alone, making such optimizations feasible for large systems.

The QM/MM approach is ideally suited to studying reactions in proteins, disordered solids, large molecules, and other systems that have well defined structures and are heterogeneous in character. However, for processes in solution, it is necessary to average extensively over the positions of the solvent molecules to properly take into account

environmental effects. Thus, QM/MM approaches are possible but computationally very expensive. An alternative that delivers reasonable accuracy at a much lower computational cost is the use of continuum models to describe the solvent. The coupling of continuum models with quantum chemical calculations using SCRF approaches (57, 58) has been implemented over the past decade in a number of widely available *ab initio* quantum chemistry programs (e.g., GAUSSIAN, JAGUAR, and QCHEM). By alternating cycles of continuum Poisson–Boltzmann (59) (or approximations of such cycles; ref. 60) and quantum chemical computations, the interacting reaction field of the solvent and charge distribution of the solute are iterated to convergence. Geometry optimization of structures in solution is also possible by using analytical gradient methods (61, 62). To achieve accurate results for solvation-free energies, parameters of the continuum model must be fit to small molecule experimental data (60, 63). SCRF-based methods can also be used effectively for pK_a prediction (64). Last, it is possible to combine QM/MM and continuum solvation approaches, although few articles combining these approaches have been published.

Applications

For small molecules in the gas phase and in solution, *ab initio* quantum chemical calculations can provide results approaching benchmark accuracy (11), and they are used routinely to complement experimental studies. A wide variety of properties, including structures (12), thermochemistry (11) (including activation barriers; ref. 65), spectroscopic quantities of various types (13, 66), and responses to external perturbations (67), can be computed effectively. As discussed above, SCRF methods (or simply ignoring the solvent entirely, an approximation that is sometimes acceptable, particularly in nonpolar solvents or when a quantity that is insensitive to the dielectric of the environment is being computed) enable a relatively straightforward extension of gas-phase quantum chemical methods to obtaining results for molecules in solution.

Materials-science applications have grown exponentially over the past decade and now involve exploration of highly complex structures and chemistries. Both bulk and surface properties can be computed for solids (68), in many cases yielding excellent agreement with experiment. Bandgaps and optical properties of solids have been addressed successfully by DFT-based methods (69). The interactions of small molecules with surfaces has also been the subject of

extensive investigation, with the focus on structures, binding energies, and catalytic chemistry (70–72). Calculations of this type are central to practical applications involving industrially relevant catalytic processes, semiconductor processing, and modeling of conductivity for microelectronics applications. Another significant area of both theoretical and practical interest is modeling photochemical applications (e.g., solar energy conversion by dyes adsorbed onto titanium dioxide; ref. 20), as implemented in the Graetzel cell (73), which has achieved a conversion yield very close to that of amorphous silicon.

Applications to nanotechnology are relatively recent but represent a focus of increasing interest. Among the systems that have been investigated are carbon nanostructures (buckyballs and nanotubes) (25, 74–76) and semiconductor quantum dots (26, 77, 78). The effects of quantum confinement and structural modification imposed by various nanostructures on chemical and electrical properties can be investigated in a rigorous fashion, which will be essential as attempts are made to use these materials in various types of devices.

Biological applications have focused principally on the modeling of enzymatic catalysis and active-site chemistry (79–88), although interesting investigations of other phenomena, such as cooperativity in backbone hydrogen bonding (89) and modeling of β -sheet formation propensities by a periodic DFT calculation (90), have been performed recently. The use of QM/MM methods permits incorporation of the full protein environment, which is crucial in an important subset of cases, as has recently been shown in cytochrome P450 (79), for example. A wide variety of enzymatic systems have been investigated, with quantitative comparison with experimental structural and thermodynamic data yielding an encouraging level of agreement in the most accurate modeling efforts. *Ab initio* calculations provide a wealth of detail that is not available from experiment and a degree of confidence in the results that is not available from more empirical approaches. In systems such as methane monooxygenase, in which a substantial number of careful calculations have been performed by several groups (81, 91–96) and close contact with experiment has been achieved for many aspects of the catalytic process, a comprehensive picture of the functioning of the enzyme is beginning to be assembled by means of the interaction of theory and experiment. Last, the coupling of *ab initio* quantum QM/MM methods with simulation and protein structure prediction techniques

permits investigation of events in which reactive chemistry is coupled to substantial conformational changes, such as the catalytic loop motion in triose phosphate isomerase (82).

Last, a key application of quantum chemical methods is in the development of molecular mechanics force fields (97–99). The quality of force fields has improved enormously in the past two decades, and a principal reason has been the ability to fit parameters to the results of high-level *ab initio* calculations for larger and more complex model systems. For fixed-charge force fields, torsional parameters can be fit to computations of conformational energy differences of model molecules, whereas charge distributions can be fit to quantum chemical electrostatic potentials. Dimer-interaction energies have typically been used as a heuristic guide to modeling nonbonded interactions; however, for a polarizable force field, these quantities can be fit directly as well, because the model is supposed to represent nonbonded interactions accurately in both the gas phase and the condensed phase (100, 101). Polarizability parameters can also be fit to quantum chemical data, although there are some issues associated with basis set size that must be considered when doing so (100).

At this point, it is useful to summarize how advances in theory and computational implementation have enabled the tremendous growth in the important applications described above. Small molecules in the gas phase are now typically addressed by high-level methods such as CCSD(T), which in many cases are more accurate than experiment. As systems increase in size, one has to at some point switch over to DFT and/or MP2. Multireference perturbation methods are typically applied when there are particular difficulties associated with near degeneracies, or for properties such as excited states for which specific multireference approaches (such as CASPT2) have proven to be superior to DFT in a wider subset of cases. The overwhelming majority of large-scale material science and biological applications have been performed with DFT; this state of affairs is mandated by the large size of the systems that are being considered, the need in some cases for periodic boundary conditions, and the availability of QM/MM and SCRF methods principally for DFT based approaches. Last, for force-field development, localized MP2 is the method of choice because of its acceptable treatment of dispersion interaction and higher accuracy for conformational energies and hydrogen bonding energies than DFT while main-

taining a similar (if somewhat larger) computational cost.

Conclusion and Future Directions

Over the next decade, we expect to see significant increases in accuracy in all four of the principal quantum chemical directions described above. Progress in developing DFT functionals has been difficult since the breakthroughs of the early 1990s; however, a number of promising approaches (102–105) and key systematic difficulties, such as problems in predicting barrier heights of small molecule radical reactions (106), have been identified. Coupled-cluster and multireference algorithms include attempts to use localized reference states that potentially could lead to the same huge reductions in scaling with system size that have been realized with localized MP2 (107–110). Continued reductions in the cost/performance of computing and improvements in algorithmic details should continue to yield shorter time to solution for increasingly larger systems.

Similarly, improvement can be expected in treatment of the condensed phase environment. Optimization of the accuracy of continuum solvation methods is far from a solved problem; furthermore, there is some evidence that the inclusion of a small number of explicit water molecule can improve results (111), but methods of this type must be formulated very carefully to avoid double counting. QM/MM methods can be made more accurate, robust, cost-effective, and easy to use. Last, sampling algorithms in the condensed phase are crucial for many large-scale applications, and significant advances can be expected as more ambitious problems are addressed.

Even if the technology were to stand still, one would expect a large number of important new applications to be carried out over the next decade; with advances in theory, software, and computational hardware, larger data sets and systems of increasing size, will be amenable to study. However, the most exciting possibility is that the parallel advances in theory and experiment will enable fully explanatory and predictive models to be constructed for the complex, condensed-phase processes that govern most of the natural world. *Ab initio* quantum chemical methods are not the only technology that will be a component of such a development, but they surely are an essential one, as demonstrated by the progress and prospects outlined above.

This work was supported in part by National Institutes of Health Grant GM 40526 and Department of Energy Grant DE FG02 90ER 141.

1. Sabeo, S. & Pulay, P. (1991) *Annu. Rev. Phys. Chem.* **44**, 213–236.
2. Bartlett, R. J. (1989) *J. Phys. Chem.* **93**, 1697–1708.
3. Roos, B. O., Andersson, K., Fulscher, M. P., Malmqvist, P. A., Serrano-Andres, L., Pierloot, K. & Merchán, M. (1996) *Adv. Chem. Phys.* **93**, 219–331.
4. Hohenberg, P. & Kohn, W. (1964) *Phys. Rev.* **132B**, B864–B871.
5. Sham, L. J. & Kohn, W. (1964) *Phys. Rev.* **145**, 561–567.
6. Kohn, W., Becke, A. D. & Parr, R. G. (1996) *J. Phys. Chem.* **100**, 12974–12980.
7. Johnson, B. G., Gill, P. M. W. & Pople, J. A. (1993) *J. Chem. Phys.* **98**, 5612–5626.
8. Becke, A. D. (1998) *Phys. Rev. A* **38**, 3098–3100.
9. Becke, A. D. (1993) *J. Chem. Phys.* **98**, 5648–5652.
10. Lee, C. T., Yang, W. T. & Parr, R. G. (1988) *Phys. Rev. B* **37**, 785–789.
11. Martin, J. M. L. & de Oliveira, G. (1999) *J. Chem. Phys.* **111**, 1843–1856.
12. Thomas, J. R., Deeeeuw, B. J. Vacek, G., Crawford, T. D., Yamaguchi, Y. & Schaefer, H. F. (1993) *J. Chem. Phys.* **99**, 403–416.
13. Stanton, J. F. & Bartlett, R. J. (1993) *J. Chem. Phys.* **98**, 7029–7039.
14. Grimme, S. (2003) *J. Chem. Phys.* **118**, 9095–9102.
15. Tsuzuki, S., Uchimaru, T. & Tanabe, K. (1998) *Chem. Phys. Lett.* **287**, 202–208.
16. Tsuzuki, S., Uchimaru, T., Matsumura, K., Mikami, M. & Tanabe, K. (1999) *J. Chem. Phys.* **110**, 11906–11910.
17. Sinnokrot, M. O., Valeev, E. F. & Sherrill, C. D. (2002) *J. Am. Chem. Soc.* **124**, 10887–10893.
18. Dutta, A. & Sherrill, C. D. (2003) *J. Chem. Phys.* **118**, 1610–1619.
19. Murray, C. W., Laming, G. J., Handy, N. C. & Amos, R. D. (1992) *Chem. Phys. Lett.* **199**, 551–556.
20. Redfern, P. C., Zapol, P., Curtiss, L. A., Rajh, T. & Thurnauer, M. C. (2003) *J. Phys. Chem. B* **107**, 11419–11427.
21. Curtiss, L. A., Raghavachari, K., Redfern, P. C. & Pople, J. A. (1997) *J. Chem. Phys.* **106**, 1063–1079.
22. Tsuzuki, S. & Luthi, H. P. (2001) *J. Chem. Phys.* **114**, 3949–3957.
23. Murphy, R. B., Beach, M. D., Friesner, R. A. & Ringnalda, M. N. (1995) *J. Chem. Phys.* **103**, 1481–1490.
24. Stratmann, R. E., Scuseria, G. E. & Frisch, M. J. (1998) *J. Chem. Phys.* **109**, 8218–8224.
25. Bauernschmitt, R., Ahlrichs, R., Hennrich, F. H. & Kappes, M. M. (1998) *J. Am. Chem. Soc.* **120**, 5052–5059.
26. Baroufalis, C. S., Zdzetsis, D. & Grimme, S. (2001) *Phys. Rev. Lett.* **87**, 276402.
27. Kohn, W. & Sham, L. J. (1965) *Phys. Rev.* **140**, 1133.
28. Car, R. & Parrinello, M. (1985) *Phys. Rev. Lett.* **55**, 2471–2474.
29. Payne, M. C., Teter, M. P., Allan, D. C., Arias, T. A. & Joannopoulos, J. D. (1992) *Rev. Mod. Phys.* **64**, 1045–1097.
30. Segall, M. D., Lindan, P. J. D., Probert, M. J., Pickard, C. J., Hasnip, P. J., Clark, S. J. & Payne, M. C. (2002) *J. Phys. Cond. Matt.* **14**, 2717–2744.
31. Kresse, G. & Furthmüller, J. (1996) *Phys. Rev. B* **54**, 11169–11186.
32. Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Montgomery, J. A., Jr., Vreven, T., Kudin, K. N., Burant, J. M., et al. (2003) GAUSSIAN 03 (Gaussian, Pittsburgh), Revision B.02.
33. Gill, P. M. W., Johnson, B. G. & Pople, J. A. (1993) *Chem. Phys. Lett.* **209**, 506–512.
34. Treutler, O. & Ahlrichs, R. (1995) *J. Chem. Phys.* **102**, 346–354.
35. White, C. A., Johnson, B. G., Gill, P. M. W. & Head-Gordon, M. (1996) *Chem. Phys. Lett.* **253**, 268–278.
36. Sierka, M., Hogeckamp, A. & Ahlrichs, R. (2003) *J. Chem. Phys.* **118**, 9136–9148.
37. Murphy, R. B., Cao, Y., Beachy, M. D., Ringnalda, M. N. & Friesner, R. A. (2000) *J. Chem. Phys.* **112**, 10131–10141.
38. Friesner, R. A., Murphy, R. B., Beachy, M. D., Ringnalda, M. N., Pollard, W. T., Dunietz, B. D. & Cao, Y. X. (1999) *J. Phys. Chem. A* **103**, 1913–1928.
39. Beachy, M., Murphy, R., Philipp, D., Mainz, D., Ringnalda, M., Cao, Y., Greeley, H., Pollard, T., Braden, D., Friesner, R., et al. (2004) JAGUAR (Schrödinger, Portland, OR), Version 4.0.
40. Ahlrichs, R., Bar, M., Haser, M., Horn, H. & Kolmel, C. (1989) *Chem. Phys. Lett.* **162**, 165–169.
41. Werner, H. J., Manby, F. R. & Knowles, P. J. (2003) *J. Chem. Phys.* **118**, 8149–8160.
42. Kong, J., White, C. A., Krylov, A. I., Sherrill, D., Adamson, R. D., Furlani, T. R., Lee, M. S., Lee, A. M., Gwaltney, S. R., Adams, T. R., et al. (2000) *J. Comput. Chem.* **21**, 1532–1548.
43. Karlstrom, G., Lindh, R., Malmqvist, P. A., Roos, B. O., Ryde, U., Veryazov, V., Widmark, P. O., Cossi, M., Schimmelpfennig, B., Neogrady, P., et al. (2003) *Comput. Mat. Sci.* **28**, 222–239.
44. Curtiss, L. A., Raghavachari, K., Redfern, P. C., Rassolov, V. & Pople, J. A. (1998) *J. Chem. Phys.* **109**, 7764–7776.
45. Lynch, B. & Truhlar, D. (2003) *J. Phys. Chem. A* **107**, 3898–3906.
46. Truhlar, D. G. (1998) *Chem. Phys. Lett.* **294**, 45–48.
47. Ochterski, J. W., Petersson, G. A. & Montgomery, J. A. (1996) *J. Chem. Phys.* **104**, 2598–2619.
48. Field, M. J., Bash, P. A. & Karplus, M. (1990) *J. Comput. Chem.* **11**, 700–733.
49. Murphy, R. B., Philipp, D. M. & Friesner, R. A. (2000) *J. Comput. Chem.* **21**, 1442–1457.
50. Das, D., Eurenios, K. P., Billings, E. M., Sherwood, P., Chatfield, D. C., Hodoscek, M. & Brooks, B. R. (2002) *J. Chem. Phys.* **117**, 10534–10547.
51. Amara, P. & Field, M. J. (2003) *Theor. Chem. Acc.* **109**, 43–52.
52. Maseras, F. & Morokuma, K. (1995) *J. Comput. Chem.* **16**, 1170–1179.
53. Gao, J. L., Amara, P., Alhambra, C. & Field, M. J. (1998) *J. Phys. Chem. A* **102**, 4714–4721.
54. Reuter, N., Dejaegere, A., Maignet, B. & Karplus, M. (2000) *J. Phys. Chem. A* **104**, 1720–1735.
55. Vreven, T., Morokuma, K., Farkas, O., Schlegel, H. B. & Frisch, M. J. (2003) *J. Comput. Chem.* **24**, 760–769.
56. Zhang, Y. K., Liu, H. Y. & Yang, W. T. (2000) *J. Chem. Phys.* **112**, 3483–3492.
57. Tannor, D. J., Marten, B., Murphy, R., Friesner, R. A., Sitkoff, D., Nicholls, A., Ringnalda, M., Goddard, W. A. & Honig, B. (1994) *J. Am. Chem. Soc.* **116**, 11875–11882.
58. Tomasi, J. & Persico, M. (1994) *Chem. Rev.* **94**, 2027–2094.
59. Honig, B. & Nicholls, A. (1995) *Science* **268**, 1144–1149.
60. Klamt, A., Jonas, V., Burger, T. & Lohrenz, J. C. W. (1998) *J. Phys. Chem. A* **102**, 5074–5085.
61. Barone, V., Cossi, M. & Tomasi, J. (1998) *J. Comput. Chem.* **19**, 404–417.
62. Cortis, C. M., Langlois, J. M., Beachy, M. D. & Friesner, R. A. (1996) *J. Chem. Phys.* **105**, 5472–5484.
63. Marten, B., Kim, K., Cortis, C., Friesner, R. A., Murphy, R. B., Ringnalda, M. N., Sitkoff, D. & Honig, B. (1996) *J. Phys. Chem.* **100**, 11775–11788.
64. Klicic, J. J., Friesner, R. A., Liu, S. Y. & Guida, W. C. (2002) *J. Phys. Chem. A* **106**, 1327–1335.
65. Guner, V., Khuong, K. S., Leach, A. G., Lee, P. S., Bartberger, M. D. & Houk, K. N. (2003) *J. Phys. Chem. A* **107**, 11445–11459.
66. Gauss, J. & Stanton, J. F. (1996) *J. Chem. Phys.* **104**, 2574–2583.
67. Sekino, H. & Bartlett, R. J. (1993) *J. Chem. Phys.* **98**, 3022–3037.
68. Milman, V., Winkler, B., White, J. A., Pickard, C. J., Payne, M. C., Akhmatkaya, E. V. & Nobes, R. H. (2000) *Int. J. Quantum Chem.* **77**, 895–910.
69. Rohlfing, M. & Louie, S. G. (2000) *Phys. Rev. B* **62**, 4927–4944.
70. Hammer, B. & Norskov, J. K. (2000) *Adv. Catal.* **45**, 71–129.
71. Wu, C. J. & Carter, E. A. (1991) *Chem. Phys. Lett.* **185**, 172–178.
72. Raghavachari, K. & Halls, M. D. (2004) *Mol. Phys.* **102**, 381–393.
73. Oregan, B. & Gratzel, M. (1991) *Nature* **353**, 737–740.
74. Zhou, Z. Y., Steigerwald, M., Hybertsen, M., Brus, L. & Friesner, R. A. (2004) *J. Am. Chem. Soc.* **126**, 3597–3607.
75. Bettinger, H. F., Yakobson, B. I. & Scuseria, G. E. (2003) *J. Am. Chem. Soc.* **125**, 5572–5580.
76. Jhi, S. H., Louie, S. G. & Cohen, M. L. (2000) *Phys. Rev. Lett.* **85**, 1710–1713.
77. Zhou, Z. Y., Friesner, R. A. & Brus, L. (2003) *J. Am. Chem. Soc.* **125**, 15599–15607.
78. Melnikov, D. V. & Chelikowsky, J. R. (2004) *Phys. Rev. Lett.* **92**, 046802.
79. Guallar, V. & Friesner, R. A. (2004) *J. Am. Chem. Soc.* **126**, 8501–8508.
80. Gherman, B. F., Goldberg, S. D., Cornish, V. W. & Friesner, R. A. (2004) *J. Am. Chem. Soc.* **126**, 7652–7664.
81. Gherman, B. F., Baik, M. H., Lippard, S. J. & Friesner, R. A. (2004) *J. Am. Chem. Soc.* **126**, 2978–2990.
82. Guallar, V., Jacobson, M., McDermott, A. & Friesner, R. A. (2004) *J. Mol. Biol.* **337**, 227–239.
83. Siegbahn, P. E. M. & Blomberg, M. R. A. (2000) *Chem. Rev.* **100**, 421–437.
84. Siegbahn, P. E. M., Blomberg, M. R. A. & Blomberg, M. L. (2003) *J. Phys. Chem. B* **107**, 10946–10955.
85. Cui, Q. & Karplus, M. (2001) *J. Am. Chem. Soc.* **123**, 2284–2290.
86. Rothlisberger, U., Carloni, P., Doclo, K. & Parrinello, M. (2000) *J. Biol. Inorg. Chem.* **5**, 236–250.
87. Wu, N., Mo, Y. R., Gao, J. L. & Pai, E. F. (2000) *Proc. Natl. Acad. Sci. USA* **97**, 2017–2022.
88. Amara, P., Volbeda, A., Fontecilla-Camps, J. C. & Field, M. J. (1999) *J. Am. Chem. Soc.* **121**, 4468–4477.
89. Wiczorek, R. & Dannenberg, J. J. (2003) *J. Am. Chem. Soc.* **125**, 8124–8129.
90. Rossmeisil, J., Kristensen, T., Gregersen, M., Jacobsen, K. W. & Norskov, J. K. (2003) *J. Am. Chem. Soc.* **125**, 16383–16386.
91. Siegbahn, P. E. M. (2001) *J. Biol. Inorg. Chem.* **6**, 27–45.
92. Basch, H., Mogi, K., Musaev, D. G. & Morokuma, K. (1999) *J. Am. Chem. Soc.* **121**, 7249–7256.
93. Hoffmann, M., Khavrutskii, I. V., Musaev, D. G. & Morokuma, K. (2004) *Int. J. Quantum Chem.* **99**, 972–980.
94. Guallar, V., Gherman, B. F., Miller, W. H., Lippard, S. J. & Friesner, R. A. (2002) *J. Am. Chem. Soc.* **124**, 3377–3384.
95. Dunietz, B. D., Beachy, M. D., Cao, Y. X., Whittington, D. A., Lippard, S. J. & Friesner, R. A. (2000) *J. Am. Chem. Soc.* **122**, 2828–2839.
96. Siegbahn, P. E. M. & Crabtree, R. H. (1997) *J. Am. Chem. Soc.* **119**, 3103–3113.
97. Kaminski, G. A., Friesner, R. A., Tirado-Rives, J. & Jorgensen, W. L. (2001) *J. Phys. Chem. B* **105**, 6474–6487.
98. MacKerell, A. D., Bashford, D., Bellott, M., Dunbrack, R. L., Evanseck, J. D., Field, M. J., Fischer, S., Gao, J., Guo, H., Ha, S., et al. (1998) *J. Phys. Chem. B* **102**, 3586–3616.
99. Wang, J. M., Wolf, R. M., Caldwell, J. W., Kollman, P. A. & Case, D. A. (2004) *J. Comput. Chem.* **25**, 1157–1174.
100. Kaminski, G. A., Stern, H. A., Berne, B. J. & Friesner, R. A. (2004) *J. Phys. Chem. A* **108**, 621–627.
101. Kaminski, G. A., Stern, H. A., Berne, B. J., Friesner, R. A., Cao, Y. X., Murphy, R. B., Zhou, R. H. & Halgren, T. A. (2002) *J. Comput. Chem.* **23**, 1515–1531.
102. Becke, A. D. (2003) *J. Chem. Phys.* **119**, 2972–2977.
103. Tozer, D. J. & Handy, N. C. (1998) *J. Chem. Phys.* **109**, 10180–10189.
104. Tao, J. M., Perdew, J. P., Staroverov, V. N. & Scuseria, G. E. (2003) *Phys. Rev. Lett.* **91**, 146401.
105. Zhao, Y., Lynch, B. J. & Truhlar, D. G. (2004) *J. Phys. Chem. A* **108**, 4786–4791.
106. Lynch, B. J., Fast, P. L., Harris, M. & Truhlar, D. G. (2000) *J. Phys. Chem. A* **104**, 4811–4815.
107. Van Voorhis, T. & Head-Gordon, M. (2002) *J. Chem. Phys.* **117**, 9190–9201.
108. Flocke, N. & Bartlett, R. J. (2003) *Chem. Phys. Lett.* **367**, 80–89.
109. Schutz, M. & Werner, H. J. (2001) *J. Chem. Phys.* **114**, 661–681.
110. Maslen, P. E., Lee, M. S. & Head-Gordon, M. (2000) *Chem. Phys. Lett.* **319**, 205–212.
111. Rempe, S. B., Asthagiri, D. & Pratt, L. R. (2004) *Chem. Phys.* **6**, 1966–1969.