Jaguar: A High-Performance Quantum Chemistry Software Program with Strengths in Life and Materials Sciences

Art D. Bochevarov,*[a] Edward Harder,[a] Thomas F. Hughes,[a] Jeremy R. Greenwood,[a] Dale A. Braden,[b] Dean M. Philipp,[b] David Rinaldo,[c] Mathew D. Halls,[d] Jing Zhang,[e] and Richard A. Friesner[e]

Jaguar is an *ab initio* quantum chemical program that specializes in fast electronic structure predictions for molecular systems of medium and large size. Jaguar focuses on computational methods with reasonable computational scaling with the size of the system, such as density functional theory (DFT) and local second-order Møller–Plesset perturbation theory. The favorable scaling of the methods and the high efficiency of the program make it possible to conduct routine computations involving several thousand molecular orbitals. This performance is achieved through a utilization of the pseudospectral approximation and several levels of parallelization. The speed advantages are beneficial for applying Jaguar in biomolecular computational modeling. Additionally, owing to its superior wave function guess for transition-metal-containing systems, Jaguar finds applications in inorganic and bioinorganic chemistry. The emphasis on larger systems and transition metal elements paves the way toward developing Jaguar for its use in materials science modeling. The article describes the historical and new features of Jaguar, such as improved parallelization of many modules, innovations in *ab initio* pKa prediction, and new semiempirical corrections for nondynamic correlation errors in DFT. Jaguar applications in drug discovery, materials science, force field parameterization, and other areas of computational research are reviewed. Timing benchmarks and other results obtained from the most recent Jaguar code are provided. The article concludes with a discussion of challenges and directions for future development of the program. © 2013 The Authors. International Journal of Quantum Chemistry Published by Wiley Periodicals, Inc.

DOI: 10.1002/qua.24481

Introduction

The Jaguar *ab initio* electronic structure program has been developed over the past 20 years with the goal of treating large systems with accurate quantum chemical methods. Currently Jaguar is commercial software produced and maintained by Schrödinger Inc. A distinguishing characteristic of Jaguar is that computational efficiency for large systems has been implemented in it via the use of the pseudospectral (PS) method, a numerical approach to the calculation of the Coulomb and exchange terms, which provides particularly significant advantages for the computation of exact exchange terms.[1–9]

The focus of the present code is on density functional theory (DFT)[10] and local second-order perturbation theory (LMP2).[11,12] The scaling with system size and prefactors in these methods are reasonable, which enables practical calculations for systems containing hundreds of atoms. The program is employed in both the biological and materials science communities for calculations on large systems, particularly in pharmaceutical applications and modeling of enzymatic reactions. Another important application of Jaguar is in force field development. As an important example, Jaguar is an intrinsic component of the Schrödinger OPLS2 package,[13] which features extensive coverage of pharmaceutically relevant chemistries via more than 10,000 torsional terms, each of which has been fitted to appropriate quantum mechanical data.

Over the past decade, our major efforts have been devoted to enhancing Jaguar performance and accuracy in DFT calculations, particularly those using hybrid functionals that require exact exchange. Three areas have seen particularly active development during the previous two decades: enhancing accuracy of DFT functionals, accelerating computational performance of the self-consistent field (SCF) algorithm by, for example, creating better initial guesses for the wave function and implementing efficient parallelization over a large number of processors, and developing methods that yield accurate

[a] A. D. Bochevarov, E. Harder, T. F. Hughes, Jeremy R. Greenwood
 Schrödinger Inc, 120 West 45th St, 17th floor, New York, New York, 10036
 E-mail: art.bochevarov@schrodinger.com

[b] D. A. Braden and D. M. Philipp
 Schrödinger Inc, 101 SW Main Street, Suite 1300, Portland, Oregon, 97204

[c] D. Rinaldo
 Schrödinger GmbH, Dynamostrasse 13, D-68165, Mannheim, Germany

[d] M. D. Halls
 Schrödinger Inc, 8910 University Center Lane, Suite 270, San Diego, California, 92122

[e] J. Zhang and R. A. Friesner
 Department of Chemistry, Columbia University, 3000 Broadway, New York, New York, 10027

Art D. Bochevarov graduated with an MS equivalent in chemistry from Kharkiv National University, Ukraine in 2001, after working for several years with Anatoliy Luzanov. Next year, Art joined the research group of David Sherrill at Georgia Institute of Technology, which he finished with a PhD in 2006. The same year he started his postdoctoral studies in the group of Richard Friesner at Columbia University, working on *ab initio* methods development and on applications of QM/MM methods to iron-containing enzymes. In 2010, he joined Schrödinger Inc, where since 2011 he has been leading the development of the program Jaguar. Art's research interests are primarily focused on improving quantum chemical solutions to practical problems in industrial applications.

Edward D. Harder was born in St. Catharines, Ontario, Canada in 1976. He received his BS degree in chemistry and physics from McMaster University in 1999. In 2004, he was awarded a PhD in chemical physics from Columbia University working in the laboratory of Bruce Berne. He then worked with Benoit Roux as a postdoctoral scholar at the University of Chicago before joining Schrödinger Inc in 2009. His present work is focused on the development of force fields for biomolecular modeling.

Thomas F. Hughes was born in Long Island, NY, in 1980. In 1984, he and his family relocated to a town close to Tampa, FL. In 2002, he graduated Summa Cum Laude from the University of North Florida with a BS in chemistry with emphasis in physics and math. During the latter years of college he worked as a modeler in a pharmacology lab at the Mayo Clinic of Jacksonville. In 2008, he received his PhD in chemistry from the Quantum Theory Project at the University of Florida under the direction of Rodney J. Bartlett. There he worked on developing quantum chemistry methods for computing the properties of large systems using the ACES program. In 2012, he completed his postdoctoral work at Columbia University under the direction of Richard A. Friesner. There he used the Schrödinger suite of software to study the properties of systems containing transition metals. He is currently a Scientist/Developer in materials science for Schrödinger Inc.

Jeremy R. Greenwood was born in Sydney, Australia, in the early 1970s. After obtaining his BSc(Hons 1st) in industrial chemistry from UNSW, studying quantum chemistry in the Radom Group at ANU Canberra, and drug design at the Victorian College of Pharmacy, he took a PhD in medicinal chemistry at The University of Sydney. Following a post-doc in the early 2000s at the Pharmacy School, University of Copenhagen, Denmark, specializing in applications of quantum chemistry to structure-based drug design, he relocated to Schrödinger Inc, where he is now Senior Fellow in the Drug Discovery Group. After more than two decades of applying quantum chemistry to multifactorial problems in molecular design, he has a continued research interest in the steady growth at the interface of experimental and computational chemistry.

Dale A. Braden is a senior scientist at Schrödinger Inc. He was born in Eugene, Oregon in 1964. He holds bachelor's degrees in both classical archaeology (Cornell University, 1987) and chemistry (Portland State University, 1993), an MS in chemistry from Portland State University (1995) and a PhD in chemistry from the University of Oregon (1999), where he worked in the research group of David Tyler. He joined Schrödinger Inc in 2000.

Dean M. Philipp was born in Milwaukee, WI, in 1969. After obtaining his BA in chemistry and physics at the University of Wisconsin-Madison in 1992, he started his graduate career working experimentally with semiconductor nanocrystallites at MIT before leaving with an MS in 1995 to switch to theoretical quantum chemistry at Columbia University. He was awarded a PhD in 1998, and then performed post-doctoral research at Caltech in QM modeling of catalysts until 2002 before accepting a position at Schrödinger Inc, to develop QM and QM/MM methodology for the company's modeling software.

David Rinaldo has been an Applications Scientist at Schrödinger since 2007. He was born in Clermont-Ferrand, France in 1975. He obtained an engineering degree in chemistry from the Ecole Nationale Superieure de Chimie de Montpellier in 1999. In 2003, he was awarded a PhD in structural and molecular physical chemistry from the University of Grenoble, working with Martin Field. Subsequently, he did postdoctoral studies at Columbia University with Prof. Friesner until he joined Schrödinger. His research focused mainly on studying inorganic and bioinorganic systems using theoretical methods such as DFT, QM/MM, and MD simulations.

Mathew D. Halls is the Director of Materials Science at Schrödinger Inc. He was awarded a PhD in quantum chemistry from Wayne State University in 2001, under the direction of Berny Schlegel. Prior to joining Schrödinger in 2012, his activities have focused on advancing and promoting the adoption of atomic-scale chemical simulation techniques in diverse industries including aerospace, electronics and specialty chemicals. His personal research contributions have made significant impact in areas such as computational spectroscopy, organic optoelectronic materials, nanocarbon-polymer interfaces, thin-film precursors and deposition processes, and battery electrolyte additives.

Jing Zhang was born in Wuhan, China in 1985. He obtained a BS in Chemistry from Peking University in 2008. Since joining Columbia University for a PhD program in chemistry, Jing has been awarded his MS degree and MPh degree in Chemistry in 2010 and 2012. He is currently working on developing a high-performance parallel implementation for an electronic structure program, and theoretical modeling of electron transport and trapping in dye-sensitized solar cells. Jing also enjoys music, hiking, and spending time with his friends.
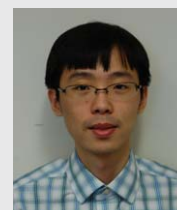
Richard A. Friesner was born in New York, NY in 1952. He received his BS degree in chemistry from the University of Chicago in 1973. He obtained his PhD in 1979 at the University of California, Berkeley, working in the laboratory of Kenneth Sauer. He then spent three years as a postdoctoral fellow working with Robert Silbey at the Massachusetts Institute of Technology. He joined the Chemistry Department at the University of Texas at Austin in 1982 as an Assistant Professor. In 1990, he became Professor of Chemistry at Columbia University. His work is currently focused on computational modeling of complex systems biology and materials sciences. Specific interests include protein structure prediction, structure based drug design, modeling of enzyme reactions, and modeling of nanosystems such as silicon nanoparticles and carbon nanotubes with a particular focus on solar energy applications.

results in the condensed phase. Each of these efforts is outlined briefly below and in more detail in the main body of the review. For condensed phase applications in solution, the use of continuum solvation models is critical to obtaining results that can be profitably compared with experimental data. Jaguar was one of the first quantum chemistry programs to develop an optimized continuum solvent model, based on a self-consistent reaction field solution of the Poisson–Boltzmann (PB) equation,[14–16] and we have continued to improve and test this approach for predicting a wide range of properties including solvation energies,[17,18] pKa's,[19,20] redox potentials,[21,22] and other aspects of chemical processes.[23]

Development of DFT functionals is being conducted by many research groups, and at present, there is a very large number of alternatives in the literature.[24–26] Many of the popular functionals have been implemented in Jaguar. We

highlight in particular the Minnesota family of density functionals, such as M06-2X,[27,28] which have displayed substantial improvements for a wide variety of properties including heats of formation, dispersion interactions, and conformational energies. In parallel, we have developed our own novel approach to improving the results of DFT calculations, which is based on the use of empirical localized orbital correction (LOC). The theoretical foundation of the DFT-LOC approach is described in Ref. [29] and is also briefly summarized in the present review in section Methods employing LOC. The basic idea is to associate errors in DFT energetics, particularly for the B3LYP functional, with specific electron pairs (chemical bonds and lone pairs) and singly occupied orbitals, and to assume the transferability of these errors for a specific chemical environment from one molecule to another. This approach has enabled chemical accuracy (~1 kcal/mole average errors) to be obtained for

molecules composed of second and third row atoms,[30,31] and somewhat larger average errors of 2-3 kcal/mole for redox potentials, spin splittings, and metal-ligand bond energies in systems containing transition metal atoms.[32–34] These results compare favorably with alternative efforts to improve DFT accuracy, and therefore we decided to include several LOC-based methods in Jaguar.

Finally, achieving high performance for large systems continues to be a crucial objective in the application of *ab initio* quantum chemical methods in materials science and biology. An essential technology for biological systems involves mixed quantum mechanics/molecular mechanics (QM/MM) calculations, which enable the treatment of proteins and other biological macromolecules to be partitioned into a reactive core of 50–300 atoms surrounded by a periphery described by MM.[35,36] For this purpose, Jaguar has been an essential component of the QM/MM package QSite.[37,38] A second effort in this area is improvement of the parallel performance in Jaguar. This has been a challenging project, but we will describe recent results in which we have been able to scale performance up to as many as 120 computer processor cores, while retaining outstanding single node performance. Parallelization allows time to solution for an arbitrarily large system (e.g., a metal oxide cluster, such as our model for TiO_2 nanoparticles in the Grätzel cell, discussed in section Optoelectronics and photovoltaics), to be substantially reduced; without such improvements in speed, many projects would not be feasible at all due to the high computational requirements to complete individual calculations. For performance of the Jaguar parallel code, see section Jaguar parallelization.

Distinguishing Characteristics of Jaguar

Jaguar shares multiple features with other electronic structure software programs. For example, some of the basis sets, density functionals, and molecular property predictions available in Jaguar are also available in several other packages. At the same time, a number of features make Jaguar stand out, namely: (i) prediction of vibrational circular dichroism (VCD) spectra, optionally combined with conformational search and statistical averaging of spectra; (ii) pKa prediction module; (iii) D3 *a posteriori* correction[26,39] with first and second derivatives with respect to nuclear displacement; (iv) novel *a posteriori*-corrected DFT methods B3LYP-MM (Ref. [40]) and PBE-ulg (Ref. [41]) that account for noncovalent interactions (NCI); (v) computation and visualization of average local ionization energy (ALIE) surfaces; (vi) visualization of NCI; (vii) Fukui functions and Fukui atomic indices;[42,43] (viii) automated prediction of hydrogen bond strength; (ix) LOCs to the B3LYP functionals treating nondynamic correlation errors of organic and *3d*-metal-containing systems. Finally, Jaguar enjoys a graphical user interface (GUI) shared with multiple other computational software programs developed by Schrödinger, such as Glide and MacroModel. This allows one, for example, to treat one molecular system on several different computational levels without having to do format translation and without having to learn or adjust to a different graphical interface.

But a more distinguishing characteristic defining the program's unique strategic placement *vis-à-vis* its competitors is the philosophy underlying directions of its development. Jaguar was conceived with a vision to make *ab initio* calculations on larger molecules routine without breaking outside of the accepted accuracy borders. Focus was put on delivering robust solutions and not on implementing as many recently published methods as possible. Typically, a decision to put a new method in Jaguar was taken only if the resulting feature would help address a specific computational problem or significantly improve on the accuracy of the previously existing solution. In this regard, it is even possible in some cases to hide or remove from Jaguar methods that are no longer useful for solving practical problems because they have been unquestionably superceded by more accurate or powerful methods.

Information overload has become detrimental in many scientific domains such as journal publications and genomics data, and is encroaching on computational chemistry. Computational chemists who cannot afford to spend a significant portion of their time to follow the latest developments of methods are often baffled when having to make a choice among dozens of available density functionals, and especially hundreds of possible combinations of density functionals and basis sets. Jaguar's policy of "not multiplying features beyond necessity", reminiscent of the famous Occam's razor principle, should help better manage the wealth of the available computational methods and protocols.

Another distinguishing characteristic of Jaguar is the emphasis on performance. We believe that a high-end performance of a few reliable and well-validated methods is a preferable feature for practically oriented computational chemists. Jaguar is unique among *ab initio* software programs in that it uses PS approximation in many computationally intensive modules. According to our testing, this reduces the total time required for the computation by a factor of 2–5, depending on the dimensionality of the molecular system and the computational task. For more information about PS approximation as used in Jaguar, see section Pseudospectral approximation. The three layers of parallelization [OpenMP,[44] Message Passing Interface (MPI),[45] and trivially parallel] which can be used simultaneously for many computational tasks in Jaguar is another leg of the high-end performance focus (see section Jaguar parallelization).

Program Organization

A computational program spanning several decades of active development is expected to have multiple components and interfaces. Figure 1 shows principal constituents of Jaguar that are accessible to the user, and ways in which these constituents interact.

At the bottom level, there is the compiled executable code. It spans a number of individual executable files, each of them being typically responsible for a specific large computational task. For example, the executable file onee (on Windows systems, the executable file names are augmented with the
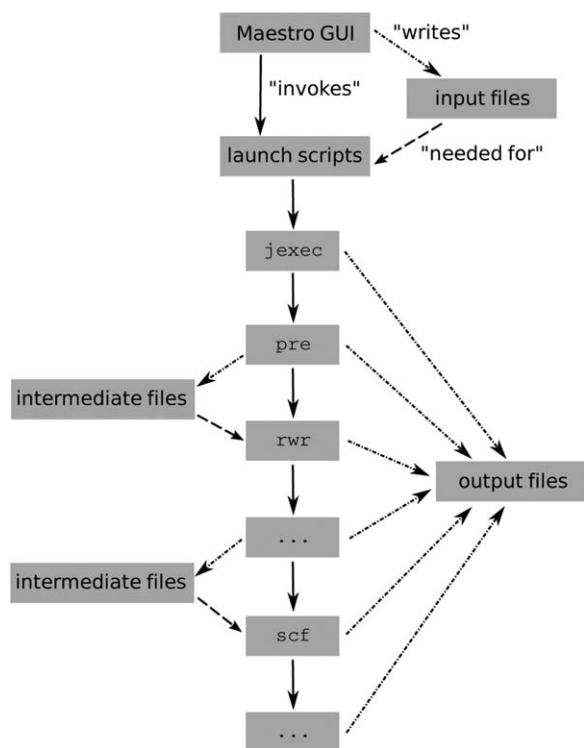
Figure 1. A scheme showing Jaguar program organization. See text for details.

extension ".exe") computes one-electron integrals, and the executable file scf performs operations related to solving SCF equations. Many executable files are accompanied by their parallel versions. For example, the executable file pscf performs parallel SCF calculations. Executable files exchange information by writing and reading text and binary files. The executable file jexec drives all other executable files.

In its turn, the "top" program jexec is invoked by a chain of job control scripts that control how signals and files propagate from one machine to another (which becomes particularly important when the Jaguar command is issued by the user on one machine, often being a desktop computer, but the corresponding computationally expensive operation is executed on another machine which is often a computational cluster). The top-level Jaguar script parses and processes the command issued from the command line or from the GUI.

Another important task executed by the job control scripts is to transport files from the launch to scratch directory and back. At the start of the calculation, the input files are copied to the scratch directory. The actual Jaguar calculation proceeds from the scratch directory, which holds all the files necessary for the calculation, including temporary and intermediate files generated in the course of the calculation. At the end of the calculation, the results are copied back to the launch directory, creating an illusion that the Jaguar calculation was carried out in the launch directory. After this, the scratch directory is erased. The user may request that the files in the scratch directory be saved after the calculation, in case the temporary files are of interest.

All Jaguar computations can be launched from the command line. The specification of the molecular system and all settings pertaining to the method, convergence criteria, and so forth is performed in an "input" file (which has the extension ".in"). Multiple options pertaining to the computational environment (such as on which computer the calculation will be performed and whether it will be executed in the serial or parallel mode) are specified directly in the command line.

Finally, if the user prefers to operate Jaguar from the graphical interface Maestro, then the latter essentially constructs and launches commands on behalf of the user. See the following section for more details on Maestro.

GUI Maestro

Even though Jaguar is fully functional in the text-based mode, so that any types of calculations can be setup and submitted through text editors and from the command line, it is often convenient to work with molecular structures, spectra, and three-dimensional (3-D) surfaces graphically. Maestro is the only officially supported graphical interface for Jaguar. It uses familiar menus, windows, buttons, panels, and high-quality shading graphics for rendering 3-D objects to create a powerful, customizable, self-contained working environment expected of a modern computer application. Operations on molecular structures and their fragments in Maestro receive full graphical support: structures can be built, visualized, rotated, superimposed, labeled, transformed, analyzed, and so forth. Snapshots from a typical Maestro work session are shown in Figure 2.

In Maestro, Jaguar calculations are launched from a Jaguar panel which allows the user to choose the level of theory and the basis set, adjust convergence criteria, select molecular properties to be computed, and so forth. Several different Jaguar panels exist for different types of calculations: single-point energy calculations, geometry optimizations, coordinate scans, transition state searches, and so on. Additionally, Maestro offers panels to set up the details of Jaguar workflows such as pKa prediction and counterpoise calculations. A special Jaguar profile or "skin" is available for the researchers who are using Maestro solely for small molecule calculations. This profile hides most features of Maestro relevant to proteins.

The proceeding Jaguar calculation can be monitored from Maestro, and once the calculation finishes, its results are "incorporated" into the Maestro project table which keeps the details of the calculation organized. Multiple calculations and structures can be processed simultaneously on different levels: grouped into "projects," launched as a batch, sorted, searched, compared, and so forth.

An important feature that distinguishes Maestro from GUIs associated with other quantum chemistry programs is that Maestro is a native interface not only for Jaguar but many other Schrödinger products that may be of interest to a computational chemist. Because of this "umbrella" coverage, results obtained from Jaguar are immediately available to programs such as the semiempirical neglect of diatomic
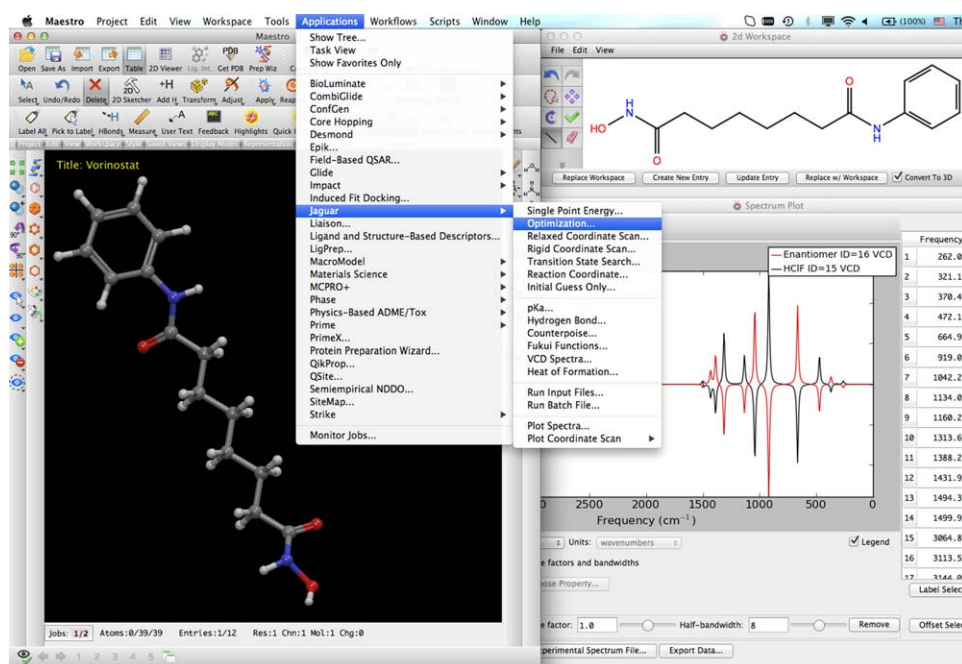
Figure 2. A snapshot of a typical Maestro 9.4 working session.

differential overlap (NDDO) module, MM MacroModel program, protein-ligand docking software Glide, tautomerization and pKa prediction tool Epik, and vice versa. Therefore, the user working with Jaguar and one or more such additional programs, does not need to convert files from one format to another or to adjust to a different working environment.

Jaguar Features

Jaguar strives to provide a complete package of features needed to solve realistic computational problems in which quantum effects play a role. We believe that in solving practical problems, especially those involving high-throughput screening, it is preferable to rely on tools that are known to deliver superior accuracy and performance. When priorities are robustness and accuracy within the expected error bars, then novel approaches should be used discreetly, to minimize unforeseen effects. At the same time, the application of new and relatively untested methods can be warranted if there are no alternative ways to solve the problem. Jaguar is committed to these two approaches: add features sparingly where good alternatives exist (and are already implemented in Jaguar), and provide novel and original methods where alternatives do not exist or are computationally prohibitively expensive.

Traditional methods

Jaguar was originally designed as an *ab initio* quantum chemical program that would be able to routinely work with larger molecules (100 atoms and more). For this reason, the developers ruled out an implementation of post-SCF methods such as Møller–Plesset perturbation theory, configuration interaction, and coupled cluster theories,[46,47] which all suffer from computational cost scaling as the fifth or higher power of the

molecular size N. Instead, our efforts were concentrated on methods which scale as N^3, at most. Only such approaches would allow one to routinely process molecular systems involving at least a thousand basis functions, which is what one expects to use in a hundred-atom calculation.

The simplest, noncorrelated *ab initio* method, which is the Hartree–Fock theory,[46] and which is considered to be the zeroth approximation in many *ab initio* programs, scales as N^4. Additional approximations can bring the cost down by two orders of magnitude. The accuracy of Hartree–Fock methods is inadequate for most applications, but these approximations can also be applied to the much more accurate DFT which has the same computational scaling as the Hartree–Fock theory.

The PS approximation, which is described in some detail in section Pseudospectral approximation, is one such approximation. It reduces the scaling of the Hartree–Fock and DFT methods to N^3, and, after some additional cutoffs in the integrals, even to N^2. Another approximation that brings significant reductions in scaling is the localized orbitals approximation.[48,49] When combined with the Møller–Plesset second-order perturbation theory (MP2) and the PS approximation, the scaling of the resultant method, known as local MP2, or LMP2, goes down to N^4. Additional integral cutoffs lead to scaling of N^3 on many realistic systems.[9] These approximations contribute significantly to the high performance of Jaguar.

DFT methods remain very popular with Jaguar users for their speed, accuracy, and simplicity of use. Numerous reviews and benchmark studies are dedicated to DFT methods.[24,25] Despite the widely held claim that (individual) DFT functionals, in contrast to multideterminant *ab initio* theories, cannot be improved in a systematic way, conceptual DFT development and improvement do not seem to slow down.[50,51] Jaguar is dedicated to providing its users with new essential DFT developments.

Several dozen functionals are available in Jaguar. Apart from the early and traditional functionals such as SVWN5, B3LYP, and PBE, Jaguar contains the popular M05 and M06 families of functionals (often called the Minnesota functionals) developed by Zhao and Truhlar.[27,28] Recently, a new wave of development has focused on a posteriori corrections which aim to address various deficiencies of DFT methods with a simple treatment which contributes essentially zero computational cost to the total time. Most often such corrections are aiming at improving the description of NCI. Of a posteriori-corrected functionals, Jaguar contains B97-D (see Ref. [52]) and a family of D3-corrected functionals[26,39] developed by Grimme et al. Several new a posteriori corrections will appear in the 2013 release of Jaguar (version 8.0). They include the LOC of Friesner et al.[29–31] (see section Methods employing LOC for more details), B3LYP-MM,[40] and PBE-ulg methods.[41] The a posteriori part of B3LYP-MM was trained on over 2000 single point energies coming from accurate methods, often of coupled cluster singles, doubles, and perturbative triples [CCSD(T)]/complete basis set (CBS) quality calculations, and includes specific corrections for dispersion, hydrogen bond, and cation-π interactions. The correction in the PBE-ulg method is based on the universal force field (UFF),[53] and supports all elements up to element 103. First and second derivatives with respect to nuclear displacements are available for the vast majority of Jaguar's DFT functionals (including the Minnesota families and D3-corrected functionals).

LMP2 calculations in Jaguar are popular with researchers who prefer working with wave function-based approaches.[54] LMP2 has also been a method of choice in the construction of force fields (see section Conclusions). However, we would like to demonstrate that nowadays some modern DFT functionals can compete with LMP2 in accuracy.

Several recent benchmarking studies show that functionals such as M06-2X-D3 yield higher-quality predictions than MP2 across a broad range of properties.[55–59] The LMP2 method, just as another popular approximation to MP2, resolution of identity MP2 (RI-MP2),[60,61] is expected to show similar performance to MP2. When LMP2 was developed, accurate functionals such as M06-2X did not exist, and a study comparing the performance of all these methods would be important. We carried out such a benchmarking study comparing LMP2 versus DFT functionals for how closely they can match different CCSD(T)/CBS conformer and rotamer energetics. The results summarized in Table 1 show that, generally, modern DFT methods such as M06-2X are as accurate or more accurate than LMP2. For conformational and rotational energetics of small molecules, the performance of LMP2 versus M06-2X is similar or slightly superior. However, for more extended peptides, the conformational energetics predicted by M06-2X and other modern DFT functionals are clearly better.

Methods employing LOC

Despite a significant progress in improving the accuracy of DFT methods, the nondynamic correlation error remains a significant source of error in DFT.[62–65] A series of novel methods that are based on the so-called LOCs and designed to address the problem of nondynamic correlation in a computationally tractable way will appear in the 2013 release of Jaguar. However, there has also been success in addressing the problem using more traditional design of density functionals.[65,66]

The premise behind the DFT-LOC approach is that the dominant errors in the current DFT functionals are determined by the chemical environment of localized electron pairs, and are transferable from one molecule to another if that environment is sufficiently similar. An analysis of high-level quantum chemical correlation methods indicates that by far the largest electron correlation terms are the intrapair terms between two electrons in a chemical bond or lone pair, followed by interactions between electrons in bonded atoms.[29] At greater distances, the correlation energy falls off rapidly, and one can assume that it is sufficiently well-described by DFT, because the most important types of nondynamical correlation (in-out or left-right correlation) are not relevant to such interactions.

The theoretical basis of DFT-LOC has been discussed in detail in Ref. [29]; here, we provide an outline of the arguments made therein. The central hypothesis is that modern DFT functionals represent nondynamical correlation via the self-interaction "error" term, particularly the component of this term in the DFT exchange functional. Evidence in favor of this hypothesis is found in a paper by Vydrov and Scuseria,[67] who approximately removed the self-interaction term from B3LYP and other functionals and examined the effect on average error in enthalpies of formation for the G2 data set of Pople and coworkers. The mean unsigned error (MUE) increased by 16.5 kcal/mole for B3LYP, with most molecules in the data set substantially underbound.

Table 1. Comparison of performance of DFT to LMP2 on a diverse test set of conformers and rotamers.

Type	# in dataset	Basis	B3LYP	B97-D3	M06-2X	PW6B95-D3	LMP2	LMP2 rel. time
Small molecule conformers	68	6-31G*	0.64	0.50	0.51	0.50	0.60	
		cc-pVTZ(-f)	0.53	0.32	0.28	0.27	0.23	
Peptide conformers	60	6-31G*	2.40	0.89	0.77	0.66	1.74	
		cc-pVTZ(-f)	2.34	0.74	0.54	0.46	1.86	
Small molecule rotamers	550	6-31G*	0.41	0.45	0.31	0.45	0.52	3.63
		cc-pVTZ(-f)	0.27	0.32	0.19	0.32	0.18	3.14

The mean unsigned energy errors, in kcal/mole, for different levels of theory are generated against energies of the CCSD(T)/complete basis set quality. LMP2 relative timings are for wall-clock time, with respect to B3LYP, over the whole test set of 678 systems.

In the B3LYP functional, the average self-interaction error is optimized by adding a component of exact exchange which minimizes the MUE of the G2 atomization energies. However, this optimization only minimizes the MUE on average. Large outliers exist, for B3LYP in the G2 and G3 data sets. The errors can be as large as 4 kcal/mole per bond (NaCl, O_3, SF_6) and range from $+8$ kcal/mole to -22 kcal/mole over the G3 data set.[64] Furthermore, there is clearly a systematic component to the error as can be seen by examining, for example, errors in hydrocarbon chains as a function of molecular weight. The M06-2X functional yields a reduced number and the magnitude of the outliers,[28,30] but significant problems are still present, for example, with multiply halogenated compounds.[30]

The DFT-LOC method amounts to applying several different types of semiempirical corrections *a posterioiri* to the self-consistent DFT energy. In the current parameterization, there are 22 basic correction types (and, correspondingly, 22 parameters) that cover most types of nondynamic correlation error in neutral systems consisting of elements H-Ar. Sixteen more parameters account for a removal or an attachment of an electron in such systems, and eight additional parameters treat transition states. Additional parameters are applied for various aspects of chemistry of *3d*-elements and are described in detail in the next section.

The basic correction types are assigned to different types of atoms and bonds. The atomic corrections depend on the element and its approximate hybridization state, and the bond corrections are assigned according to the bond order, polarity, and (in the case of atoms C—F) length. In our parameterization, not all elements and hybridization states have parameters associated with them. This, and using more general bond types than what a more elaborate scheme would generate, protects the model from an excessive number of parameters to fit. Although all the DFT functionals can benefit from LOC parametrization, to date B3LYP consistently yields the smallest MUEs, and hence has been the main focus of our optimization effort.

Even though the basic parameterization model contains only 22 parameters (which is fewer than that in some modern functionals, for example, M06-2X), it is still critical to test for the possibility of overfitting. This can be done in a number of ways. For instance, one can fit the parameters of the model to the G2 set, and then perform a test on the 75 molecules in the G3 set (which are larger and more complex than those in the G2 set, making fortuitous agreement particularly unlikely). The MUE for the G3 set in this case is 1.0 kcal/mole, as compared to the 0.8 kcal/mole obtained when the parameters are fitted directly to the full G3 data.[29] The negligible change in MUE (and in the great majority of parameter values) demonstrates that the parameters are robust. A second test involves the use of test data not used in the original parameter fitting. Such data are presented in (Bochevarov and Friesner, unpublished), and again, the errors are dramatically reduced compared to B3LYP and other functionals (including M06-2X, with which direct comparisons are made).

Subsequent to the original DFT-LOC publication, we have greatly expanded the range of functionality and chemistry covered by LOC parametrization. Included are ionization potentials and electron affinities, transition state energetics, and a model that covers transition metals (redox potentials, spin splittings, and atomization energies). The errors remain consistent with those discussed above; MUEs of ~1 kcal/mole for second and third row atoms, and 2–3 kcal/mole for transition metals.[32–34] The increased error in case of transition metals is due to a combination of the greater complexity of the electronic structure of transition metal atoms and chemical bonding, and the paucity and lower quality of experimental data for transition-metal-containing systems. A more extensive review of the transition metal results are given in the next section, as accurate calculations on such systems has proven to be very difficult for conventional DFT approaches.

Automated implementation of LOC in Jaguar requires essentially valence bond assignments of hybridization states, charge distribution, bond order and type, and so forth. Such assignments are typically straightforward, but some cases are challenging because they involve alternative (and sometimes very different) resonance structures. An additional problem with this simple parameter scheme is the discontinuous nature of the corrections which do not depend directly on atomic coordinates. For this reason, it is impossible to develop explicit gradients of the corrections with respect to nuclear displacements. The possibility of interpolation between two end-point structures, in this way introducing an explicit dependence on the atomic coordinates, has been demonstrated for LOC.[68] Nevertheless, the LOC model is yet to be cast in a form that would permit differentiation in a more traditional setting, when a single initial structure is available. Thus, in the current implementation of B3LYP-LOC geometry optimizations, LOC gradients are assumed to be zero. This is a reasonable approximation because most geometry optimizations starting from close-to-equilibrium geometries do not go through changes in the valence-bond structure in the course of the optimization.

LOC for 3d transition metals (DBLOC)

Transition metals present severe challenges for first principles electronic structure methods, due in great part to the fact that the valence orbitals include a *d*-shell (*3d* for the first-row transition metals). The *d*-electron manifold gives rise to complex chemical bonding and many low-lying spin states, which can lead to significant multireference character. Converged coupled cluster results at the CCSD(T) level exist for the bare atoms,[69,70] and yield good agreement with experimental spectroscopic data. However, even for small molecules containing only a few atoms, comparisons between theory and experiment at the CCSD(T) level can be problematic. For the larger systems that are relevant to biology and materials science, only DFT methods are computationally tractable. The accuracy of DFT methods for transition–metal-containing species is difficult to estimate reliably. Siegbahn[71] has proposed that B3LYP methods typically achieve results with a 3–5 kcal/mole error, but this estimate is based on a relatively small set of calculations. Errors in the range of 10–15 kcal/mole are in

fact readily demonstrated, as is noted below in more detail. Modern functionals such as M06 display modest improvements over B3LYP for small (two to four atom) systems,[28] but these species are coordinatively unsaturated, and hence the conclusions drawn from their study may not be transferable to more normal coordination complexes or solid state systems.

The low accuracy of the existing methods, the scarcity of the experimental data to compare and validate against, and often large experimental error bars all make the problem of treating transition metal systems with DFT methods a formidable one. What is even more worrying is that there is little hope for greatly improving the existing quality of predictions for transition metal systems in a few direct steps; it is more likely that the progress will be incremental and slow. Therefore, we developed semiempirical corrections in the spirit of the LOC approach presented in the previous section, aimed at treating various aspects of chemistry of 3d transition metal systems. These corrections are referred to as DBLOC (after d-block LOC) and are combined with the B3LYP functional, for the reasons similar to those that make us prefer the B3LYP-LOC combination. Although semiempirical in nature, we believe that the B3LYP-DBLOC approach provides the first methodology capable of handling large, transition-metal-containing systems with consistent benchmark errors for relevant data sets in the range of 1–3 kcal/mole MUE. The transferability of this parameterization to materials and biological problems remains to be investigated. Two initial results are available. We used the redox corrections in studying the electronic state of TiO_2 nanoparticles,[72] and found that simply using the corrections for the oxygen anion ligands surrounding a Ti atom produces a correction term of 0.8 eV, which brought theory into agreement with experiment for a number of key comparisons, including the open circuit voltage of the system. Second, by computing the redox correction for the iron porphyrin cofactor in cytochrome P450,[32] we have been able to explain the long standing problem of DFT methods computing a barrier height for the key hydrogen abstraction step in the P450 catalytic cycle that is much too large compared to experiment. With the DBLOC correction terms, the resulting barrier is consistent with the experimental observation of ultrafast reactivity of the Compound I catalytic species. The rest of this section describes the implementation of the B3LYP-DBLOC approach in Jaguar.

The method determines the corrections by applying a set of operators to bring the d-electrons of a standard reference electronic configuration to that of the desired target electronic configuration which is determined using ligand field theory. The reference state ligand field diagram is defined to be the electronic state of lowest spin multiplicity for the metal cation in its most common oxidation state. Each operator is associated with a B3LYP-DBLOC correction parameter which has been fit to large and diverse databases of experimental spin-splitting,[33] redox,[32] and ligand removal[34] energetics.

One problem that arises when using quantum chemical methods in applications involving transition metals is how to ensure convergence to the desired electronic state. In contrast to relatively simple organic molecules, which even on a rather strong perturbation of the initial guess wave function maintain their convergence to the same ground state, systems containing occupied d-orbitals are more sensitive to the initial guess because the energy gaps between successive states can be smaller. This means that it is possible, with only a small change in the initial guess, for a calculation to converge to a low-lying excited state rather than the desired ground state. For this reason, Jaguar has a series of initial guess options that are tailored for molecules containing transition metals. One of the initial guess options, namely the ligand field theory guess, includes a description of d-d repulsion. It involves an automated process of selecting unoccupied d-orbitals deemed beneficial for mixing with occupied d-orbitals, similar in spirit to the well-known strategy of mixing in virtual orbitals to help convergence to the ground state for otherwise difficult to converge molecules. For transition metal complexes with complicated electronic structures, an ab initio exploration of the electronic energy levels may be required. Such a procedure typically involves using a series of automated calculations, which search over different formal charge assignments for the metal and ligands, different spin states for the metal, and different electronic configurations for the metal.

B3LYP-DBLOC was primarily trained against spin-splitting, redox, and ligand removal energetics of mononuclear, largely coordinately saturated octahedral, transition metal complexes containing 3d-metals with common oxidation states and singlet to sextet multiplicities, and a variety of closed shell ligands.[32–34] Many important applications involving transition metals in materials science and biology fall into this category where B3LYP-DBLOC is expected to provide accurate energetics. The DBLOC model can in principle be used for other kinds of systems containing transition metals. However, due to a lack of sufficient experimental data, its accuracy has been validated on a much fewer number of systems than that used in the original training procedure. Therefore, the method should be used with caution for such systems, particularly: (i) molecules containing more than one transition metal atom, for example, a transition metal cluster; (ii) transition metal complexes with ground states containing 4s- or higher orbital occupation, for example, as in calculations on the complex $[Ni(II)(phen)_3]^{+2}$, which is a ground-state triplet, using a quintet multiplicity that necessarily occupies the 4s- or higher-orbital; (iii) transition metal complexes with an incomplete noble gas electronic configuration, for example, as in calculations on the complex $[Cr(III)(phen)_3]^{+3}$, which is a ground-state quartet, using a sextet multiplicity that necessarily leaves incomplete the noble gas electronic configuration; (iv) transition metal complexes which are neither octahedral nor tetrahedral; (v) complexes featuring uncommon coordinating ligands, for example, open-shell ligands. The final B3LYP-DBLOC method[32–34] was parameterized against a total of 184 spin-splitting, redox, and ligand removal energies calculated using 368 different transition metal complexes, and currently uses five spin-splitting parameters, seven redox parameters, and six ligand removal parameters. The B3LYP-DBLOC method provides considerable improvement over conventional B3LYP, see Figures 3–5, for example, bringing the MUE over the spin-

splitting database from 10.14 ± 4.56 kcal/mole to 1.98 ± 1.62 kcal/mole, over the redox database from 0.40 ± 0.20 V to 0.12 ± 0.09 V, and over the ligand removal database from 3.74 ± 3.51 kcal/mole to 0.94 ± 0.68 kcal/mole.[32–34]

The B3LYP-DBLOC method is implemented in Jaguar as part of the B3LYP-LOC model, the former being activated if the target calculation contains a transition metal atom. For certain unforeseen cases, the user may need to specify in the input file the oxidation state and multiplicity of the transition metal center. The B3LYP-DBLOC program will then determine certain properties of the metal center and its coordinating ligands
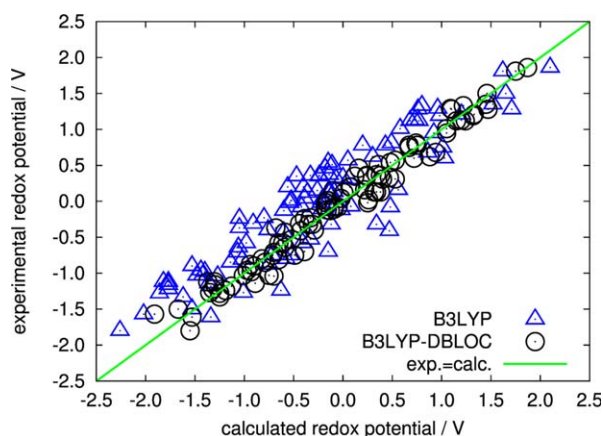

Figure 5. Experimental versus B3LYP/LACV3P and B3LYP-DBLOC/LACV3P calculated ligand removal (30 cases) energetics, in kcal/mole.[34]

metal center and its interactions with the ligands. Although a restricted open shell reference wave function is often the desired reference function for simple open shell molecules like organic doublet radicals, an unrestricted open shell reference function can be more advantageous for molecules containing transition metals for two reasons: (i) it helps alleviate convergence issues; (ii) it relaxes the constraint that the wave function be a spin eigenfunction, which is beneficial for certain applications where the concept of mixed spin states becomes physically important.[33] For this reason, the B3LYP-DBLOC method for transition metals is meant to be used with an unrestricted reference wave function.

Figure 3. Experimental versus B3LYP/LACV3P and B3LYP-DBLOC/LACV3P calculated redox potentials (59 cases), in V.[32]
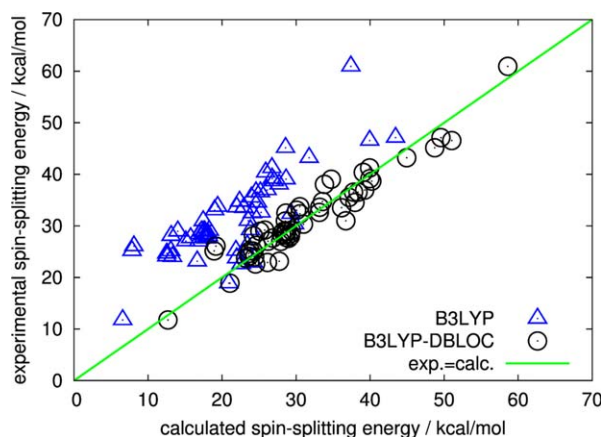

Figure 4. Experimental versus B3LYP/LACV3P and B3LYP-DBLOC/LACV3P calculated spin-splitting (59 cases) energetics in kcal/mole.[33]

and output the ligand field diagrams of the target complex and its reference state. Last, it will determine the necessary corrections for generating the target state from the reference state and print out a table detailing all of the applied DBLOC corrections and their weights. The final B3LYP-DBLOC energy is a sum of the total B3LYP energy, the LOC correction, which is for the ligands, and the DBLOC correction, which is for the
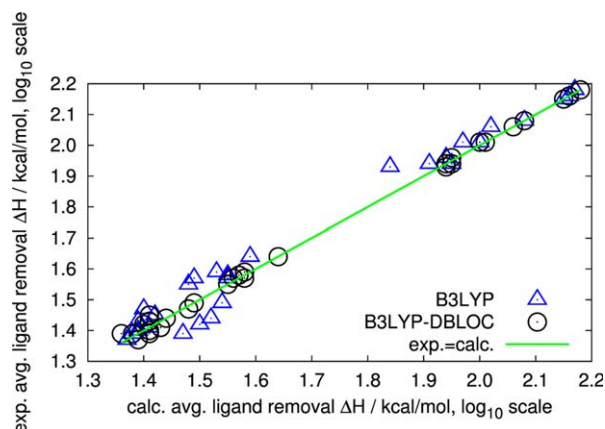
Basis sets

Like most electronic structure software programs, Jaguar uses standard basis sets consisting of Gaussian-type atomic orbitals. The Pople-style basis sets of segmented contractions commonly known as STO-3G, 3-21G, 6-31G, 6-311G, and 6-311G(3df,3pd) (Refs. [73–91]), as well as the Dunning-style, generally contracted basis sets cc-pVDZ, cc-pVTZ, and cc-pVQZ(-g) (Refs. [92–95]) are available. The less common MIDI! basis set of Truhlar and coworkers[96–98] is also included with Jaguar. For transition metal systems, the Los Alamos effective core potentials (ECP) of Hay and Wadt[99] are used in conjunction with double-zeta or triple-zeta basis sets, known as LACVP and LACV3P,† respectively. The all-electron split-valence basis set of Rappoport and Furche[100] may be used for systems containing the elements hydrogen through krypton. For lanthanides, Jaguar offers the CSDZ ECP basis set of Cundari and Stevens.[101] Lanthanides, actinides, and transition metals are covered by the ERMLER2 ECP basis set of Ermler et al.[13,102,104–108] All these basis sets, except for STO-3G, 6-311G(3df,3pd) and ERMLER2, may be used in PS calculations (see section Pseudospectral approximation). Users may also input their own basis sets. In addition to that, Jaguar provides a utility script which converts the basis sets

---

†The double-zeta LACVP basis set was decontracted to create the triple-zeta LACV3P basis set.

downloaded from the EMSL Basis Set Exchange website[109] into Jaguar's native format.

The creation of PS grids and dealiasing functions is not a straightforward procedure because it involves constructing a training set and then fitting the PS parameters to the energies in an iterative fashion, until a balance between the accuracy and the broad coverage of chemical space is achieved. However, a description of this procedure is available, and the users can follow it to produce their own PS basis sets.

The default basis set in Jaguar is 6-31G**, which covers the elements hydrogen through argon. This is a popular basis set used in most electronic structure programs, especially for geometry optimization and for the calculation of vibrational frequencies. For more accurate computations of the energy, such as those usually following geometry optimization calculations, we recommend cc-pVTZ(-f). When working with transition metals, we recommend the LACVP** for geometry optimization and LACV3P** for (subsequent) energy evaluation. When building a structure using the Maestro graphical interface, the basis sets which are defined for the elements in the structure are shown in a menu so that users are always aware of what basis sets they may choose from.

For energies and gradients, angular momentum up to f functions in the basis set is supported, whereas for second derivatives the limit on angular momentum is d. For ECP, the angular momentum limit on potential is g. Because the DFT methods in Jaguar do not require basis functions with angular momentum as high as that used for explicitly correlated methods like MP2 and CCSD(T), these limits are not restrictive and adequately cover the chemistry of most of the periodic table. However, lanthanides and actinides, in which the f orbitals are occupied, would likely benefit from the availability of g functions to polarize the f orbitals.

At this time, Jaguar can perform calculations on molecules with up to 1000 atoms and 12,000 basis functions. The largest system that we have tested Jaguar on was a B3LYP/6-31G** energy and forces calculation on a scorpion protein toxin containing 989 atoms and 10,121 basis functions. This calculation was conducted several months ago on a single core of a presently somewhat obsolete AMD Opteron 2427 processor, and required 229 hours to complete. The peak memory usage was 7.0 Gb, which was for the calculation of the atomic forces. The SCF part of the calculation took 173 hours and memory use peaked at 4.7 Gb.

Pseudospectral approximation

PS numerical methods are used when it is convenient to represent the solution to a partial differential equation (such as the Hartree–Fock or DFT electronic structure equations) as both a basis set expansion and as values on a numerical grid. PS methods were originally developed to model fluid mechanics in periodic systems, and used fast Fourier transforms enabling efficient use of a plane wave basis set to evaluate the Laplacian of the system, and a grid representation to evaluate the nonlinear multiplicative terms in the Navier–

Stokes equation (see, e.g., Ref. [110]). Plane wave-based electronic structure methods have now become routinely used in solid-state DFT calculations (see, e.g., Ref. [111]) and in fact are based on a straightforward implementation of the original PS approach. The application of PS methods to Gaussian-based functions is much more challenging numerically, due to the nonorthogonality and spatial nonuniformity of the basis set; however, it is just these properties of the basis functions which make them efficient in *ab initio* electronic structure approaches.

Rather than a simple matrix transformation between basis functions and grid points, it is necessary to develop technology to deal with basis set overcompleteness and noise due to grid nonuniformity. These numerical instabilities are known in the fluid mechanics literature under the term aliasing errors in a PS implementation. Basis set overcompleteness (especially common in basis sets with diffuse functions) can be addressed by diagonalizing the basis set overlap matrix and removing components corresponding to small eigenvalues. Noise suppression can be achieved by a combination of increasing the ratio of grid points to basis functions, using a least squares algorithm to transform from the grid representation to the basis set representation, and including extra basis functions (dealiasing functions) which are used in the least squares fitting process but then discarded. A significant amount of work is required to develop suitable grids and dealiasing functions for any given quantum chemical basis set; however, as we explained in section Basis set, a protocol for achieving stable and accurate results has been established, and this is a key component of Jaguar technology.

PS methods in Jaguar are not used to evaluate every term in the Hamiltonian; for example, one electron nuclear attraction integrals are readily computed analytically at negligible computational cost. A subset of the two electron integrals, including all one and two center integrals, and selected three center integrals, are also calculated analytically. This restricts the noise induced by the numerical methods to terms which are several orders of magnitude smaller than the biggest contributors to the total energy. The required calculations thus involve evaluation of portions of the Fock matrix involving three- and four-center integrals.

A derivation of the Hartree–Fock and DFT PS expressions can be found in Ref. [54]. Below, we reproduce expressions for the Coulomb and exchange matrix elements (these expressions do not take into account the partitioning of the operators into analytical and PS components; this introduces some technical complications, but does not change the fundamental scaling with system size or operational complexity).

The Coulomb matrix in the molecular orbital space is given by the formula

$$J_{\mu v} = \sum_g Q_\mu(g) \left[\sum_{kl} A_{kl}(g) D_{kl} \right] R_v(g), \qquad (1)$$

and the corresponding exchange matrix is

$$K_{\mu v} = \sum_g Q_\mu(g) \left[\sum_n A_{vn}(g) \sum_m D_{nm} R_m(g) \right]. \qquad (2)$$

Here, g goes over a spatial grid, $||D_{kl}||$ is the density matrix, and the one-electron integral $||A_{kl}||$ is defined through molecular orbitals $\{\phi\}$, and the Coulomb interaction operator $1/r$:

$$A_{kl}(g) = \int \frac{\phi_k(g)\phi_l(g)}{r_{1g}} \, dr_{1g}. \qquad (3)$$

Finally, $||Q_\mu(g)||$ and $||R_v(g)||$ are auxiliary matrices which depend only on the spatial grid and the basis set and which are computed prior to the SCF procedure.

In traditional, Hartree–Fock and DFT calculations, the evaluation of the four-center integrals with the subsequent construction of the Coulomb and exchange matrices is the most time-consuming operation. No four-center integrals are needed in a PS scheme because they are approximated by summations over one-electron three-center integrals. Further, if the number of grid points g is proportional to the number of basis functions $\{\phi\}$, as is usual, then the formal scaling for computing J and K (without the use of cutoffs) is N^3, as opposed to the formal N^4 scaling of traditional approaches that use only basis functions. Cutoffs on integral thresholds and wave function amplitudes can be used profitably in PS assembly of the Fock matrix, and this reduces the formal scaling to N^2 for large systems. The Coulomb term can be reduced to linear scaling by various well-known techniques; in Jaguar, we use multipole methods which lead to linear scaling for large molecules. In principle, one can also achieve linear scaling for the exchange term, but this is typically relevant only for unusual systems such as chains of water molecules or very long linear polymers, so we view this particular limit as not being relevant to the great majority of practical applications.

For DFT, numerical integration is used to evaluate the grid-based parts of the exchange-correlation operator; the details of our numerical integration algorithm can be found in Ref. [6]. PS methods have also been developed for local MP2 (LMP2) calculations,[9] where they provide a formal scaling of N^4 and a practical scaling (with cutoffs) of N^3 for typical relatively large 2- or 3-D systems. The key algorithms here are those for performing the four index transform of integrals from atomic orbitals to molecular orbitals.

The computational efficiency of PS algorithms, for both serial and parallel computation, is demonstrated by performing various types of test cases using the Jaguar implementation of the algorithms discussed above. The prefactor of these calculations, as well as the scaling, is important in determining performance for specific types of calculations. We present timing results for both serial and parallel computation using the latest version of Jaguar in section Jaguar parallelization.

In Jaguar, our principal focus is on hybrid DFT methods for large systems, which provide a good balance of accuracy and efficiency, particularly, for some properties, in conjunction with

our LOC (and DBLOC) described in sections Methods employing LOC and LOC for 3d transition metals (DBLOC). The ability of PS methods to accelerate calculation of the exchange term is critical to this performance. Alternative approaches which are quite similar to the PS methodology, such as RI[60] algorithms, have been developed and successfully applied to evaluation of the Coulomb term in gradient corrected DFT. However, RI methods do not provide the formal scaling improvement for the exchange term that is available from PS techniques. Hence, for DFT calculations, we expect that the greatest competitive advantage will be seen in hybrid calculations.

Initial Guess for Transition Metals

For certain species involving transition metals, it is difficult to arrive at an SCF solution representing the ground state of the system. This is mainly due to the fact that these systems have several low-lying energy states. For such cases, the initial guess wave function has an important impact on the convergence character of the SCF equations. Indeed, for these systems, the standard initial guess algorithms may often populate wrong molecular orbitals, creating instabilities in the SCF procedure or leading to converged solutions representing excited states.

To avoid these convergence issues as much as possible, Jaguar uses a special algorithm to generate high-quality initial guess wave functions for transition-metal-containing systems.[112] This algorithm relies on building successive effective Hamiltonians based on the ligand field theory. The theory regards the electronic structure of an inorganic complex as a perturbation of the electronic structures of its constituents (the fragments of the complex). Jaguar generates the first initial guess wave function by filling orbitals in a way that is consistent with the fragments' formal charges and multiplicities provided by the user or automatically inferred based on the force field. Because different nonbonding d-orbitals from the metal are degenerate, this degeneration is then lifted by assigning a large penalty to orbitals with a large overlap with occupied ligand orbitals, and a large reward to orbitals with a large overlap with unoccupied ligand orbitals. The effects on the ligand orbitals of the unoccupied s-orbital of the metal are subsequently included. Finally, for highly unsaturated complexes, the d-d repulsion of the metal orbitals can also be included.

This high-quality initial guess wave functions algorithm has shown dramatic improvements in convergence success rates: both in convergence toward the correct ground state and in decrease in the number of iterations required to reach convergence.[112] For DFT calculations, it has been shown that the high-quality initial guess wave functions algorithm allows one to find the right ground state in 92% of the cases, whereas it is only possible in 42% of the cases with the standard Jaguar Atomic Overlap initial guess algorithm or in 9% of the cases using a Huckel guess.[112] These statistics help explain why this algorithm has been widely used in both inorganic and bioinorganic applications by different groups. In fact, even when using alternative quantum chemical software to study metal-containing systems, Jaguar initial guess is still often generated to perform SCF calculations, as in cases of Gaussian03,[113–117] Gaussian09,[118,119] and NWChem[120,121] programs.

Important catalysts in the industry such as the Wilkinson's or the Ziegler–Natta catalysts, contain metal atoms.[122] Quantum chemistry is very often used to understand the complex reaction mechanisms involving such catalysts.[123] High-quality initial guess wave functions used for 3d and other transition metals in Jaguar can have a major impact in those studies because control over the spin state of the different species typically involved in the study is of high importance. For instance, Hughes and Friesner[124] used Jaguar's initial guess in a study of the catalytic cycle of a ruthenium oxygen evolving complex to extensively sample the electronic configuration of the metal d orbitals. Such sampling can be important when developing new functionals, particularly to insure that the reported energy corresponds to the lowest energy state found by the functional. This configuration search has also been used in the developments of the DBLOC correction for the B3LYP functional,[32,33,71,125] and is particularly important for bare metal atoms[126] as well as for unsaturated metallic complexes.[127]

One active area of research in inorganic chemistry is the field of molecular magnetism because of the potential applications in the development of nanoscopic storage information units. Indeed, such metallic complexes are not only used to store information in "classical" computing devices (classical bits of computers), but can also be used in quantum computing devices via the so-called qbit storage.[127] The complexes involved generally include several metal atoms per complex, and the molecular magnetic properties of those entities can be characterized by the exchange interactions between the metals and the exchange coupling constant J. To compute J, it is therefore very important to be able to tune not only the multiplicity on the different metal centers but also the relative orientation of their net spin (i.e., ferromagnetic or antiferromagnetic coupling). Such a fine tuning is possible in the initial guess algorithm employed in Jaguar, and it is therefore often used in such cases.[119,127–131]

The Jaguar initial guess algorithm for transition metal atoms has also been frequently applied in computational studies of metalloenzymes.[132–134] The algorithm was very useful, for instance, in the study of cytochrome P450. It helped steer the SCF equations toward the ground-state solution,[32,135–137] and also helped find another low-lying state that is claimed to be relevant to the explanation of the P450 reactivity.[138] The algorithm has also been used to study challenging polynuclear metalloproteins such as methane monooxygenase,[139–141] making possible the computation of the exchange coupling constant between the two iron atoms, and SAM proteins,[142] helping find the ground state of the [Fe$_4$S$_4$] cluster.

Atomic and molecular properties

From the generated wave function, Jaguar is able to compute a large range of chemical properties both in gas phase and in solution. The predicted properties include observable properties such as infrared (IR), VCD, or nuclear magnetic resonance (NMR) spectroscopies, multipole moments, polarizability and hyperpolarizability, as well as nonobservable properties used by chemists such as atomic charges and reactivity indices.

Vibrational Frequencies. Vibrational frequencies is a very important property from both a chemical and theoretical point of view. Frequencies are obtained from the second derivatives of the energy with respect to the atomic coordinates, and as such, are routinely used to characterize the nature of the stationary points found on the potential energy surface. Thus, reactants, products, and intermediates are minima and should only have positive vibrational frequencies, whereas transition states should have one and only one imaginary frequency (in practice reported as negative frequencies). Frequency calculations are computationally intensive, especially for large molecules. An analytic method to compute second derivatives in combination with PS methods makes those calculations far more tractable. The computed frequencies, particularly the very small ones, are subject to numerical errors. The numerical errors can be as high as 50 cm^{-1}, and thus modes with negative values in this range should be visually inspected in Maestro to confirm that they are irrelevant. Additionally, vibrational frequencies are used to compute thermochemical corrections to the energy, and the very low vibration frequencies have a strong impact on the entropic contribution as the harmonic approximation is not generally valid in this region. Those small frequencies should generally be discarded when thermochemical corrections are computed (by default, Jaguar ignores negative frequencies and frequencies with values less than 10 cm^{-1}). Beside numerical errors at low frequencies, vibrations also suffer from systematic errors that can be empirically corrected by scaling factors such as those implemented in Jaguar.[143,144] Recently, we have computed and added to Jaguar a number of optimal scaling factors for modern density functionals such as M06-2X.

Vibrational frequencies are also of interest for direct comparison with experiment because they correspond to IR/Raman frequencies. Jaguar can compute IR intensities, and the corresponding spectra can be plotted in Maestro and compared to experiment. The computed frequencies are also of interest in materials science because, together with polarized Raman spectroscopy, they have recently been used to determine unit cell orientation of oligothiophene crystals.[145] They have also served as chemical descriptors in various quantitative structure–activity relationship (QSAR) models.[146] This is not surprising because specific IR bands have been used to characterize functional groups in molecules, and thus vibrational frequencies can be regarded as encoding chemical features of the molecule in a way similar to fingerprints.

Vibrational Circular Dichroism. Aside from IR spectra, Jaguar is also able to compute VCD spectra.[147] The computed VCD spectra can be used in conjunction with experimental spectra to determine chirality of molecules. Because these spectra are very sensitive to the conformation of the molecule, it is very important to compute a Boltzmann conformational average prior to generating the final spectrum. A script combining MacroModel conformational search and Jaguar is available to compute and plot Boltzmann-averaged spectra in Maestro (see section Automated workflows).

NMR Shielding Constants. In organic chemistry, NMR spectra are routinely generated to determine the structure of the synthesized compounds. It is thus often necessary to compute NMR shifts of known structures for subsequent comparison with experimental spectra. Jaguar computes NMR shielding constants σ very efficiently using a PS gauge-including atomic orbitals algorithm.[148] Once the shielding constants are computed, NMR shifts can be easily obtained using the following formula:

$$\delta_x = \frac{\sigma_{\text{ref}} - \sigma_x}{1 - \sigma/10^6},\qquad(4)$$

where δ_x is the chemical shift of the nucleus in the units of ppm, σ_x is the shielding constant in ppm as reported by Jaguar, and σ_{ref} is the shielding constant of a reference nucleus (^{1}H or ^{13}C of tetramethylsilane, for instance).

Beside helping in the assignment of spectra, computed spectra have also been used to determine the stereochemistry of natural compounds. Generally, such a task is very difficult, but computational NMR calculations can simplify this process and reliably assign stereochemistry with quantifiable confidence.[149,150]

Multipole Moments. The electrostatic field at a long distance from a molecule can be reasonably approximated by the first nonzero term in its multipole expansion. As such, multipole moments can be important molecular descriptors. Jaguar can compute moments up to the hexadecapole. These calculations can take a significant amount of time as they require solving coupled perturbed Kohn–Sham (CPKS) equations. In practice, it is important to remember that the moments are dependent on the global translation/rotation of the molecule when they are not the first nonzero term in the multipole expansion. Thus, the computed dipole moment, which can be visualized in Maestro, will only be independent of the position when the molecule is not electrically charged. For charged systems, the dipole moment can depend on translation or rotation of a molecule.

Polarizability and Hyperpolarizability. Nonlinear optical properties of materials are used in laser technology, telecommunications, optical devices, and high-resolution spectroscopy.[151] Polarizability and hyperpolarizability are difficult to measure experimentally but are straightforward to compute. Jaguar is able to compute polarizability and first- and second-order hyperpolarizabilities with an efficient CPKS algorithm based on PS methods.[152] Recently, an even more efficient computation of polarizability and hyperpolarizability has been implemented in Jaguar through OpenMP parallelization.

Atomic Charges. The concept of partial atomic charges is very popular among chemists because it helps represent the electronic distribution in a molecule by atom-centered values. Yet, atomic charges are not quantum chemical observables and several schemes with different drawbacks exist to derive those quantities from quantum calculations.

Mulliken atomic charges are easily computed from the wave function but generally suffer from basis set dependency. Löwdin charges, accessible in Jaguar when using the SM8 solvation model,[153] somewhat reduce this dependency, but

remain unstable for very large basis sets. On the other hand, natural population analysis charges derived from a natural bond orbital (NBO) analysis[154] do not suffer from basis set dependency but tend to be the largest atomic charges in magnitude.

The atomic charges that are most commonly used, particularly in force field development, are electrostatic potential (ESP) charges.[155] In Jaguar, they can be derived by fitting atomic charges so that they best reproduce the ESP of the molecule, and (if requested) its multipole moments. Those charges can be used in other programs like our docking program Glide when the OPLS-AA charges are not optimal for certain molecule.[156] Nonetheless, ESP charges are dependant on the molecular conformation used, and restrained electrostatic potential charges[157,158] which offer the possibility to perform averages over several conformations are sometimes preferred.[159]

Another method to derive partial charges is to partition the electron density and to assign different regions of space to atoms. Jaguar can compute these so-called Stockholder charges[160] by first assigning the density to an atom. This density is the integral of the electronic density weighted by the ratio of the density of this isolated atom to the sum of the densities of all the isolated atoms of the molecule. Stockholder charges are generally much smaller than other types of partial charges, but are less sensitive to basis set effects.

Fukui Functions and Atomic Fukui Indices. To quantify the reactivity of a molecule toward various types of agents, Jaguar provides Fukui functions[161] and atomic Fukui indices.[42,43] Fukui indices are derived from an approximation of the Fukui functions and are condensed to atoms in the same way as Mulliken charges are computed.

Fukui functions, dependent on the total electronic density, are generally regarded as more accurate than atomic Fukui indices which are based on frontier orbitals only. For the identification of a nucleophilic center, the user must look for the areas where most negative values of the Fukui function f^- are concentrated. Likewise, for the identification of an electrophilic center, the areas with the maximal positive values of f^+ are of interest. See section Jaguar use in Drug Discovery for details on applications of Fukui functions to prediction of sites of metabolism in computational drug design, and Figure 6 for an illustration of a f^+ Fukui function.

Atomic Fukui indices are convenient because they represent reactivity of atoms in the molecule through scalar characteristics. Jaguar provides four indices (f_NN, f_NS, F_SN, F_SS) for both the HOMO and the LUMO. Those coefficients describe the variations in density N or spin multiplicity S relative to small variations in density N or in spin multiplicity S. The indices f_NN for HOMO and LUMO are therefore likely to be the most useful descriptors as they report the sensitivity of atoms toward electrophilic and nucleophilic attacks, respectively. Fukui indices are valuable in many applications, for example, for determining the regioselectivity of a reaction,[162] and comparing the reactivity among molecules.[163] For examples of
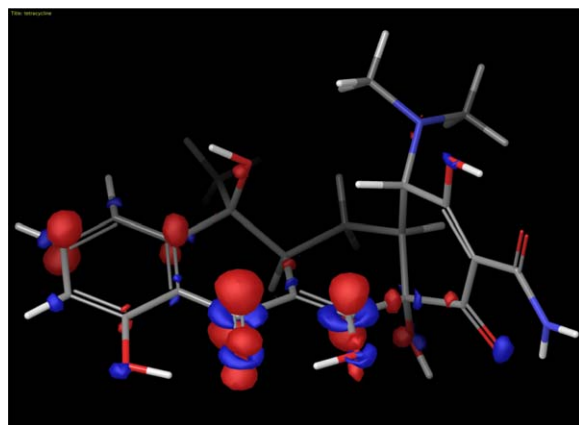
Figure 6. The f^+ Fukui function of tetracycline, indicating where charge builds up on adding 0.01 electron to the neutral molecule. Image generated by Maestro 9.3.
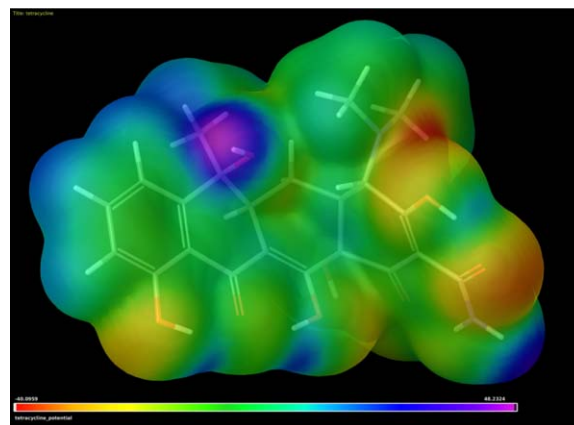

Figure 7. The ESP of tetracycline mapped onto the isodensity surface of 0.001 electrons/bohr³. Image generated by Maestro 9.3.

practical usage of Fukui indices implemented in Jaguar, see Refs. [164] and [165].

In such applications, however, one must be aware that Fukui indices are not "size-consistent." For instance, if one computes the f_NN HOMO Fukui indices on the water molecule alone and a system of two water molecules separated by 20 Å (which is essentially a noninteracting dimer), then one can find that the f_NN HOMO on oxygen of the water dimer is half the value found for water alone. This is obvious even from the fact that Fukui indices are normalized and sum up to 1.0. So the use of Fukui indices can prove not straightforward in certain cases, and one may prefer atomic descriptors such as the average local ionization potential which can be obtained by generating the corresponding surface.

Molecular surfaces

Jaguar calculates many volumetric and surface properties which can be visualized using Maestro. These properties include molecular orbitals, electron charge and spin densities, ESP, ALIE, and Fukui functions. The properties are evaluated on automatically generated rectilinear grids, and the data are written to files in HDF5 format.[166] Maestro can display the data as isosurface contours, as is commonly done for molecular orbitals or electron densities, or as color-coded surface maps, which is typical for ESP or ALIE. Maestro offers many options for controlling surfaces, including colors, lighting, transparency, mesh, wire-frame, and solid rendering.

An ESP map generated by Maestro using Jaguar's calculated values is shown in Figure 7 for tetracycline. The f^+ Fukui function, which shows how the electron density shifts when a fraction of an electronic charge (in Jaguar, by default 0.01 of the electric charge) is added to the neutral molecule, is shown in Figure 6.

For bimolecular systems, NCI can be visualized using the method of Johnson et al.[167] In this method, the magnitude of the electron density signed by the second eigenvalue of the density Hessian is mapped onto regions of very low-reduced density gradient. Van der Waals interactions are indicated by

regions of low gradient and low density, while stronger interactions such as hydrogen bonds tend to have higher density values. In addition, Jaguar can locate critical points in the electron density, which allows Maestro to draw bond paths between the interacting atoms. The significance of critical points and bond paths is described by Bader.[168,169] Figure 8 shows an NCI plot for the equilibrium geometry of the *N*-methylacetamide dimer. The light yellow regions, where the electron density is relatively high and the sign of the second eigenvalue of the density Hessian is negative, indicate regions of favorable intermolecular interaction, which is confirmed by the presence of the bond paths (dotted lines) passing through these regions and connecting pairs of interacting atoms. Plots like this are of interest to chemists because they make it clear where the actual bonding interactions take place in molecular complexes.

Another fundamental property in the Bader's quantum theory of atoms in molecules is the Laplacian of the electron
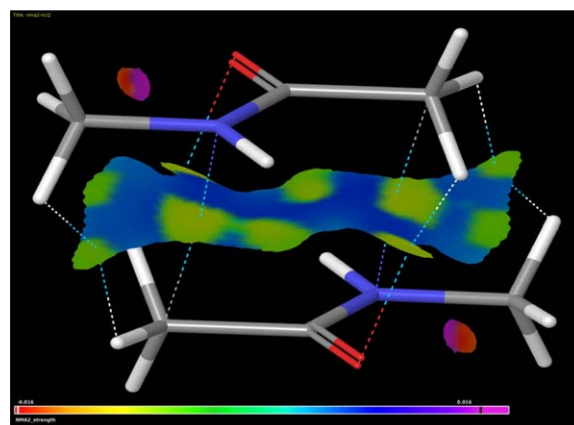

Figure 8. An NCI plot for the equilibrium geometry of *N*-methylacetamide dimer. Weak internal C—H···O hydrogen bonds are indicated as disks of purple and red, whereas the intermolecular bonding is indicated by the yellowish regions, in which there are bond critical points in the density. The bond paths through these critical points, connecting bonded atoms, are indicated with dashed lines.

density, which shows regions in molecular systems where charge is locally concentrated or depleted. This in turn can reveal reactive sites in molecules, places for nucleophilic or electrophilic attack. An example is shown in Figure 9, where the surface for which the Laplacian is zero is shown for methyl acrylate. Regions of charge depletion are seen as holes in this surface, at the terminal alkenyl carbon (where the nucleophile in a Michael reaction would attack, for example), at the carbonyl carbon, and in the tetrahedral faces of the methyl group. The core electrons on the carbon and oxygen atoms are also apparent as nearly spherical shapes inside the outer envelope.

Solution phase calculations

Multiple computational chemistry applications require inclusion of solvation effects, and the literature in the development and application domains of solvation studies is vast (see, e.g., reviews on continuum solvation in Refs. [170–172]). Jaguar
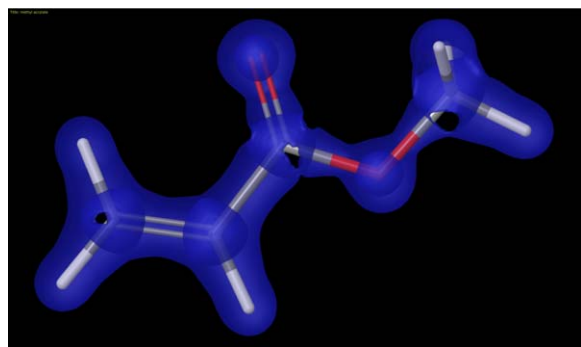

Figure 9. The surface for which the Laplacian is zero for methyl acrylate. Image generated by Maestro 9.3. See the text for details.

offers essentially two continuum solvation models. The first one, SM8,[153] is based on the Born model and is a replacement of the earlier SM6 model[173] (which is also available in Jaguar). The second one, called PBF (which stands for "Poisson–Boltzmann Finite" elements) is based on an iterative solution of the PB equation, until the electric potential on the molecular surface, with contributions from the solvent and the solute, reaches equilibrium.[14–16]

Both SM8 and PBF models can predict similar properties, contain parameters for a number of solvents, and are actively used in Jaguar applications (see, e.g., Refs. [174,175]) However, from the Jaguar user's point of view, there are important differences. The PBF model works in conjunction with either Hartree–Fock, DFT, or LMP2 theories, whereas SM8 is only available for the first two of these. Then, geometry optimizations in solution are only possible with PBF, as our current implementation does not contain SM8 energy gradients. It must be noted, however, that the PBF gradients, as implemented in Jaguar, are not fully analytic. This is a consequence of an implicit dependence of the solution-phase energy on the numerically generated grid which is used to solve the PB equation. The numerical approximations that are required to compute some gradient terms might result

in a certain level of noise for particularly challenging and flexible molecules, but special techniques and options are available for Jaguar users to minimize or eliminate this undesirable effect. Finally, PBF does not have restrictions for what basis set to use, whereas atomic charges in the SM8 model can only be computed efficiently with several recommended basis sets (if some other basis set is used, then less accurate Löwdin charges will be applied). Both models use empirical corrections to account for first solvation shell effects which are not captured by pure implicit solvation models. Recently, we improved parameters to treat first shell effects in ions in aqueous PBF calculations, so that the average solvation energy error for ionic species, evaluated on standard test sets,[153] is about 3 kcal/mole, which is close to the experimental error. The experimental error for solvation of neutral species is smaller, around 1 kcal/mole, reflected by similar error bars produced by PBF for these species.

Jaguar parallelization

Parallelization has become a ubiquitous feature of quantum chemistry codes.[176–178] Jaguar calculations can also be conducted using multiple central processing unit (CPU) cores for significantly increased computational speed. Batch calculations on multiple structures may be distributed across multiple cores, providing essentially 100% efficiency. But calculations on single inputs may also be carried out across multiple cores using either the MPI[45] or OpenMP threads,[44] or a combination of both (support for OpenMP is new and will be available in the 2013 release of Jaguar).[73] An MPI calculation may use multiple host computers, whereas each team of OpenMP threads must be launched on a single compute host. A combined MPI/OpenMP implementation allows one to perform calculations on very large molecules containing hundreds of atoms and thousands of basis functions, by distributing the calculation across multiple compute hosts with MPI, where each MPI process then launches a team of OpenMP threads.

Jaguar is distributed with the Open MPI implementation[179] of the MPI 2.0 specification. However, Jaguar does not make direct calls from its Fortran routines into the MPI libraries, but calls MPI wrapper functions instead. This allows users to replace the Open MPI library in the Jaguar installation with another MPI implementation which can provide enhanced performance in cases where the local MPI library has been tuned for a particular network interface. With Open MPI, Jaguar offers out-of-the-box support for common high-performance network interconnects such as Infiniband and Myrinet, plus standard TCP/ethernet switches.

Similarly to the other *ab initio* quantum chemistry codes, Jaguar's operations are CPU, bandwidth, and I/O intensive. Most of Jaguar's CPU-intensive, PS algorithms are parallelized using either MPI[180,181] or OpenMP, or both. These include HF and DFT energies, gradients, integral second derivatives, charge-fitting, grid generation, LMP2 energies, CIS/TDDFT excitation energies, and the generation of many of the matrices that are used in the PS algorithm. The CPHF/CPKS algorithms have also been parallelized, which significantly speeds up the

calculation of vibrational frequencies, NMR shielding constants, VCD intensities, polarizabilities, and hyperpolarizabilities.

Figures 10–12 show Jaguar's performance using either MPI or OpenMP in single-point energy calculations, gradient calculations, and vibrational frequency calculations, respectively. These calculations were performed using development versions of Jaguar 8.0 on a cluster in which each compute node contained two sixteen-core AMD Opteron 6274 processors. Each node was equipped with its own local scratch disk and an on-board gigabit ethernet network interface. The MPI calculations were conducted with no more than four processes per compute host, to reduce the cost of I/O contention for the scratch disk.[†] In these calculations, the unused processors were kept idle.

The performance of Jaguar's OpenMP and MPI implementations is similar up to four processes/thread, but for higher process/thread counts, the OpenMP parallelization shows better performance. Some of the difference in speed is due to the fact that MPI must pass messages over the network. We per-
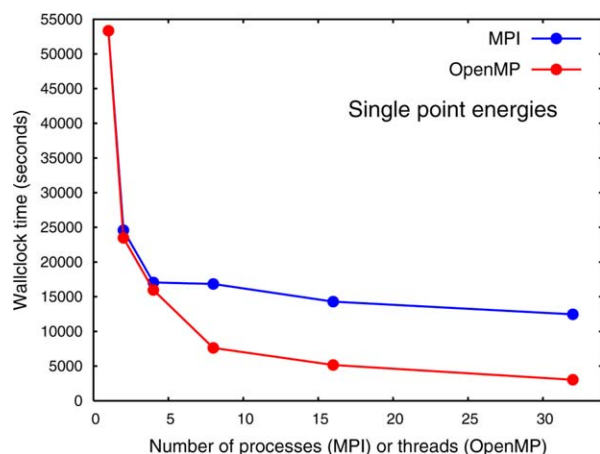

Figure 11. Total calculation walltimes for M06-2X/LACVP* geometry optimization calculations on ramoplanin (335 atoms, 2846 basis functions, no symmetry) on a computer cluster equipped with AMD Opteron 6274 processors.

Figure 10. Total calculation walltimes for M06-2X/LACVP* single-point energy calculations on ramoplanin (335 atoms, 2846 basis functions, no symmetry) on a computer cluster equipped with AMD Opteron 6274 processors.

Figure 12. Total calculation walltimes for M06-2X/LACVP* vibrational frequency calculations on tetramethrin (49 atoms, 386 basis functions, no symmetry) on a computer cluster equipped with AMD Opteron 6274 processors.

formed additional tests which allowed more than four MPI processes per host and found that the calculations were significantly slower. This is due to the contention of so many MPI processes for the scratch disk, which overwhelms any gain in speed due to confining the MPI processes to the same compute host. The MPI implementation is still very efficient given separate scratch disk across different nodes, or sharing using high-performance RAID disks.

[†]We have found that the I/O time increases significantly when 8 or more MPI processes are using the same scratch disk during the same calculation. Solid-state drives and redundant arrays of independent disks (RAID) improve performance somewhat, but contention for the disk still yields poor scaling. It is best to distribute MPI processes across compute hosts, each of which is equipped with its own local scratch disk.
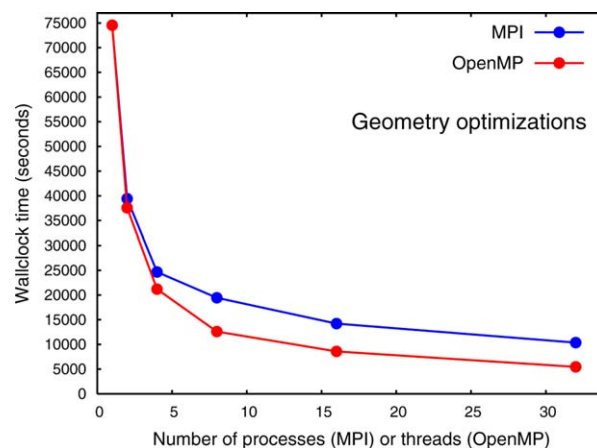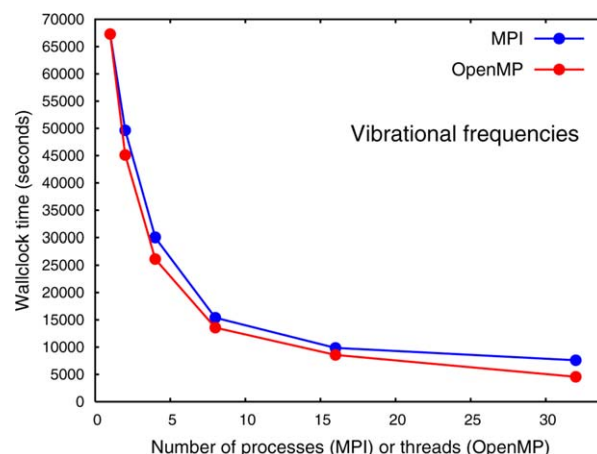
We also conducted mixed MPI/OpenMP calculations to investigate whether this would improve performance. When using both MPI and OpenMP, each MPI process creates its own team of OpenMP threads which runs locally on the same host as the parent MPI process. The results are shown in Table 2. In this table, "total cores" is the product of the number of MPI processes and the number of OpenMP threads, and serves as an estimate of the hardware resource demand. In nearly all cases, using OpenMP alone is the most efficient means of running these types of calculations, and where a mix of MPI and OpenMP is faster, it is only by a small margin.

Sixteen-core AMD Opteron 6274 processors are relatively inexpensive in that they deliver good value per core (performance considerations aside). However, it should be possible to retrieve higher performance from Jaguar on more costly and higher-performance hardware. As an example, in Table 3, we

Table 2. Total calculation walltimes for Jaguar single-point, geometry optimization, and vibrational frequency calculations using MPI and OpenMP on a Linux cluster equipped with AMD Opteron 6274 processors.

OpenMP threads	MPI processes	Total cores	Energy walltime (s)	Gradient walltime (s)	Vibrational frequency walltime (s)
1	1	1	53,350	67,294	74,544
2	1	2	23,485	45,118	37,567
1	2	2	24,586	49,651	39,415
4	1	4	15,969	26,133	21,195
2	2	4	15,729	25,483	30,717
1	4	4	17,061	30,071	24,659
8	1	8	7618	13,568	12,604
4	2	8	8816	13,869	15,545
2	4	8	9393	13,044	17,203
1	8	8	16,841	15,387	19,449
16	1	16	5162	8572	8577
8	2	16	5177	8114	9202
4	4	16	5555	7774	9758
2	8	16	6371	9355	10,107
1	16	16	14,293	9857	14,211
32	1	32	3028	4550	5457
16	2	32	3417	4506	6507
8	4	32	3661	5108	6432
4	8	32	4169	6243	6970
2	16	32	4962	7154	7254
1	32	32	12,468	7594	10,359

For the rest of the conditions see captions to Figures 10–12.

Table 3. Total calculation walltimes for Jaguar B3LYP geometry optimizations of a nanoparticle system with the brutto-formula $Ti_{61}H_{116}O_{180}$ carrying 3014 basis functions.

OpenMP threads	MPI processes	Total cores	Gradient walltime (s)	Speedup
6	1	6	49,567	1.0
12	1	12	27,603	1.8
12	2	24	15,094	3.3
12	4	48	8569	5.8
12	6	72	6190	8.0
12	8	96	5490	9.0
12	10	120	4348	11.4

The calculations using no-symmetry conditions were carried out on the Lonestar Linux cluster equipped with Intel Xeon 5680 processors. The speedup is given with respect to the calculation carried out with six OpenMP threads.

are presenting wall-clock timings and speedups obtained with a development version of Jaguar for geometry optimizations of a molecular cluster consisting of 61 molecules of TiO_2 and 58 molecules of H_2O. The molecular system is illustrated in Figure 13. The calculations presented in the table were carried out on the Lonestar Linux cluster located at Texas Advanced Computer Center. Each node on the cluster carried two 6-core Intel Xeon 5680 processors. Table 3 presents Jaguar calculations performed on as many as 120 total cores simultaneously. Without this scale of parallelization, projects such as investigation of electronic structure of TiO_2 nanoparticles[72] would be impossible to carry out in reasonable time.

The Cartesian coordinates of all the systems used for benchmarking in this section are given in the Supporting Information.

Automated workflows

Some combinations of computational tasks occur in one's work multiple times. Examples of such "workflows" are geometry optimizations followed by single-point energy calculations in a larger basis set, or a counterpoise correction computation which requires a series of computations on a molecular complex and its substituents. One should be able to easily automate such workflows.

Jaguar provides two types of automated workflows. The first type is represented by the so-called batch scripts (files with the extension ".bat"). These are essentially text files with simple syntax rules which allow the user to specify which individual calculations must be launched, what their order should be,

and what results of previous calculations should be used in subsequent calculations. A number of batch files performing standard quantum chemical operations are distributed with Jaguar. The second type of workflows is standard workflows. Those are typically Python programs which are included with Jaguar and which address some of the most common complex computational tasks, such as counterpoise correction and Fukui function calculations. Standard workflows come with GUI support, contain multiple options, and often perform operations that go beyond the simple functionality of batch scripts. The rest of this section briefly describes the function and capabilities of the standard workflows.

The counterpoise correction workflow computes a counterpose-corrected binding energy of a complex comprising two noncovalently bound fragments. Additionally, the counterpoise correction itself is reported. The user has the option to change the level of theory, the basis set, and other settings, although reasonable default settings are provided.

The hydrogen bond workflow computes the binding energy of a complex of two molecules interacting via a hydrogen bond. The total binding energy, corrected for the basis set superposition error, and extrapolated to CCSD(T) reference energy, is reported. According to our tests, the extrapolated interaction energy should be within 0.5 kcal/mole of the CCSD(T) reference value. The underlying computational protocol is based, by choice, either on LMP2 or X3LYP computations in large basis sets, and is essentially equivalent to a method from Ref. [182]. Details of geometry needed for the workflow can be controlled by the user. For the examples of usage of the script in practice, see Refs. [183,184].

Another standard workflow performs scans of a potential energy surface by distributing independent datapoint calculations across multiple processors. One- and two-dimensional scans can be performed. The user can choose between relaxing all degrees of freedom except the scanned one (relaxed scan) or keeping degrees of freedom "frozen" (rigid scan). Types of degrees of freedom that can be scanned are Cartesian coordinates, interatomic distances, bond angles, and dihedral angles. The workflow has a mechanism to keep track of
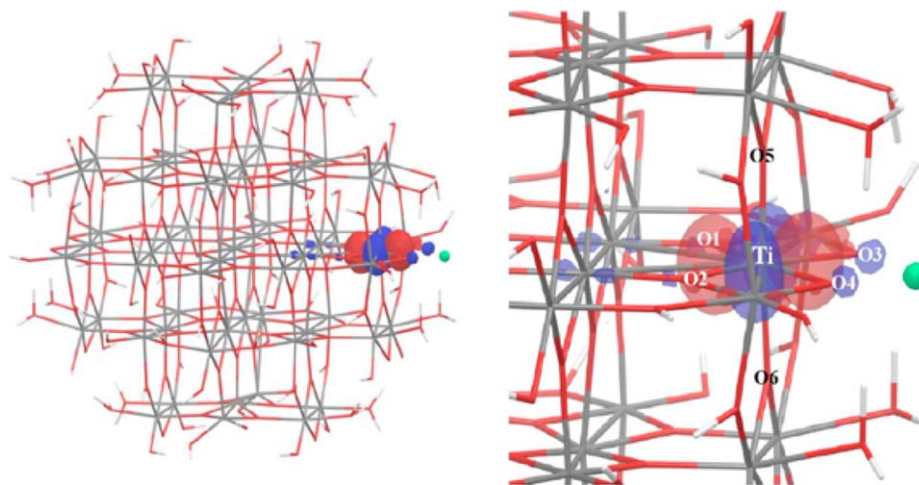
Figure 13. B3LYP-LOC-simulated TiO$_2$ 5 × 5 × 5 cluster with surface-absorbed Li$^+$ ions and an excess electron in continuum water (left); close view of location of excess electron, lithium in green (right). (Reproduced from Ref. [72] with permission from American Chemical Society).

individual scan points that may have failed for some reason, and can easily restart those calculations.

Fukui atomic indices can be computed in a regular Jaguar calculation, but Fukui functions are computed with a special automated workflow. The user can choose which Fukui function, f^- or f^+, or both, should be calculated, and can also specify the levels of theory and control some details of the algorithm.

VCD spectra can be computed with a standard workflow which has the capability to perform a conformational search prior to computing the spectra (see section Vibrational Circular Dichroism on VCD computations themselves). In the workflow, the user can vary a number of conformational search options such as the force field, the number of conformers to retain, and so forth, and also choose between a vacuum or a solvated calculation.

A new workflow that is included in Jaguar in 2013 can be used to determine the ground states of molecules that might have open shell character, and is specifically designed to alleviate some of the difficulties in calculations on molecules containing transition metals. The script will create initial files with several possible multiplicities and different initial electronic configurations, then launch the jobs, and comparing the final energies will determine which of those guesses produces the ground state. For example, the script will determine that the ground state of the O$_2$ molecule is triplet, and that of the [Fe(III)(H$_2$O)$_6$]$^{+3}$ complex is sextet.

A workflow for computing enthalpies of formations and atomization energies at the DFT level is available in the 2013 Jaguar release. The workflow uses precomputed atomic energies for several combinations of atoms and basis sets for atomization energies. Zero-point vibrational energies with the optimal scaling factors for the given functional are used when constructing the enthalpy. The B3LYP-LOC functional in combination with the 6-311G-3df-3pd++ basis set is recommended for the most accurate predictions.

Finally, a pKa prediction workflow is available. It will be described in considerable detail in the next section.

All the automatic workflows can be executed on multiple processors, with individual calculations distributed over separate machines, and can additionally use the MPI and OpenMP levels of parallelism.

Schrödinger also provides a diverse set of Jaguar automated operations that can be used in a KNIME[185] setup as nodes. For instance, a single KNIME workflow using such nodes can automatically convert a number of structures from one format to another, submit them to a MacroModel conformational search calculations, filter the generated structures, submit the resulting conformations for geometry optimization in Jaguar, and then use the final optimal structures for single-point calculations with a large basis set.

pKa prediction

Prediction of pKa is one of the key tasks in the practice of computational drug design. Acidity of individual sites or whole small molecules is a factor when considering blood-brain barrier penetration, preparing protein structures, as a descriptor in SAR models, and so forth.[186–189] Prediction of pKa is also used in organic synthesis design and in various applications in materials science.[190,191]

Jaguar offers a sophisticated module for pKa prediction which will be referred to as Jaguar pKa. It is an automated workflow based on a simple thermodynamic cycle shown in Figure 14. Legs A, B, C of the workflow are modeled with a DFT calculation, so that the whole approach is *ab initio* in nature, in contrast to numerous empirical approaches.[192,193]

Prediction of pKa based on *ab initio* calculations remains difficult even with the increasing quality of the available theoretical methods, mainly because 1 pKa unit of error corresponds at 298 K to a mere 1.37 kcal/mole of the total error in energy in whatever thermodynamic cycle is used. Although the first error is usually considered large because experimentally pKa can be measured with an accuracy of about 0.01 pKa unit, the second one is quite acceptable in most DFT calculations. To
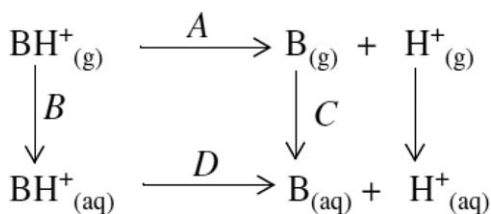
Figure 14. The thermodynamic cycle used in the Jaguar pKa automatic workflow. The final "raw" pKa value is computed according to the equation pKa$=D/(2.3RT)$, where D is a leg of the thermodynamic cycle shown in the figure. D is assembled from legs A, B, C, which are in turn computed by Jaguar.

bring down the error of DFT-based pKa predictors to about 0.1 pKa unit, a small number of empirical corrections in addition to *ab initio* energies is often used.

Compared to *ab initio* pKa predictors, empirical, QSAR-like algorithms are much faster and can be more accurate when acting on common structures and functional groups.[194] However, they perform much less accurately when confronted with structures which are not well represented in the underlying look-up databases, and usually cannot handle very unusual structures such as those encountered in materials science. The presence of multiple conformations, nontrivial tautomeric equilibria, and resonance forms exacerbates the quality of empirical pKa predictions.

Higher-level methods such as coupled cluster have been used for *ab initio* pKa prediction.[195–197] Unfortunately, these approaches are much more computationally expensive, and in any case evidence remains to be seen if they can be applied to predicting pKa of diverse chemical structures as accurately without the need to adjust parameters of protocol from one functional group to another.

Jaguar pKa uses two parameters per functional group, exploiting the often excellent linear correlation between the raw pKa predicted by the thermodynamic cycle and the experimental pKa. Versions of Jaguar pKa from 2012 and earlier included parameters for a couple dozen functional groups common in organic synthesis and drug discovery. If the parameterization for the desirable functional group was not available, the calculation would not be able to proceed.[193] Although arguably being more reliable than empirical pKa predictors for such functional groups as nitrogen-containing "heterocycles," Jaguar pKa used to have a somewhat limited coverage of chemical space. According to our estimations, such coverage in the previous versions of Jaguar amounted to less than 50 percent of protonation/deprotonation-prone chemical groups in a typical database of drug-like compounds. The new version of Jaguar pKa features an improved algorithm that provides an essentially 100 percent coverage of functional groups, due to its ability to "zoom out" on the next-best, less specific functional group if a certain functional group is not found in the program's parameterization database. This pKa prediction scheme, dubbed "shell model," is illustrated in Figure 15. Our multiple benchmarks showed that this approach gives quite accurate, and certainly very useful in practice, pKa predictions even for some of the least common functional

groups such as Si—OH. Besides, the shell model in the 2013 release of Jaguar enjoys parameters for an additional 27 or so highly specific functional groups. The model and its performance will be described in detail elsewhere but some illustrative results are presented in Figure 16 and in Table 4.

Figure 16 shows that the Jaguar pKa model available as part of the Jaguar release in 2012 would not be able to predict pKa of the common drugs phenylbutazone, vitamin B$_6$, and acetarsone, because it lacked specific parameters for keto-CH acids, phosphates, and arsenates. The new shell model con-
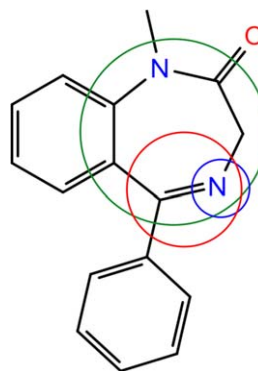

Figure 15. This figure illustrates the idea behind the Jaguar pKa shell model. In this example, three pKa parameterizations are available for the protonation of the imine nitrogen. Each circle shows the specificity of the parameterization. The green circle matches a larger part of the structure and thus corresponds to a more specific (and usually more accurate) pKa parameterization. This parameterization was trained only on molecules sharing the structural element indicated by the green circle. The red circle indicates only a CRR′$=$NR″ match and is therefore less specific. The least specific match is generic: it will match any atom (indicated by the blue circle), but it is the least accurate. The pKa predictions corresponding to all the matched models can be printed along with their least squares' RMSD values. Only the prediction coming from the parameterization with the lowest RMSD is usually of interest.

tains parameters for those groups and predicts the pKa's of the aforementioned drugs in good agreement with the experiment.

Several important areas of the large scientific field of computational pKa prediction remain complicated subjects and are still waiting to be properly addressed or automated. They include handling zwitter-ions, predicting pKa in complex equilibria involving multiply charged species and tautomers, predicting macro-pKa's, assigning entropic corrections due to symmetrically equivalent functional groups, predicting pKa in complex protein environments such as those involving transition metal ions, and so forth. Because of the high priority of the problem of pKa prediction in various areas of chemistry, Jaguar is dedicated to continuing improvement of its pKa module to address these questions. For instance, the 2013 version of Jaguar pKa is able to apply specific parameters for high-quality pKa predictions of some of the most common zwitter-ions such as NR$_3^+$—…—COO$^-$ and also alpha-amino acids. It was necessary to alter the details of our workflow to handle amino acids: geometries optimized in solution had to be used for leg A in Figure 14. The new workflow for amino

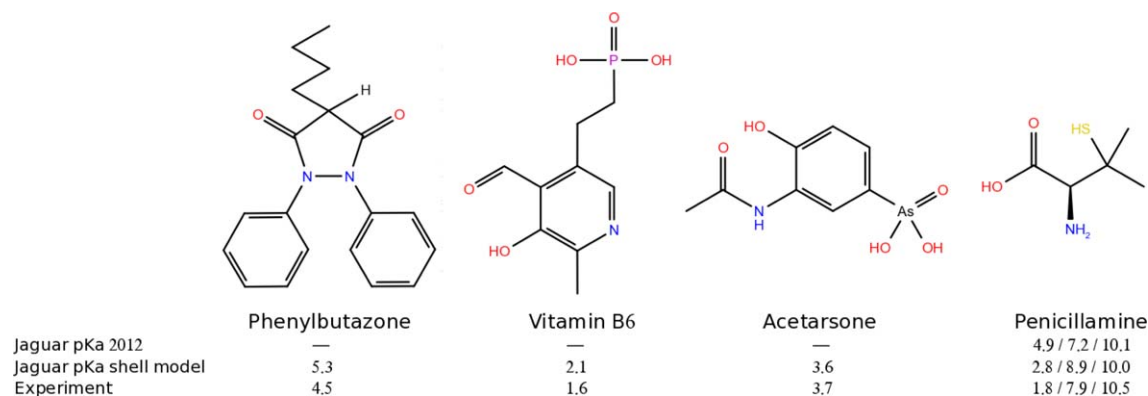|	Phenylbutazone	Vitamin B6	Acetarsone	Penicillamine
Jaguar pKa 2012	—	—	—	4.9 / 7.2 / 10.1
Jaguar pKa shell model	5.3	2.1	3.6	2.8 / 8.9 / 10.0
Experiment	4.5	1.6	3.7	1.8 / 7.9 / 10.5

Figure 16. Prediction of pKa values of actual drugs. Jaguar pKa 2012 lacked specific parameters for the functional groups that are deprotonated in these compounds, and therefore was unable to generate a prediction. The structures shown were not in the training set of the Jaguar pKa shell model.

Table 4. Performance of Jaguar pKa 2012 and a development version of the Jaguar pKa shell model on a test set comprising 193 molecules.

	Jaguar pKa 2012	Jaguar pKa 2013
Coverage of CACDB sample	42%	100%
Number of supported functional groups	28	59
MUE on test set, pKa units	1.12	0.58
Number of outliers in test set	100	31

Note that a few molecules from this test set were included in training the pKa shell model. The CACDB is a database of drug-like molecules. A random sample used in this study contained about 50,000 molecules and approximately 183,000 protonation sites. Comparison of predicted pKa values is with respect to the experiment. The prediction was considered an outlier if the predicted value differed from the experimental value by at least $\pm$ 1.0 pKa unit.

acids predicts both the amino-group and the carboxyl-group pKa's more statistically accurately, which is illustrated in Figure 16 by the results for the drug penicillamine.

Jaguar use in Drug Discovery

The modest accuracy of current empirical methods used in drug design and the large number of factors that contribute to the eventual success of a drug do not allow one to reliably "design" a drug yet.[198] The improvement and more frequent explicit use of quantum mechanical methods is believed to be one of the most promising paths toward reducing the level of uncertainty inherent in computational modeling.[198,199]

Jaguar is involved in a great number of applications in drug discovery modeling. For many, if not most of these applications, it would not be possible to substitute Jaguar with tools based on MM force fields or empirical methods. Occasionally, MM methods are reported to yield better performance than ab initio methods, but such conclusions are typically results of suboptimal use of ab initio methods.[200]

Analysis of ligand conformation on binding to target is of high importance for predicting potential bioactive conformation, binding energy, and strain penalty associated with the binding process. Such calculations require the determination of global minimum energy conformation in solution phase and its associated gas-phase energy.[201–203]

Accurate determination of positions of hydrogen atoms, usually unavailable from experimental x-ray data, coupled with prediction of the relative stability of various tautomeric forms, is another application of Jaguar in drug discovery and other areas.[204,205] Cores may be redesigned to stabilize preferred tautomers.

The Jaguar pKa module is used actively for predicting pKa of novel design patterns, particularly involving nitrogen atoms in unusual heterocycles. Oftentimes, very accurate predictions of pKa are not required, and the universality of application to any types of structures, even rare, and undescribed in the literature, becomes much more important. In such cases, pKa predictions based on empirical lookup tables cannot be carried out, at least with sufficient reliability, because of the paucity of the training data. As an illustration, consider a design of a drug with pKa of around 8.0, so that the cationic form predominates, but the basicity is not as high as to block membrane permeation. When the design structure involves a novel pH-sensitive structural element, a prediction by the Jaguar pka shell model with a relatively high error of $\pm$ 1.0 pKa units (for some nonspecific functional groups) can still be useful information. Jaguar pKa predictions of 6.0 and lower or 10.0 and higher would likely result in abandoning the candidate, whereas predictions in the range of 7.0–9.0 pKa units would be favorable. In contrast, an empirical method which was not trained on structures of this rare type (which can be known or unknown to the user) can either exit with no prediction, or yield a prediction that would be so unreliable as to obliterate its usefulness.

Identifying sites of reactivity is another frequent application of Jaguar in medicinal chemistry, especially in connection with predicting metabolic pathways.[206,207] Jaguar is well-suited for predicting the so-called "intrinsic" reactivity, that is, reactivity of an isolated structure for which it is assumed that the reactive site is not sterically hindered. Descriptors of intrinsic reactivity available in Jaguar are HOMO/LUMO orbitals, Fukui functions and atomic indices, ESP surfaces, ALIE surfaces, and

sometimes partial atomic charges. The ready availability of the reactive site typically assumed when computing intrinsic reactivity may be a drastic approximation even for some isolated structures in which bulky elements of the structure itself can present an obstruction. Structures that must fit tightly in the active site of an enzyme before the reaction can take place often actually react at sites which, according to intrinsic reactivity, should be inert. Thus, cytochrome P450 often oxidizes aliphatic hydrogen atoms, which are, according to Fukui functions, unreactive.[208] In such cases, it is advisable to combine intrinsic reactivity calculations with docking, if the actual reactivity is to be obtained. Basically, if every site of the molecule is exposed and the fit is loose, the most intrinsically reactive site is likely to attack or to be attacked. However, if the fit is tight, then only the correctly positioned site should be considered for a reaction. Of course, one need not forget about entropic factors which might complicate the prediction. To assist with prediction of P450 sites of oxidation, Schrödinger provides a computational workflow called "P450 Site of Metabolism" which combines intrinsic reactivity prediction by Epik and induced fit docking by Glide to yield the combined reactivity. It is expected that in the future Jaguar will play a role in this workflow by providing intrinsic reactivity, as an option. In some complex situations, it is still recommended to forego simple descriptors of reactivity and properly model the energetics of reactions.

NMR and VCD spectral predictions allow Jaguar to contribute to deciphering or confirming structures of organic compounds that serve as precursors to drug candidates. These computational features are used by chemists who work with modelers.

Force fields usually give high-accuracy geometric predictions for well-known and, consequently, well-parameterized structural elements. When the structural element is rare, then force fields might lack parameters for it or give an unreliable prediction. In such cases, the use of quantum mechanical methods is invaluable. One example of using Jaguar in the area only sparsely covered in force field development is (hetero)cyclic flexible structures, which can be partially or fully saturated and potentially fused. These scaffolds offer high novelty of composition of matter and can be extremely valuable for some drug targets. Even when the parameters are available for some unusual ring systems, getting all the energetics of the (pseudo)axial versus (pseudo)equatorial bond flips for a given novel ring structure and substitution pattern just right is a great challenge when not using *ab initio* calculations. Additionally, noncovalent interactions, particularly in situations when electronic effects of substituents have to be considered, when there are cation-π interactions, and when heavier atoms such as Cl and S are involved, are best accounted for by *ab initio* methods.[209] Jaguar offers a series of DFT methods to treat such interactions: apart from sophisticated modern functionals such as M06-2X, *a posteriori* corrections D, D3, ulg, and MM, address specifically noncovalent interactions. See section Jaguar use in Force Field Development for more detail on the use of Jaguar in force field development.

Common force fields are usually totally inapplicable to interactions or reactions involving transition metals. In such cases,

the use of Jaguar or QSite for prediction of geometries, coordination numbers, spin states, ligand protonation states, strengths of interactions, and other details of reaction energetics, is very common.

Finally, hydrogen bond strength is an important characteristic of many interactions in medicinal chemistry, such as interactions between the water molecule with the ligand. Jaguar offers two automatic methods to predict the energetics of such interactions: the hydrogen bond script (see section Automated workflows) and the B3LYP-MM functional.

Jaguar's current use in drug discovery modeling is considerable, which can be seen at least from the fact that many of the world's largest pharmaceutical companies routinely apply Jaguar in their research and development projects (see, e.g., Refs. [175,207,210–213]). Among advantages that make Jaguar very well-suited for pharmaceutical *ab initio* modeling are its speed, automation of key modeling tasks, integration with the rest of Schrödinger modeling software through Maestro, and a high level of technical support. Many expectations for the greater use of *ab initio* methods in drug discovery are placed on the future development of more accurate computational methods and the continued progress in creating faster computer hardware. However, even with the accuracy and speed that are available now, one can imagine directions for improvement that would be capable of bringing Jaguar to a new level of usability in pharmaceutical modeling. These improvements include: (i) implementation of computational methods that are still missing from Jaguar and that are indispensable in research; (ii) increasing the ease of use through more harmonious software design (e.g., a more convenient setup of transition state and QM/MM calculations); (iii) automation of a greater number of key tasks; (iv) enhancing the robustness of key components, particularly initial guess, SCF convergence, geometry optimization; (v) ranking methods and protocols by accuracy and area of applicability through validation and benchmarking; (vi) providing medicinal chemists and modelers with training materials to keep them up to date with the developments in Jaguar and in the scientific field of computational chemistry. The enumerated points of concern are not specific to Jaguar: most other currently available *ab initio* software programs would benefit from laying more emphasis on them.

Jaguar use in Materials Science

The PS method and parallel implementation give Jaguar advantages in efficiency, that when combined with modern computational resources, pushes its realm of application from the treatment of small molecules to the study of increasingly larger and realistic systems with chemically reliable accuracy. Unlike classical force field based-methods, Jaguar can be used to study complex systems without empirical parameters containing elements from across the periodic table. This enables treatment of systems comprised of alkali elements, main-group elements, and transition metals on the same footing, allowing meaningful comparisons and analyses to be made. Jaguar can be an invaluable tool for studying the chemical reactions and properties

that are implicated in the assembly, operation or failure of materials, or for the discovery and optimization of new materials solutions. In this section, we provide examples of Jaguar's use in a range of materials science applications.

Catalysis

DFT has become a standard level of theory for analyzing the reaction mechanisms and energy profiles for elementary reaction steps in catalytic processes, and for elucidating the intrinsic chemical properties of the active sites that give rise to desired characteristics. The geometric effects, orbital, and electrostatic interactions that provide the basis for catalyst stability, selectivity, and activity can be analyzed. These details are difficult or impossible to gain by experiment alone. Calculations using Jaguar provide the insight needed to enable the rational design of improved catalysts.

Jaguar has been used extensively to study a wide range of homogeneous catalytic systems. Examples include Ru-alkylidene catalysts for enyne metathesis,[15,16,214] Ir(III) CO_2 hydrogenation catalysts,[217] CH bond activation by Pt, Ir, and Os complexes,[217–222] dioxirane-catalyzed asymmetric epoxidations,[223] Pd-monophosphine catalysts for Suzuki–Miyaura coupling,[224] Au(I) catalysts,[225–227] oxy-functionalization of Re—C bonds,[228,229] and Pd oxidation catalysts.[230]

Olefins that contain deactivating groups, such as esters or other electron-withdrawing moieties, are difficult to metathesize, but can be treated with ruthenium-alkylidene catalysts coordinated to N-heterocyclic carbenes. Fomine et al.[214] carried out hybrid DFT calculations to analyze the reaction pathways and chelation effects of carbonyl groups on the metathesis activity of cis-1,4-diacetoxy-2-butene using the second generation Ru Grubbs catalyst with the ligand 1,3-dimesityl-4,5-dihydroimidazol-2-ylidene (SIMes). The driving force in the metathesis of cis-1,4-diacetoxy-2-butene is predicted to be the formation of an internal Ru—O bond in the metallocarbene complex, which increases the activation free energy for this system. Comparison with the dimethyl maleate reaction highlights the role substrate conjugation can play in determining the relative metathesis reactivity of carbonyl containing olefins.

Organogold intermediates have a unique reactivity that has been leveraged to develop new carbon-carbon bonding reactions. There has been disagreement on how to best characterize the catalytic intermediate Au—CR_2^+, as either carbocation or carbene in nature. To provide a clear and consistent framework to rationalize differences in Au(I) catalytic pathways, Toste and coworkers carried out a systematic computational study[225] characterizing key intermediates relevant to Au catalysis. They found that the nature of the bonding in the Au—CR_2^+ intermediate included contributions of both σ- and π-bonding, varying from an Au-stabilized carbene to Au-coordinated carbocation, dictated by the carbene substituents and ancillary ligand. These findings are consistent with experiment and provide a basis for rational optimization and design of gold-catalyzed reactions.

Development of organometallic complexes that can be used for the direct selective activation of C—H bonds is under investigation to enable valuable transformations of this historically unreactive group. Goddard and co-workers[222] reported the first example of stoichiometric, catalytic, intermolecular CH activation by a discrete Os catalyst. They discovered that the complex $(\kappa^2$-acac-O,O)$_2$Os^{IV}(Ph)Cl reacts with C_6D_6 to form $(\kappa^2$-acac-O,O)$_2$Os^{IV}(Ph-d_5)Cl, and catalyzes the H/D exchange reaction between toluene-d_6 and benzene. First-principles simulations using hybrid DFT were used to gain insight into the mechanism of benzene CH activation and catalytic H/D exchange. Calculations indicate that the active species is a trace Os^{III} intermediate having a reaction pathway with energy barriers consistent with experimental data. Experiments adding reductants confirmed that the active species is an oxidation state lower than Os^{IV}, supporting the predictions made using Jaguar.

In addition to computing the energetics for reaction and establishing catalytic mechanisms, simulations with Jaguar can be used to construct QSAR models which represent direct predictive relationships between intrinsic catalyst properties and observed catalyst characteristics and performance.[231,232] Tabares-Mendoza and Guadarrama[231] carried out a study on pincer metal catalysts for the C—C coupling Heck reaction. The initial step in the mechanism for these transformations is an oxidative addition in which the metallic complex works as a nucleophile, so properties that reflect nucleophilicity should also reflect catalytic efficiency. Calculations of the atomic charge on the metal center (using NBO analysis), the HOMO energy, and the chemical hardness [$\eta = \frac{1}{2}(E_{LUMO} - E_{HOMO})$] were carried out. The pincer catalyst's chemical hardness and experimental turn-over number were found to be correlated. The regression model based on hardness was used to predict the activity for pincer Heck catalysts, providing a useful tool for evaluating new catalysts under fixed conditions.

Heterogeneous catalysts have also been studied using Jaguar. By using a cluster model to represent the local reaction site at the catalyst surface, the atomistic details, energetics, and mechanisms can be evaluated for heterogeneous catalytic processes using identical methods as used for homogeneous catalysts. In a cluster model, a limited region of the surface is cleaved from the extended surface and appropriate boundary conditions are enforced to minimize or eliminate nonphysical effects caused from neglecting the rest of the solid. An example of this approach is work by Lund and coworkers,[233] who investigated the active site structure for the high-temperature water-gas shift reaction on ferrochrome catalysts. Experimental kinetics does not provide specific details of the nature of the active surfaces and intermediates in the primary reaction steps in a catalytic reaction. A two-step microkinetic model was used to determine the experimental value for the O surface adsorption energy on the active oxide surface. Comparison of the DFT-calculated O adatom binding energies on Fe_3O_4 cluster models representative of the {100}, {110}, and {111} surfaces, suggests that the active catalyst is dominated by the {111} surface. The {111} surface has a binding energy of −668 kJ/mole, which is

comparable to the experimental value of −611 kJ/mole. The O-covered {111} Fe₃O₄ cluster model is illustrated in Figure 17.

The use of zeolites for catalytic transformations has been extensively developed for applications in the petrochemical industry. As for studies of surface active sites, truncated cluster models can also be constructed to represent the active site and inner phase of zeolite frameworks for quantum chemical analysis using Jaguar. Calculations have been carried out to investigate the oxidation of benzene and methane;[234,235] adsorption, dehydration, and cracking of branched alkanes;[236,237] and N_2O decomposition[238] on the zeolite microporous catalyst ZSM-5.

For example, Milas and Nascimento used a "double-ring" 20T cluster to represent the Brønsted acid site and cavity of
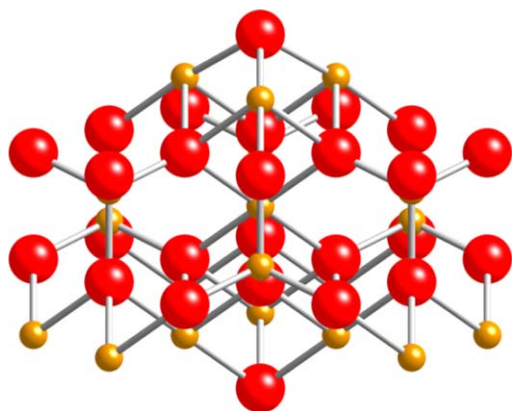

Figure 18. Transition state structure for isobutane dehydrogenation in the "double-ring" 20T cluster model of the HZSM-5 zeolite. (Reproduced from Ref. [237] with permission from Elsevier).

Figure 17. Truncated Fe_3O_4 cluster used to model the oxygen-covered active site for water-gas shift on {111} surface. Oxygen atoms are red, iron atoms are orange. (Reproduced from Ref. [233] with permission from Elsevier).

the HZSM-5 zeolite framework to determine the reaction energetics for the dehydrogenation and cracking reactions of isobutane.[237] A hybrid DFT functional in Jaguar was used to locate and characterize the reactants and transition structures for zeolite-catalyzed alkane reactions. The isobutane dehydrogenation transition structure in HZSM-5 is shown in Figure 18. The dehydrogenation and cracking reactions were found to be competitive, with activation energies of 46.3 and 47.4 kcal/mole, respectively. The small difference in barrier heights is consistent with experimental measurements showing that the dehydrogenation and cracking reactions have similar reaction rates.

Energy storage and generation

The realization of new batteries and super capacitors that exhibit the desired performance characteristics will require the discovery and optimization of new component materials including electrodes, solvents, and functional additives. First-principles simulations using Jaguar can be used to analyze the chemical mechanisms and controlling energetics for the operation and failure modes for candidate energy storage materials.

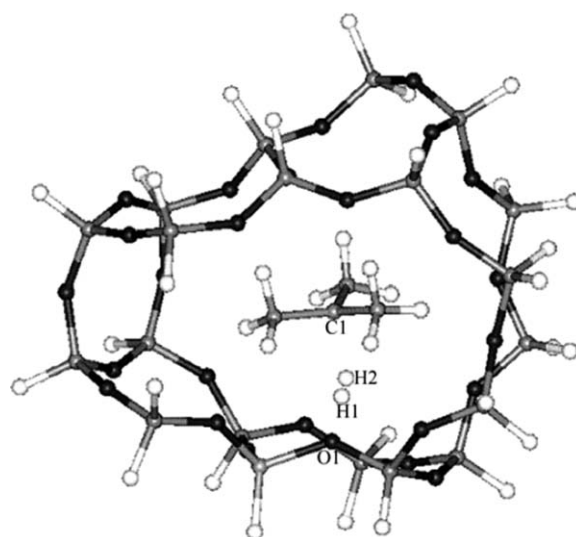Song and coworkers[239] synthesized and characterized nitrogen-doped carbon nanotubes (CNTs) for use as anodes in

Li-ion batteries. They found that the first discharge capacity of CN_x electrodes increased with nitrogen content. At higher nitrogen incorporation levels, an increase in irreversible capacity was observed. DFT calculations carried out by Song and coworkers provided insight into the atomic-scale features giving rise to the loss in discharge capacity. Simulations showed that the irreversible capacity was attributable to pyridinic nitrogen incorporation sites which have a Li ion binding energy 4.5 times higher than graphitic nitrogen sites.

West et al.[240] were able to demonstrate reversible intercalation of fluoride-anion receptor complexes in graphite cathodes, which enables batteries based on dual (de)intercalation of Li^+ at the anode and F^- at the cathode. In these cells, the LiF salt is dissolved in a nonaqeous solvent with the aid of the anion receptor additives, tris(pentaflurophenyl) borane and tris(hexafluoroisopropyl) borate. The characteristics of the anion receptor determined the extent of graphite fluorination, with optimal receptors having favorable anion binding and release properties in solvent. To facilitate the discovery of optimal anion receptors, an *ab initio*-based scheme was developed using Jaguar to evaluate F^- binding and release energetics using implicit solvation simulations that employ a PB continuum model.

The realization of Li-air batteries would enable next-generation electric vehicles with dramatically increased range due to their superior energy densities. A core challenge hampering progress is the development of organic electrolytes that are stable in the chemical environment of an operating Li-air battery. Researchers at Liox Power have used hybrid DFT along with the PB solvation model in Jaguar to determine the mechanism and energetics for solvent decomposition pathways to establish a basis for the selection and design of stable electrolyte components.[241–243] A comparative analysis of the energy barriers and overall reaction energies for nucleophilic attack by superoxide and for autooxidation by molecular

oxygen over a wide range of solvents was carried out. This work determined which chemical functionalities are most suitable for use as solvents and revealed strategies for further enhancing solvent stability. Organic carbonates, sulfonates, aliphatic carboxylic esters, lactones, phosphinates, phosphonates, phosphates, and sulfones were found to be highly susceptible to nucleophilic attack by O_2^- (energy profile for propylene carbonate shown in Fig. 19), whereas ethers and N,N-dialkylamides were identified as the most stable solvents against superoxide attack. Chemical strategies to increase the resistance to autooxidation include fluorination, using sterics to block the reaction site, and engineering heteroatom lone pair/aromatic conjugation.

Nanoelectronics

With each generation of microelectronic device, the lateral dimensions of the active elements and the thickness of the active layers decrease. The ultimate limit of this aggressive miniaturization trend is the development of molecular electronic elements. In the development of nanoelectronic devices, quantum-based simulation can provide insight into nanomaterial synthetic mechanisms, the features of the electronic struc-

Figure 19. Hybrid DFT calculated reaction profile ($T=298.15$ K) for attack of O_2^- on propylene carbonate. (Reproduced from Ref. [241] with permission from American Chemical Society).

ture tied to device properties, and the role structure and chemistry play in tuning key characteristics. Jaguar has been used to investigate the mechanism of carbon nanotube growth,[244] the conformation dependence of molecular conduction,[245] the electronic structure of molecular rectifiers,[246] switching in mechanically interlocked molecules,[247] and interference effects in conduction through arene molecular wires.[248]

Since their discovery in 1991, CNTs have received worldwide attention because of their unique electronic properties. Nanotube conductivity can vary from semiconducting to metallic depending on structural details. Key to realizing the use of CNT in applications is first realizing synthetic control through the selection and development of efficient catalysts. Goddard and coworkers[244] carried out calculations examining the energetics of the steps in CNT formation using a two-

stage mechanism; (1) CNT nucleation and (2) CNT growth and defect repair. Analysis of the metal dependence on the relative energetics for nucleation and growth correctly predicts monometallic and bimetallic catalytic efficiencies in good agreement with experiment. Through simulation, they identified a new bimetallic CNT catalyst, Ni/Mo that should outperform Ni/Co, a standard CNT catalyst.

Conduction on the molecular scale depends critically on the details of the underlying electronic structure, showing specific resonances with the molecular energy levels. One factor that clearly differentiates molecular conduction is the role of quantum interference, which can depend heavily on differences in molecular structure. Baer and coworkers[248] used a Green's function approach with Jaguar to compute the input DFT Fock matrix and obtain the electron transmission spectra for a series of polycyclic aromatic hydrocarbon wires. Interference effects were found to be most significant in naphthalene-based molecular wires, in comparison to wires containing benzene, anthracene, or tetracene, with clear loss of transmission resonance features.

Optoelectronics and photovoltaics

Organic semiconductors are under development to complement or displace inorganics in optoelectronic and photovoltaic applications. Advantages over inorganic materials include an enormous chemical design space, atomically precise structures using organic synthesis, and large-scale materials integration by thermal evaporation or solvent deposition. Molecular properties such as electronic energies, multipole moments, linear and higher-order polarizabilities, ionization and reduction potentials, and charge reorganization energies can be evaluated computationally to aid in the selection or design of organic optoelectronic materials.

Jaguar has been used to analyze a variety of organic semiconductors including derivatized oligothiophenes,[249] cyanated tetracenes,[250] and N-heteropentacenes.[251] An example is work by Wang and Friesner,[249] evaluating the effect of oligomer backbone length, side chains, end groups, and spacers on the electronic structure and redox potentials of thiophenes. Their analysis confirmed that polarons are the dominant charge carrier for thiophene oligomers. Doubly oxidized molecules favored the formation of two separate polarons. Intermediate volume sidechains lead to backbone twisting that forms a number of planar segments which facilitate the formation of polarons, suggesting sidechain structure as a parameter for tailored thiophene conductivity.

Organic light-emitting diodes (OLEDs) have become the most successful area of application for organic semiconductor materials. These devices normally consist of at least one hole-transport layer and one electron-transport layer forming an organic/organic heterojunction. Holes from the anode and electrons from the cathode travel through the transport layers until they form a singlet exciton that relaxes, giving rise to electroluminescence. The emissive materials in OLED devices are typically low-molecular-weight organometallic complexes, dispersed in a host material. The active layer materials and
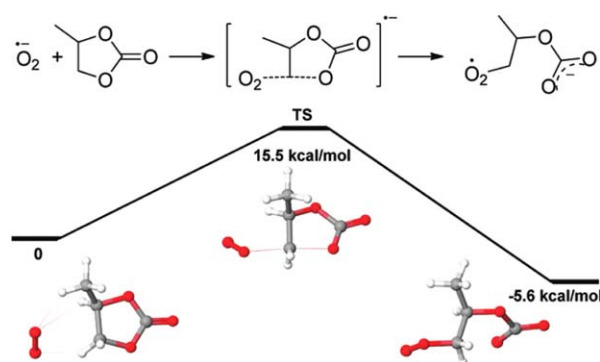
emitters are selected with close attention to their redox properties, triplet energies, and electronic structure. Molecular orbital calculations using Jaguar have been carried out for a variety of OLED materials including a highly emissive phosphine-$Cu_2(\mu$-$NAr_2)_2$ complex,[252] mixed light-emitting layer host materials,[253] and various phosphorescent cyclometalated Ir(III) complexes.[254–256]

Cyclometalated Ir(III) complexes have phosphorescent efficiencies approaching unity and have short radiative lifetimes, making them well-suited as emitter materials for OLEDs. Thompson and coworkers[256] carried out a study of the photophysical properties for a series of cyclometalated fac-Ir(III) complexes with emission energies ranging from the near-UV to green, complementing experimental measurements with theoretical calculations. They found that the quantum yields were temperature-dependent, indicating thermal deactivation to a nonradiative state. Calculations showed that the defect triplet state corresponds to a five-coordinate trigonal bipyramidal species formed through the rupture of an Ir—N bond, which is illustrated in Figure 20. Spin density analysis indicates that the nonradiative product state has triplet metal-center (^{3}MC) character with an increase in spin density on the Ir center, compared to the emissive state.

Dye-sensitized solar cells (DSSC) are under development to compete with traditional inorganic p-n junction photovoltaic devices. In a DSSC device, light is adsorbed by a thin film of high efficiency chromophores adsorbed on the surface of a nanostructured oxide substrate (mostly TiO_2). The photoexcited electrons from the dye layer transfer into the oxide conduction band and then to the collection anode. A redox electrolyte completes the system, which regenerates the dye
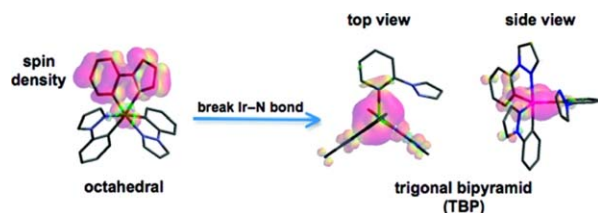

Figure 20. Jaguar-predicted structures and spin density surfaces for the emissive and nonradiative triplet forms of fac-Ir(ppz)$_3$. (Reproduced from Ref. [256] with permission from American Chemical Society).

and is itself reduced from electron transfer at the cathode. Jaguar can be used to analyze the DSSC constituent materials and properties affecting carrier transport. Examples include the calculations of the surface dipole contributions of high-adsorption coefficient metal-free indoline dyes,[257] the orbital changes linked to increased efficiency in direct transfer electron-donor derivatized catechols,[258] and the structures and energetics of TiO_2 nanoparticle intermediates relevant to electron transport in DSSC devices.[72]

Recently, Zhang et al.[72] illustrated the effectiveness of Jaguar for reliably predicting the properties of complex DSSC components in their work analyzing TiO_2 nanoparticles. Using a realistically terminated TiO_2 nanocluster model and

employing the B3LYP-LOC method (Section Methods employing LOC), correcting for cluster size and solvation effects, enabled the calculation of properties related to electron transport and trapping in agreement with experiment to within 0.1–0.2 eV. The working TiO_2 cluster model consisted of a H_2O-passivated 5 × 5 × 5 rutile cluster, with a stoichiometry of $Ti_{61}H_{116}O_{180}$. Using a PB continuum H_2O solvation model, the conduction band and valence band potentials were predicted to be −0.88 and +2.89 eV (vs. the standard hydrogen electrode), which are comparable to experimental values of −0.7 and +2.3 eV. A Li cation was found to trap an excess electron in a state almost completely localized on the nearest Ti d-orbital (as shown in Fig. 13) for the case of surface-adsorbed Li$^+$. Electron trap energies were calculated to be 0.35 and 0.50 eV for surface and interstitial cation, respectively, which compares well with the experimental range of 0.3–0.5 eV. The activation barriers for elementary ambipolar diffusion pathways were predicted to be 0.12–0.25 eV, consistent with experimental barrier heights of 0.10–0.27 eV from temperature-dependent electron transport rates in DSSC devices. These results support the plausibility that ambipolar diffusion plays a role in electron-transport through TiO_2 in DSSCs.

In this section, we have briefly touched on representative examples highlighting the use of Jaguar for a diverse range of technologically important chemical systems. From catalysis to optoelectronics, Jaguar has been shown to provide critical insight into the structures and processes for these materials science applications. Jaguar's capabilities make it an efficient and robust tool for the routine treatment of realistic chemical models. An exciting possibility is the potential for the *ab initio* design or high-throughput computational screening for new materials with novel or enhanced properties. Using automated workflows, the transfer of the virtual screening paradigm from drug discovery to problems in materials science is a direction that is currently under active development.

Jaguar use in Molecular Biology

QSite is a QM/MM program which provides an interface between the two molecular regions handled by QM and MM. Mixed QM/MM calculations are essential tools in modeling large entities, particularly enzymes and other biologically relevant systems. This type of methodology is most useful when one is treating a reactive region that is restricted in size (typical current models for enzymes use 100–200 atoms in the reactive region) surrounded by a superstructure that provides electrostatic interactions and structural constraints, which can total thousands, or even tens of thousands of atoms. The reactive region is treated via QM methods (invariably DFT driven by Jaguar) and the remainder of the system is treated at the MM level (driven by another program managed by Schrödinger, Impact[259]).

A key component of a QM/MM model is the interface between the QM and MM regions. QSite has two types of interface models that it uses, under the control of the user. The first is a frozen orbital method which is specifically tuned to model an interface for protein backbone and side chain regions. This method has been extensively tested for such

applications, and can be expected to provide a high degree of reliability provided that the interface is sufficiently distant from the chemical reaction of interest. The second method is a more conventional link atom approach. This method is not quite as accurate for protein systems, but has the virtue of applicability to a wide range of chemistries without further modification. It is comparable in accuracy to other link atom methods in the literature.

QSite has been applied to a large number of enzymatic calculations, including many systems containing transition metals (see, e.g., Refs. [140,260–264]). These calculations have demonstrated considerable success in reproducing experimental barrier heights in biochemical reactions. In a few cases, such as the hydrogen atom abstraction reaction in cytochrome P450, significant discrepancies with experiment remain.[265] As discussed above in section LOC for *3d* transition metals (DBLOC), we now believe that this discrepancy is due to problems with the DFT functional used (B3LYP), which can be addressed effectively via our LOC methodology.

Jaguar use in Force Field Development

Though considerable progress has been made to improve the computational efficiency of quantum chemical software, systems with a characteristic size much greater than the order of hundreds of atoms, remain outside the scope of current *ab initio* models.

For example, the size of biological systems, consisting of proteins and membranes, routinely exceed 10,000 atoms in applications.[266,267] Alternative models, such as force fields, try to mediate these scales by using simple classical potential functions that aim for computational expediency while retaining a sufficient level of model accuracy. Applications of force fields are wide ranging where just a few examples include the structural characterization of proteins[268] and biological membranes,[269,270] the design of small molecule drugs,[271] and the simulation of condensed phase materials.[272,273]

The typical force field potential function uses harmonic oscillators to represent intramolecular vibrations, the Lennard–Jones potential to represent electronic repulsion, and atomic scale partial charges to approximate electrostatic interactions.[274] Each term depends on a set of parameters empirically fit against suitable reference data. For parametrical components sensitive to short-lived structural states (i.e., barriers of rotation) or relatively rare chemical moieties, suitable data are often not available from experimental sources. In this regard, *ab initio* models can play an integral role in the development of a force field. Requisite physical observables can be resolved via the *ab initio* model and subsequently used in force field parameter fitting. In this section, we describe how Jaguar and its associated quantum chemical methods have been used in just this manner to develop and refine force field models.

Refining protein force fields

The original formulation of some of the most popular protein force fields, including CHARMM22,[275] OPLSAA,[276] and

Amber,[277] came at a time when limitations in hardware and quantum chemical software prevented a more thorough evaluation of relevant quantum observables. Perhaps most notably, in lieu of suitable reference data, determining parameters associated with some of the protein's torsional coordinates were relegated to *ad hoc* assumptions. Most often represented by a truncated Fourier series, the torsional component of the potential function plays a significant role in modulating conformational propensities. These approximations have been cited as the cause of observed structural artifacts, including the overstabilization of alpha helices in Amber[278] and the overstabilization of π-helices in CHARMM.[279] Advances, such as the development of the PS LMP2 method,[9] have since made exhaustive sampling of the torsional coordinate space of protein backbone and side chain analogs fairly routine. In particular, recent variations of three of the most commonly used force fields have all relied heavily on refinements against LMP2 evaluated conformational energies (see also Ref. [280] for comparison of different force field results against LMP2 results).

Kaminski et al.[281] used extensive sampling of analogs to all 20 amino acid side chains to refine the torsional component of the OPLS-AA protein model. The revised model has also been found to give more accurate predictions of preferred side chain conformations.[282] Employing a similar strategy, Hornak et al.[278] reparameterized the backbone torsional component of the Amber ff94 model against QM conformational energies also resolved at the LMP2/cc-pVTZ(-f) level for alanine and glycine tetrapeptide. Relative to X-ray and NMR benchmarks, the revised model more consistently represents secondary structure propensities observed experimentally.

To address overstabilization of π-helices in the CHARMM force field, Feig and coworkers[283] extensively sampled the 2-D quantum energy surface of alanine and glycine dipeptide at the LMP2/cc-pVQZ(-g) level. By parameterizing backbone torsions, including dihedral cross terms, the authors were able to reproduce to a high level of precision the quantum energy surface, leading to improved agreement with the conformational propensities observed experimentally.

Most commonly used force fields (including those cited above) do not explicitly account for electronic polarization in their potential function. The underlying assumption in such models is that the electron density of the composite molecules experiences relatively small fluctuations. The assumption belies the electrostatic diversity found in biomolecular systems, ranging from hydrophobic, like the interior of proteins, to polar, like water solvent. To address this concern, recent efforts have begun to explicitly incorporate polarizability. Based almost exclusively on *ab initio*-generated data, Kaminski et al.[284] developed a polarizable force field for proteins. Torsional parameters were fit using analogous data to the refinement of the OPLS-AA protein force field outlined above. Parameters associated with the electrostatic degrees of freedom (partial charges and point dipole polarizabilities) were fit to match ESP data resolved using density-functional B3LYP/cc-pVTZ(-f). And finally, parameters associated with the Lennard–Jones component of the force field were fit to reproduce dimer interaction energies in the CBS approximation using

data generated at the LMP2/cc-pVTZ(-g) and LMP2/cc-pVQZ(-g) levels.

Small molecule force fields

Unlike proteins, where chemical diversity is limited to a small number of naturally occurring amino acids, the primary challenge in developing accurate and reliable force fields for small molecules is parametrically covering the vast space of chemical functionalities that compose them. Components of the model that are particularly sensitive to their chemical environment (torsions and partial charges) pose the most significant challenge.

The recently developed OPLS2.0 force field[13] aims to address this issue by leveraging a growing database of quantum mechanically derived observables that probe small molecule torsional and electrostatic coordinates. Similar in nature to the database used in the development of the OPLS2005 force field,[258] the QM reference data consist of sampled torsional coordinates energetically resolved at the LMP2/cc-pVTZ(-f) level, as well as molecular ESP determined at the HF/6-31G* level. Although the OPLS2005 database consisted of QM observables developed for approximately 600 compounds, the data used in the parameterization of OPLS2.0 are based on more than 10,000 compounds, covering a much larger fraction of small molecule chemical space. An example of a torsional energetic profile taken from the OPLS2.0 database is shown in Figure 21. The figure compares a QM model to the OPLS2.0 force field and provides a contrast against two force fields where this torsional coordinate is missing from their parameterization (OPLS2005 and MMFF). An aggregate accuracy comparison, relative to the QM model, is shown in Table 5, and again compares against the MMFF and OPLS2005 force fields.

Water solvation plays an integral role in the ligand/protein binding process and, as such, measures of the free energy required to solvate molecules in liquid water constitute an important benchmark for the quality of a small molecule force field. In a recent study, Shivakumar et al.[285] compared the relative performance of several force fields in predicting absolute water solvation free energies over a set of 239 diverse small molecules, finding OPLS2.0 performed demonstrably better relative to its competitors.
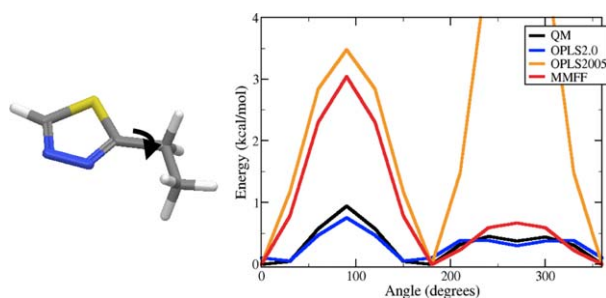

Figure 21. The torsional energetic profiles obtained with different methods for a sample molecule. The atoms in the molecule are marked in different colors: carbon is grey, hydrogen is white, nitrogen is blue, and sulfur is yellow. QM stands for quantum mechanical LMP2/cc-pVTZ(-f) energies obtained with B3LYP/6-31G*-optimized structures.

Table 5. Rotamer energy RMSD (with respect to LMP2/cc-pVTZ(-f) energies obtained with B3LYP/6-31G*-optimized structures) computed over a 12,000 compound database.

Model	Rotamer RMS (kcal/mole)
OPLS2.0	0.5
OPLS2005	2.6
MMFF	2.5

More directly assessing the capacity of a force field to model protein-ligand binding, Wang et al.[286] applied several variants of the OPLS force field in a computational study measuring relative protein-ligand binding affinities. The calculations used a molecular simulation sampling algorithm[287] that sensitively probes advances made to the electrostatic and torsional components of the force field, finding significant improvements in accuracy when using the OPLS2.0 force field model.[286]

Reactive force fields

Classical force fields such as those cited above are predominantly used for describing chemical species around their equilibrium geometries. For a wealth of applications, however, force fields capable of treating nonequilibrium species and reactions are needed. Jaguar has been a core component for parameterizing one of the most popular such reactive force fields, ReaxFF, created by van Duin, Goddard, and coworkers (see, e.g., Refs. [288–290]). The force field contains parameters for numerous elements and bond types in a wide range of bond lengths, and found hundreds of published applications in diverse areas of materials science often using molecular dynamics simulations.[291,292]

Conclusions

Ab initio quantum chemistry programs have become an important tool in chemical research. They have taken off from being the theorist's prerogative, and penetrated into various chemical disciplines, such as organic, inorganic, and medicinal chemistry, biochemistry, materials science, and chemical education. Despite the substantial advances in making *ab initio* programs more available, more user-friendly, faster, and more robust over the last decade, much work remains to be done before these theoretical tools can truly aim at complementing experimental research. Among the principal objections to using *ab initio* modeling more often are insufficiently high speed for realistically large systems and the difficulty in modeling numerous factors in often complex experimental conditions and chemical environments. Another obstacle to using *ab initio* programs is a large number of theoretical methods and settings available in them. To operate these programs appropriately and avoid caveats and unpleasant surprises, the user is often required to possess a specialized education or considerable experience. The following paragraph briefly summarizes how Jaguar developers are addressing these problems.

First of all, we are concentrating on developing *ab initio* methods that permit work with larger and larger systems as the hardware speed increases. This is in contrast to such

methods as coupled cluster, which are bound to be useful only for small systems for a long time, owing to their prohibitive scaling being the sixth, seventh, and even larger powers of the number of atoms. The extreme complexity of these codes as well as their extreme demand for computer resources do not make them easily parallelizable or adoptable for computations on graphical processing units. Improving parallel performance of the code is another important direction for keeping up with advances in hardware development. A greatly improved performance of Jaguar, in main part due to a more extensive parallelization, was demonstrated in section Jaguar parallelization. Several other *ab initio* quantum chemistry programs have recently improved their performance. Regrettably, it is quite impossible for us to report on performance of Jaguar relative to that of other *ab initio* programs because license agreements of most such programs do not permit disclosing direct comparisons of timings. For this reason, we have provided the Cartesian coordinates and descriptions of the systems used for benchmarking in Figures 10–12 and Tables 2 and 3, so that the reader can use this information for his or her own comparison.

To bring the simulations closer to the experimental conditions, Jaguar provides methods that go beyond treating isolated molecules in vacuum. Our PBF method permits simulations of chemical phenomena in solution, and can even be accompanied with conformational search (as in case of the pKa prediction and VCD spectra workflows). We also believe in the importance of improving the user's interaction with the program, making it accessible to non-specialists, and automating the key tasks. A central constituent in this strategy is the sophisticated and powerful GUI Maestro, which supports all operations available in Jaguar and enables Jaguar to connect seamlessly with numerous other software programs developed by Schrödinger. Several automated workflows with multiple customization options, for example, the computation of Fukui functions, enthalpy of formation prediction, and counterpoise calculations, simplify computational modeling even further. We recognize that not every Jaguar user is an expert in quantum chemistry and that even experts might need the possibility to interact with the developers and require detailed information about the program operation. For this reason, Jaguar customer support is available. Besides, Jaguar is accompanied with a detailed yet legible manual which contains multiple explanations and recommendations about the methods and settings used in the program.

Finally, we would like to outline directions for the future development of Jaguar. The last paragraph of section Jaguar use in Drug Discovery already mentioned several desirable improvements of the program for expanding its use in drug discovery applications. A summary of those development vectors is given below, followed by ideas for intended enhancements in other areas of research.

Jaguar must focus on delivering a complete toolkit of fast, reliable, and validated methods to provide a solution to computational chemical problems. Note that this goal is quite different from the direction adopted by several other popular *ab initio* quantum mechanical programs, for the toolkit does not mean the same thing as a garage full of instruments. Jaguar has been shown to be a powerful tool for studying discrete chemical

models that can provide critical insight into the properties and processes at the heart of numerous materials science applications. In the future, Jaguar's capabilities will be extended to further enable the analysis and optimization of materials systems. More attention should be paid to methods required for simulating transition metal chemistry, as transition metals are a frequent motif in materials applications. We have already started addressing some of the common problems in 3*d*-metal-containing systems with the DBLOC corrections described in section LOC for *3d* transition metals (DBLOC). More variety in basis sets for transition metal elements as well as improving SCF convergence algorithms for metal-containing systems are also of interest to us. Next, in spite of the limited applicability of coupled cluster methods to realistic chemical problems referred to above, such methods are nevertheless useful and even irreplaceable in certain focused validations and parameterizations, and might find their place in the toolkit.

Although the objectives of higher accuracy and performance are obvious and were built in Jaguar design from its inception, some other directions for improvement mentioned in this conclusion are more subtle. In a sense, they are "emerging properties" which would have gone unnoticeable had not the critical mass of an already achieved progress accumulated. Computational practices and phenomena which now constitute an everyday reality (high-throughput screening of drug and materials candidates, large volumes of computational data, databases of chemical information, routine modeling of protein-ligand interactions, calculations on complex systems containing transition metal elements, a greater emphasis on graphical visualization, etc.) would have been of the least concern a couple of decades ago when most high-profile *ab initio* software programs originated. These areas of activity are dictating new patterns of Jaguar development and will have to be taken seriously in the next few years.

[1] R. A. Friesner, *Chem. Phys. Lett.* **1985**, *116*, 39.

[2] R. A. Friesner, *J. Chem. Phys.* **1986**, *85*, 1462.

[3] R. A. Friesner, *J. Chem. Phys.* **1987**, *86*, 3522.

[4] R. A. Friesner, *J. Phys. Chem.* **1988**, *92*, 3091.

[5] M. N. Ringnalda, Y. Won, R. A. Friesner, *J. Chem. Phys.* **1990**, *92*, 1163.

[6] M. N. Ringnalda, M. Belhadj, R. A. Friesner, *J. Chem. Phys.* **1990**, *93*, 3397.

[7] Y. Won, J.-G. Lee, M. N. Ringnalda, R. A. Friesner, *J. Chem. Phys.* **1991**, *94*, 8152.

[8] R. A. Friesner, *Ann. Rev. Phys. Chem.* **1991**, *42*, 341.

[9] R. B. Murphy, M. D. Beachy, R. A. Friesner, M. N. Ringnalda, *J. Chem. Phys.* **1995**, *103*, 1481.

[10] R. G. Parr, W. Yang, Density-Functional Theory of Atoms and Molecules; Oxford: New York, **1989**.

[11] P. Pulay, S. Saebø, *Theor. Chim. Acta* **1986**, *69*, 357.

[12] S. Saebø, P. Pulay, *Ann. Rev. Phys. Chem.* **1993**, *44*, 213.

[13] OPLS2.0, Schrödinger Inc.; Portland, OR, **2012**.

[14] C. M. Cortis, R. A. Friesner, *J. Comput. Chem.* **1997**, *18*, 1570.

[15] C. M. Cortis, R. A. Friesner, *J. Comput. Chem.* **1997**, *18*, 1591.

[16] C. M. Cortis, J.-M. Langlois, M. D. Beachy, R. A. Friesner, *J. Chem. Phys.* **1996**, *105*, 5472.

[17] D. J. Tannor, B. Marten, R. Murphy, R. A. Friesner, D. Sitkoff, A. Nicholls, M. Ringnalda, W. A. Goddard, III, B. Honig, *J. Am. Chem. Soc.* **1994**, *116*, 11875.

[18] B. Marten, K. Kim, C. Cortis, R. A. Friesner, R. B. Murphy, M. N. Ringnalda, D. Sitkoff, B. Honig, *J. Phys. Chem.* **1996**, *100*, 11775.

[19] J. J. Klicić, R. A. Friesner, S.-Y. Liu, W. C. Guida, *J. Phys. Chem. A* **2002**, *106*, 1327.

[20] V. S. Bryantsev, M. S. Diallo, W. A. Goddard, III, *J. Phys. Chem. A* **2007**, *111*, 4422.

[21] L. Zhang, R. A. Friesner, *J. Phys. Chem.* **1995**, *99*, 16479.

[22] M.-H. Baik, R. A. Friesner, *J. Phys. Chem. A* **2002**, *106*, 7407.

[23] Z. Yu, M. P. Jacobson, J. Josovitz, C. S. Rapp, R. A. Friesner, *J. Phys. Chem. B* **2004**, *108*, 6643.

[24] S. F. Sousa, P. A. Fernandes, M. J. Ramos, *J. Phys. Chem. A* **2007**, *111*, 10439.

[25] L. A. Burns, A. Vasquez-Mayagoitia, B. G. Sumpter, C. D. Sherrill, *J. Chem. Phys.* **2011**, *134*, 084107.

[26] L. Goerigk, S. Grimme, *Phys. Chem. Chem. Phys.* **2011**, *13*, 6670.

[27] Y. Zhao, D. G. Truhlar, *J. Phys. Chem. A* **2006**, *110*, 13126.

[28] Y. Zhao, D. G. Truhlar, *Theor. Chem. Acc.* **2008**, *120*, 215.

[29] R. A. Friesner, E. H. Knoll, Y. Cao, *J. Chem. Phys.* **2006**, *125*, 124107.

[30] D. A. Goldfeld, A. D. Bochevarov, R. A. Friesner, *J. Chem. Phys.* **2008**, *129*, 214105.

[31] M. L. Hall, D. A. Goldfeld, A. D. Bochevarov, R. A. Friesner, *J. Chem. Theory Comput.* **2009**, *5*, 2996.

[32] T. F. Hughes, R. A. Friesner, *J. Chem. Theory Comput.* **2012**, *8*, 442.

[33] T. F. Hughes, R. A. Friesner, *J. Chem. Theory Comput.* **2011**, *7*, 19.

[34] T. F. Hughes, J. N. Harvey, R. A. Friesner, *Phys. Chem. Chem. Phys.* **2012**, *14*, 7724.

[35] V. Gogonea, D. Surez, A. van der Vaart, K. M. Merz, Jr., *Curr. Opin. Struct. Biol.* **2001**, *11*, 217.

[36] M. H. M. Olsson, G. Y. Hong, A. Warshel, *J. Am. Chem. Soc.* **2003**, *125*, 5025.

[37] D. M. Philipp, R. A. Friesner, *J. Comput. Chem.* **1999**, *20*, 1468.

[38] R. B. Murphy, D. M. Philipp, R. A. Friesner, *J. Comput. Chem.* **2000**, *21*, 1442.

[39] S. Grimme, J. Antony, S. Ehrlich, H. Krieg, *J. Chem. Phys.* **2010**, *132*, 154104.

[40] S. Schneebeli, A. D. Bochevarov, R. A. Friesner, *J. Chem. Theory Comput.* **2011**, *7*, 658.

[41] H. Kim, J.-M. Choi, W. A. Goddard, III, *J. Phys. Chem. Lett.* **2012**, *3*, 360.

[42] E. Chamorro, P. Pérez, *J. Chem. Phys.* **2005**, *123*, 114107.

[43] R. R. Contreras, P. Fuentealba, M. Galván, P. Pérez, *Chem. Phys. Lett.* **1999**, *304*, 405.

[44] Available at: http://www.openmp.org. Access date: June 15, 2013.

[45] Available at: http://www.mcs.anl.gov/research/projects/mpi/. Access date: June 15, 2013.

[46] A. Szabo, N. S. Ostlund, Modern Quantum Chemistry; Dover, New York, **1989**.

[47] T. Helgaker, P. Jorgensen, J. Olsen, Molecular Electronic-Structure Theory; Wiley, Chichester, **2000**.

[48] C. Edmiston, K. Ruedenberg, *Rev. Mod. Phys.* **1963**, *35*, 457.

[49] J. M. Foster, S. F. Boys, *Rev. Mod. Phys.* **1960**, *32*, 300.

[50] A. Karton, A. Tarnopolsky, J. F. Lamere, G. C. Schatz, J. M. L. Martin, *J. Phys. Chem. A* **2008**, *112*, 12868.

[51] R. Peverati, D. G. Truhlar, *J. Phys. Chem. Lett.* **2011**, *2*, 2810.

[52] S. Grimme, *J. Comput. Chem.* **2006**, *27*, 1787.

[53] A. K. Rappé, C. J. Casewit, K. S. Colwell, W. A. Goddard III, W. M. Skiff, *J. Am. Chem. Soc.* **1992**, *114*, 10024.

[54] R. A. Friesner, R. B. Murphy, M. D. Beachy, M. N. Ringnalda, W. T. Pollard, B. D. Dunietz, Y. Cao, *J. Phys. Chem. A* **1999**, *103*, 1913.

[55] P. Hemberger, A. Bodi, C. Schon, M. Steinbauer, K. H. Fischer, C. Kaiser, I. Fischer, *Phys. Chem. Chem. Phys.* **2012**, *14*, 11920.

[56] J. C. Flick, D. Kosenkov, E. G. Hohenstein, C. D. Sherrill, L. V. Slipchenko, *J. Chem. Theory Comput.* **2012**, *8*, 2835.

[57] R. Kang, H. Chen, S. Shaik, J. Yao, *J. Chem. Theory Comput.* **2011**, *7*, 4002.

[58] H. S. Biswal, E. Gloaguen, M. Mons, S. Bhattacharyya, P. R. Shirhatti, S. Wategaonkar, *J. Phys. Chem. A* **2011**, *115*, 9485.

[59] L. Goreigk, H. Kruse, S. Grimme, *Chem. Phys. Chem.* **2011**, *11*, 3421.

[60] M. Feyereisen, G. Fitzgerald, A. Komornicki, *Chem. Phys. Lett.* **1993**, *208*, 359.

[61] F. Weigend, M. Haser, H. Patzelt, R. Ahlrichs, *Chem. Phys. Lett.* **1998**, *294*, 143.

[62] D. Cremer, *Mol. Phys.* **2001**, *99*, 1899.

[63] C. J. Cramer, B. A. Smith, *J. Phys. Chem.* **1996**, *100*, 9664.

[64] L. A. Curtiss, K. Raghavachari, P. C. Redfern, J. A. Pople, *J. Chem. Phys.* **1997**, *106*, 1063.

[65] F. L. Liu, E. Proynov, J.-G. Yu, T. R. Furlani, J. Kong, *J. Chem. Phys.* **2012**, *137*, 114104.

[66] E. R. Johnson, A. D. Becke, *Can. J. Chem.* **2009**, *87*, 1369.

[67] O. A. Vydrov, G. E. Scuseria, *J. Chem. Phys.* **2004**, *121*, 8187.

[68] M. L. Hall, J. Zhang, A. D. Bochevarov, R. A. Friesner, *J. Chem. Theory Comput.* **2010**, *6*, 3647.

[69] N. B. Balabanov, K. A. Peterson, *J. Chem. Phys.* **2006**, *125*, 074110.

[70] D. Rinaldo, L. Tian, J. N. Harvey, R. A. Friesner, *J. Chem. Phys.* **2008**, *129*, 164108.

[71] P. E. M. Siegbahn, *J. Biol. Inorg. Chem.* **2006**, *11*, 695.

[72] J. Zhang, T. F. Hughes, M. Steigerwald, L. Brus, R. A. Friesner, *J. Am. Chem. Soc.* **2012**, *134*, 12028.

[73] W. J. Hehre, R. F. Stewart, J. A. Pople, *J. Chem. Phys.* **1969**, *51*, 2657.

[74] W. J. Hehre, R. Ditchfield, R. F. Stewart, J. A. Pople, *J. Chem. Phys.* **1970**, *52*, 2769.

[75] W. J. Pietro, B. A. Levi, W. J. Hehre, R. F. Stewart, *Inorg. Chem.* **1980**, *19*, 2225.

[76] W. J. Pietro, E. S. Blurock, R. F. Hout, Jr., W. J. Hehre, D. J. DeFrees, R. F. Stewart, *Inorg. Chem.* **1980**, *20*, 3650.

[77] J. B. Collins, P. v. R. Schleyer, J. S. Binkley, J. A. Pople, *J. Chem. Phys.* **1976**, *64*, 5142.

[78] J. S. Binkley, J. A. Pople, W. J. Hehre, *J. Am. Chem. Soc.* **1980**, *102*, 939.

[79] M. S. Gordon, J. S. Binkley, J. A. Pople, W. J. Pietro, W. J. Hehre, *J. Am. Chem. Soc.* **1982**, *104*, 2797.

[80] W. J. Pietro, M. M. Francl, W. J. Hehre, D. J. DeFrees, J. A. Pople, J. S. Binkley, *J. Am. Chem. Soc.* **1982**, *104*, 5039.

[81] R. Ditchfield, W. J. Hehre, J. A. Pople, *J. Chem. Phys.* **1971**, *54*, 724.

[82] W. J. Hehre, J. A. Pople, *J. Chem. Phys.* **1972**, *56*, 4233.

[83] J. S. Binkley, J. A. Pople, *J. Chem. Phys.* **1977**, *66*, 879.

[84] P. C. Hariharan, J. A. Pople, *Theor. Chim. Acta* **1973**, *28*, 213.

[85] W. J. Hehre, R. Ditchfield, J. A. Pople, *J. Chem. Phys.* **1972**, *56*, 2257.

[86] M. M. Francl, W. J. Pietro, W. J. Hehre, J. S. Binkley, M. S. Gordon, D. J. DeFrees, J. A. Pople, *J. Chem. Phys.* **1982**, *77*, 3654.

[87] V. A. Rassolov, J. A. Pople, M. A. Ratner, T. L. Windus, *J. Chem. Phys.* **1998**, *109*, 1223.

[88] T. Clark, J. Chandrasekhar, G. W. Spitznagel, P. v. R. Schleyer, *J. Comput. Chem.* **1983**, *4*, 294.

[89] M. J. Frisch, J. A. Pople, J. S. Binkley, *J. Chem. Phys.* **1984**, *80*, 3265.

[90] R. Krishnan, J. S. Binkley, R. Seeger, J. A. Pople, *J. Chem. Phys.* **1980**, *72*, 650.

[91] A. D. McLean, G. S. Chandler, *J. Chem. Phys.* **1980**, *72*, 5639.

[92] T. H. Dunning, Jr., *J. Chem. Phys.* **1989**, *90*, 1007.

[93] R. A. Kendall, T. H. Dunning, Jr., R. J. Harrison, *J. Chem. Phys.* **1992**, *96*, 6796.

[94] D. E. Woon, T. H. Dunning, Jr., *J. Chem. Phys.* **1993**, *98*, 1358.

[95] D. E. Woon, T. H. Dunning, Jr., *J. Chem. Phys.* **1994**, *100*, 2975.

[96] R. E. Easton, D. J. Giesen, A. Welch, C. J. Cramer, D. G. Truhlar, *Theor. Chim. Acta* **1996**, *93*, 281.

[97] J. D. Thompson, P. Winget, D. G. Truhlar, *Phys. Chem. Comm.* **2001**, *4*, 72.

[98] J. Li, C. J. Cramer, D. G. Truhlar, *Theor. Chem. Acc.* **1998**, *99*, 192.

[99] P. J. Hay, W. R. Wadt, *J. Chem. Phys.* **1985**, *82*, 299.

[100] D. Rappoport, F. Furche, *J. Chem. Phys.* **2010**, *133*, 134105.

[101] T. R. Cundari, W. J. Stevens, *J. Chem. Phys.* **1993**, *98*, 5555.

[102] M. Hurley, L. F. Pacios, P. A. Christiansen, R. B. Ross, W. C. Ermler, *J. Chem. Phys.* **1986**, *84*, 6840.

[103] L. Lajohn, P. A. Christiansen, R. B. Ross, T. Atashroo, W. C. Ermler, *J. Chem. Phys.* **1987**, *87*, 2812.

[104] R. B. Ross, J. M. Powers, T. Atashroo, W. C. Ermler, L. Lajohn, P. A. Christiansen, *J. Chem. Phys.* **1990**, *93*, 6654.

[105] R. B. Ross, S. Gayen, W. C. Ermler, *J. Chem. Phys.* **1994**, *100*, 8145.

[106] W. C. Ermler, R. B. Ross, P. A. Christiansen, *Int. J. Quantum Chem.* **1991**, *40*, 829.

[107] C. S. Nash, B. E. Bursten, W. C. Ermler, *J. Chem. Phys.* **1997**, *106*, 5133.

[108] S. A. Wildman, G. A. DiLabio, P. A. Christiansen, *J. Chem. Phys.* **1997**, *107*, 9975.

[109] Available at: https://bse.pnl.gov/bse/portal. Access date: June 15, 2013.

[110] D. Gottlieb, S. Orszag, Numerical Analysis of Spectral Methods: Theory and Application; SIAM, Philadelphia, **1977**.

[111] J. Hafner, *J. Comput. Chem.* **2008**, *29*, 2044.

[112] G. Vacek, J. K. Perry, J.-M. Langlois, *Chem. Phys. Lett.* **1999**, *310*, 189.

[113] M. S. El Fallah, R. Vicente, J. Tercero, C. Elpelt, E. Rentschler, X. Solans, M. Font-Bardia, *Inorg. Chem.* **2008**, *47*, 6322.

[114] P. Comba, S. Hausberg, B. Martin, *J. Phys. Chem. A* **2009**, *113*, 6751.

[115] Y.-Q. Zhang, C.-L. Luo, *J. Phys. Chem. A* **2011**, *115*, 7778.

[116] S. Sasmal, S. Hazra, P. Kundu, S. Dutta, G. Rajaraman, C. E. Sañudo, S. Mohanta, *Inorg. Chem.* **2011**, *50*, 7257.

[117] N. Marino, D. Armentano, G. De Munno, J. Cano, F. Lloret, M. Julve, *Inorg. Chem.* **2012**, *51*, 4323.

[118] S. Speed, R. Vicente, D. Aravena, E. Ruiz, O. Roubeau, S. J. Teat, S. M. El Fallah, *Inorg. Chem.* **2012**, *51*, 6842.

[119] S. Sasmal, S. Hazra, P. Kundu, S. Majumder, N. Aliaga-Alcalde, E. Ruiz, S. Mohanta, *Inorg. Chem.* **2010**, *49*, 9517.

[120] E. Cremades, J. Cano, E. Ruiz, G. Rajaraman, C. J. Milios, E. K. Brechin, *Inorg. Chem.* **2009**, *48*, 8012.

[121] E. Cremades, E. Ruiz, *Inorg. Chem.* **2010**, *49*, 9641.

[122] D. Steinborn, Fundamentals of Organometallic Catalysis; Wiley: Weinheim, **2011**.

[123] G. Fayet, P. Raybaud, H. Toulhoat, T. de Bruin, *J. Mol. Struct.: THEOCHEM* **2009**, *903*, 100.

[124] T. F. Hughes, R. A. Friesner, *J. Phys. Chem. B* **2011**, *29*, 9280.

[125] T. F. Hughes, J. N. Harvey, R. A. Friesner, *Phys. Chem. Chem. Phys.* **2012**, *14*, 7724.

[126] M. C. Holthausen, *J. Comput. Chem.* **2005**, *26*, 1505.

[127] D. Aravena, E. Ruiz, *Chem. Eur. J.* **2011**, *17*, 8841.

[128] E. Pardo, C. Train, R. Lescouëzec, K. Boubekeur, E. Ruiz, F. Lloret, M. Verdaguer, *Dalton Transact.* **2010**, *39*, 4951.

[129] S. Gomez-Coca, T. Cauchy, E. Ruiz, *Dalton Transact.* **2010**, *39*, 4832.

[130] J. Esteban, E. Ruiz, M. Font-Bardia, T. Calvet, A. Escuer, *Chem. Eur. J.* **2012**, *18*, 3637.

[131] A. I. S. Neves, I. C. Santos, L. C. J. Pereira, C. Rovira, E. Ruiz, D. Belo, M. Almeida, *Eur. J. Inorg. Chem.* **2011**, *31*, 4807.

[132] A. Bassan, M. R. A. Blomberg, P. E. M. Siegbahn, L. Que, Jr., *J. Am. Chem. Soc.* **2002**, *124*, 11056.

[133] D. Suarez, N. Diaz, K. M. Merz, Jr., *J. Am. Chem. Soc.* **2003**, *125*, 15324.

[134] E. M. Sproviero, K. Shinopoulos, J. A. Gascon, J. P. McEvoy, G. W. Brudvig, V. S. Batista, *Phil. Transact. Royal Soc. B* **2008**, *363*, 1149.

[135] V. Guallar, R. A. Friesner, *J. Am. Chem. Soc.* **2004**, *126*, 8501.

[136] L. Tian, R. A. Friesner, *J. Chem. Theory Comput.* **2009**, *5*, 1421.

[137] J. Li, S. T. Schneebeli, J. Bylund, R. Farid, R. A. Friesner, *J. Chem. Theory Comput.* **2011**, *7*, 3829.

[138] F. Ogliaro, N. Harris, S. Cohen, M. Filatov, De S. P. Visser, S. Shaik, *J. Am. Chem. Soc.* **2000**, *122*, 8977.

[139] B. D. Dunietz, M. D. Beachy, Y. X. Cao, D. A. Whittington, S. J. Lippard, R. A. Friesner, *J. Am. Chem. Soc.* **2000**, *122*, 2828.

[140] D. Rinaldo, D. M. Philipp, S. J. Lippard, R. A. Friesner, *J. Am. Chem. Soc.* **2007**, *127*, 3135.

[141] B. F. Gherman, M.-H. Baik, S. J. Lippard, R. A. Friesner, *J. Am. Chem. Soc.* **2004**, *126*, 2978.

[142] Y. Nicolet, P. Amara, J.-M. Mouesca, F.-J. C. Camps, *Proc. Nat. Acad. Sci. USA* **2009**, *106*, 14867.

[143] A. P. Scott, L. Radom, *J. Phys. Chem.* **1996**, *100*, 16502.

[144] J. Baker, A. A. Jarzecki, P. J. Pulay, *Phys. Chem. A* **1998**, *102*, 1412.

[145] J. C. Heckel, A. L. Weisman, S. T. Schneebeli, M. L. Hall, L. J. Sherry, S. M. Stranahan, K. H. Dubay, R. A. Friesner, K. A. Willets, *J. Phys. Chem. A* **2012**, *116*, 6804.

[146] A. M. Ferguson, T. Heritage, P. Jonathon, S. E. Pack, L. Phillips, J. Rogan, P. J. Snaith, *J. Comp.-Aided Mol. Des.* **1997**, *11*, 143.

[147] F. J. Devlin, P. J. Stephens, J. R. Cheeseman, M. J. Frisch, *J. Phys. Chem. A* **1997**, *101*, 6322.

[148] Y. Cao, M. D. Beachy, D. A. Braden, L. Morrill, M. N. Ringnalda, R. A. Friesner, *J. Chem. Phys.* **2005**, *122*, 224116.

[149] S. G. Smith, J. A. Channon, I. Paterson, J. M. Goodman, *Tetrahedron* **2010**, *66*, 6437.

[150] S. G. Smith, J. M. Goodman, *J. Am. Chem. Soc.* **2010**, *132*, 12946.

[151] D. Lu, B. Marten, Y. Cao, M. N. Ringnalda, R. A. Friesner, W. A. Goddard, III, *Chem. Phys. Lett.* **1995**, *242*, 543.

[152] Y. Cao, R. A. Friesner, *J. Chem. Phys.* **2005**, *122*, 104102.

[153] A. V. Marenich, R. M. Olson, C. P. Kelly, C. J. Cramer, D. G. Truhlar, *J. Chem. Theory Comput.* **2007**, *3*, 2011.

[154] A. E. Reed, L. A. Curtiss, F. Weinhold, *Chem. Rev.* **1988**, *88*, 899.

[155] L. E. Chirlian, M. M. Francl, *J. Comput. Chem.* **1987**, *8*, 894.

[156] T. Kuntothom, M. Raab, I. Tvaroška, S. Fort, S. Pengthaisong, J. Cañada, L. Calle, J. Jiménez-Barbero, J. R. K. Cairns, M. Hrmova, *Biochemistry* **2010**, *49*, 8779.

[157] W. D. Cornell, P. Cieplak, C. I. Bayly, P. A. Kollmann, *J. Am. Chem. Soc.* **1993**, *115*, 9620.

[158] P. Cieplak, W. D. Cornell, C. I. Bayly, P. A. Kollman, *J. Comput. Chem.* **1995**, *16*, 1357.

[159] M. C. Owen, L. Tóth, B. Jojárt, I. Komáromi, I. G. Csizmadia, B. J. Viskolcz, *J. Chem. Theory Comput.* **2012**, *8*, 2569.

[160] F. De Proft, C. Van Alsenoy, A. Peeters, W. Langenaeker, P. Geerlings, *J. Comput. Chem.* **2002**, *23*, 1198.

[161] R. G. Parr, W. T. Yang, *J. Am. Chem. Soc.* **1984**, *106*, 4049.

[162] A. K. Chandra, M. T. Nguyen, *J. Chem. Soc. Perkin Trans. 2* **1997**, 1415.

[163] G. Berthier, M. Defranceschi, *Theor. Chem. Acc.* **2009**, *123*, 311.

[164] F. Shah, P. Mukherjee, J. Gut, J. Legac, P. J. Rosenthal, B. L. Tekwani, M. A. Avery, *J. Chem. Inf. Mod.* **2011**, *51*, 852.

[165] L. Klaić, P. C. Trippier, R. K. Mishra, R. I. Morimoto, R. B. Silverman, *J. Am. Chem. Soc.* **2011**, *133*, 19634.

[166] Available at: http://www.hdfgroup.org. Access date: June 26, 2013.

[167] E. R. Johnson, S. Keinan, M.-P. Sanchez, J. Contreras-Garcia, A. J. Cohen, W. Yang, *J. Am. Chem. Soc.* **2010**, *132*, 6498.

[168] R. F. W. Bader, Atoms in Molecules—A Quantum Theory; Oxford University Press: Oxford, UK, **1990**.

[169] R. F. W. Bader, *J. Phys. Chem. A* **1998**, *102*, 7314.

[170] J. Tomasi, B. Mennucci, R. Cammi, *Chem. Rev.* **2005**, *105*, 2999.

[171] B. Zu, Z., Y. C. Zhou, M. J. Holst, J. A. McCammon, *Comm. Comp. Phys.* **2008**, *3*, 973.

[172] C. J. Cramer, D. G. Truhlar, *Chem. Rev.* **1999**, *99*, 2161.

[173] C. P. Kelly, C. J. Cramer, D. G. Truhlar, *J. Chem. Theory Comput.* **2005**, *1*, 1133.

[174] T. Borowski, H. Noack, M. Radon, K. Zych, P. E.M. Siegbahn, *J. Am. Chem. Soc.* **2010**, *132*, 12887.

[175] B. Kuhn, P. Mohr, M. Stahl, *J. Med. Chem.* **2010**, *53*, 2601.

[176] I. S. Ufimtsev, T. J. Martinez, *J. Chem. Theory Comput.* **2009**, *5*, 1004.

[177] C. van Wullen, *J. Comput. Chem.* **2011**, *32*, 1195.

[178] H. J. J. van Dam, de Jong, W. A. De Bylaska, Govind, N. K. Kowalski, T. P. Straatsma, M. Valiev, **2011**, *Wiley Interdiscipl. Rev.—Comput. Mol. Sci. 1*, 888.

[179] Available at: http://www.open-mpi.org. Access date: June 26, 2013.

[180] D. Chasman, M. D. Beachy, L. Wang, R. A. Friesner, *J. Comput. Chem.* **1998**, *19*, 1017.

[181] M. D. Beachy, D. Chasman, R. B. Murphy, R. A. Friesner, *J. Comput. Chem.* **1998**, *19*, 1030.

[182] G. A. Kaminski, J. R. Maple, R. B. Murphy, D. A. Braden, R. A. Friesner, *J. Chem. Theory Comput.* **2005**, *1*, 248.

[183] B. Abu Thaher, P. Koch, V. Schattel, S. Laufer, *J. Med. Chem.* **2009**, *52*, 2613.

[184] R. Selig, M. Goettert, V. Schattel, D. Schollmeyer, W. Albrecht, S. Laufer, *J. Med. Chem.* **2012**, *55*, 8429.

[185] M. R. Berthold, N. Cebron, F. Dill, T. R. Gabriel, T. Kotter, T. Meinl, P. Ohl, C. Sieb, K. Thiel, B. Wiswedel, KNIME: The Konstanz Information Miner. In Studies in Classification, Data Analysis, and Knowledge Organization (GfKL 2007); Springer, Heidelberg, **2007**.

[186] Krieger, E, R. L. Dunbrack, R. W. Hooft, B. Krieger, *Methods Mol. Biol.,* **2012**, *819*, 405.

[187] H. A. Rajapakse, P. G. Nantermet, H. G. Selnick, J. C. Barrow, G. B. McGaughey, S. Munshi, S. R. Lindsley, M. B. Young, P. L. Ngo, M. K. Holloway, M. T. Lai, A. S. Espeseth, X. P. Shi, D. Colussi, B. Pietrak, M. C. Crouthamel, K. Tugusheva, Q. Huang, M. Xu, A. J. Simon, L. Kuo, D. J. Hazuda, S. Graham, J. P. Vacca, *Bioorg. Med. Chem. Lett.* **2010**, *20*, 1885.

[188] K. Lanevskij, P. Japertas, R. Didziapetris, A. Petrauskas, *J. Pharm. Sci.* **2009**, *98*, 122.

[189] D. G. Sprous, R. K. Palmer, J. T. Swanson, M. Lawless, *Curr. Top. Med. Chem.* **2010**, *10*, 619.

[190] R. W. Adler, *J. Am. Chem. Soc.* **2005**, *127*, 7924.

[191] J. Cheng, M. Sprik, *J. Chem. Theory Comput.* **2010**, *6*, 880.

[192] M. Meloun, S. Bordovska, *Anal. Bioanal. Chem.* **2007**, *389*, 1267.

[193] C. Z. Liao, M. C. Nicklaus, *J. Chem. Inf. Mod.* **2009**, *49*, 2801.

[194] G. T. Balogh, B. Gyarmati, B. Nagy, L. Molnar, G. M. Keseru, *QSAR Comb. Sci.* **2009**, *28*, 1148.

[195] M. D. Liptak, G. C. Shields, *J. Am. Chem. Soc.* **2001**, *123*, 7314.

[196] M. D. Liptak, K. C. Gross, S. Feldgus, G. C. Shields, *J. Am. Chem. Soc.* **2002**, *124*, 6421.

[197] K. E. Gutowski, D. A. Dixon, *J. Phys. Chem. A* **2006**, *110*, 12044.

[198] M. Sagall, *J. Comput.-Aided Mol. Des.* **2012**, *26*, 121.

[199] X. Barril, J. F. Luque, *J. Comput.-Aided Mol. Des.* **2012**, *26*, 81.

[200] T. Ercanli, D. B. Boyd, *J. Chem. Inf. Model.* **2005**, *45*, 591.

[201] Z. Fu, Li, X, K. M. Merz, *J. Comput. Chem.* **2011**, *32*, 2587.

[202] M. Sitzmann, I. E. Weidlich, I. V. Filippov, C. Liao, M. L. Peach, W.-D. Ihlenfeldt, R. G. Karki, Y. V. Borodina, R. E. Cachau, M. C. Nicklaus, *J. Chem. Inf. Model.* **2012**, *52*, 739.

[203] J. Poehlsgaard, K. Harpsøe, F. S. Jørgensen, L. Olsen, *J. Chem. Inf. Model.* **2012**, *52*, 409.

[204] R. Gobetto, C. Nervi, B. Romanin, L. Salassa, *Organometallics* **2003**, *22*, 4012.

[205] J. Skopalík, P. Anzenbacher, M. Otyepka, *J. Phys. Chem. B* **2008**, *112*, 8165.

[206] J. Kirchmair, M. J. Williamson, J. D. Tyzack, L. Tan, P. J. Bond, A. Bender, R. C. Glen, *J. Chem. Inf. Model.* **2012**, *52*, 617.

[207] R. Sharma, H. Eng, G. S. Walker, G. Barreiro, A. F. Stepan, K. F. McClure, A. Wolford, P. D. Bonin, P. Cornelius, A. S. Kalgutkar, *Chem. Res. Toxicol.* **2011**, *24*, 2207.

[208] M. Beck, *J. Chem. Inf. Model.* **2005**, *45*, 273.

[209] D. Lupyan, Y. A. Abramov, W. Sherman, *J. Comp.-Aided Mol. Des.* **2012**, *26*, 1195.

[210] D. J. P. Pinto, J. M. Smallheer, D. L. Cheney, R. M. Knabb, R. R. Wexler, *J. Med. Chem.* **2010**, *53*, 6243.

[211] V. Guagnano, P. Furet, C. Spanka, V. Bordas, M. Le Douget, C. Stamm, J. Brueggen, M. R. Jensen, C. Schnell, H. Schmid, M. Wartmann, J. Berghausen, P. Drueckes, A. Zimmerlin, D. Bussiere, J. Murray, D. G. Porta, *J. Med. Chem.* **2011**, *54*, 7066.

[212] P. Bamborough, H. Diallo, J. D. Goodacre, L. Gordon, A. Lewis, J. T. Seal, D. M. Wilson, M. D. Woodrow, C.-W. Chung, *J. Med. Chem.* **2012**, *55*, 587.

[213] J. Boström, A. Hogner, A. Llinàs, E. Wellner, A. T. Plowright, *J. Med. Chem.* **2012**, *55*, 1817.

[214] S. Fomine, S. Gutierrez, M. A. Tlenkopatchev, *J. Organomet. Chem.* **2009**, *694*, 3287.

[215] R. L. Lord, H. Wang, M. Vieweger, M.-H. Baik, *J. Organomet. Chem.* **2006**, *691*, 5505.

[216] D. Benitez, E. Tkatchouk, W. A. Goddard, III, *Chem. Commun.* **2008**, 6194.

[217] M. S. G. Ahlquist, *J. Mol. Cat. A* **2010**, *324*, 3.

[218] M. Ahlquist, R. A. Periana, W. A. Goddard III, *Chem. Commun.* **2009**, 2373.

[219] M. Ahlquist, R. J. Nielsen, R. A. Periana, W. A. Goddard III, *J. Am. Chem. Soc.* **2009**, *131*, 17110.

[220] S. M. Bischof, D. H. Ess, S. K. Meier, J. Oxgaard, R. J. Nielsen, G. Bhalla, W. A. Goddard III, R. A. Periana, *Organometallics* **2010**, *29*, 742.

[221] D. E. Ess, R. J. Nielsen, W. A. Goddard, III, R. A. Periana, *J. Am. Chem. Soc.* **2009**, *131*, 11686.

[222] K. J. H. Young, O. A. Mironov, R. J. Nielsen, M.-J. Cheng, T. Stewart, W. A. Goddard, III, R. A. Periana, *Organometallics* **2011**, *30*, 5088.

[223] S. T. Schneebeli, M. L. Hall, R. Breslow, R. Friesner, *J. Am. Chem. Soc.* **2009**, *131*, 3965.

[224] J. Jover, N. Fey, M. Purdie, L.-G. C. Jones, J. N. Harvey, *J. Mol. Cat. A* **2010**, *324*, 39.

[225] D. Benitez, N. D. Shapiro, E. Tkatchouk, Y. Wang, W. A. Goddard, III, F. D. Toste, *Nature Chem.* **2009**, *1*, 482.

[226] D. Benitez, E. Tkatchouk, A. Z. Gonzalez, W. A. Goddard, III, F. D. Toste, *Org. Lett.* **2009**, *11*, 4798.

[227] A. Z. Gonzalez, D. Benitez, E. Tkatchouk, W. A. Goddard, III, F. D. Toste, *J. Am. Chem. Soc.* **2011**, *133*, 5500.

[228] S. M. Bischof, J. Mu-Cheng, R. J. Nielsen, T. B. Gunnoe, W. A. Goddard, III, R. A. Periana, *Organometallics* **2011**, *30*, 2079.

[229] W. J. Tenn, III, B. L. Conley, C. H. Hovelmann, M. Ahlquist, R. J. Nielsen, D. H. Ess, J. Oxgaard, S. M. Bischof, W. A. Goddard III, R. A. Periana, *J. Am. Chem. Soc.* **2009**, *131*, 2466.

[230] J. M. Keith, W. A. Goddard III, *J. Am. Chem. Soc.* **2009**, *131*, 1416.

[231] C. Tabares-Mendoza, P. Guadarrama, *J. Organomet. Chem.* **2006**, *691*, 2978.

[232] G. Fayet, P. Raybaud, H. Toulhoat, T. de Bruin, *J. Mol. Struct.: THEOCHEM* **2009**, *903*, 100.

[233] R. M. Van Natter, J. S. Coleman, C. R .F. Lund, *J. Mol. Cat. A* **2008**, *292*, 76.

[234] J. A. Ryder, A. K. Chakraborty, A. T. Bell, *J. Catal.* **2003**, *220*, 84.

[235] P. P. Knops-Gerrits, W. A. Goddard, III, *J. Mol. Cat. A* **2001**, *166*, 135.

[236] I. Milas, A. M. Silva, M. A. C. Nascimento, *Appl. Cat. A: General* **2008**, *336*, 17.

[237] I. Milas, M. A. C. Nascimento, *Chem. Phys. Lett.* **2003**, *373*, 379.

[238] C. R. F. Lund, *J. Catal.* **2006**, *243*, 438.

[239] L. G. Bulusheva, A. V. Okotrub, A. G. Kurenya, H. Zhang, H. Zhang, X. Chen, H. Song, *Carbon* **2011**, *49*, 4013.

[240] W. C. West, J. F. Whitacre, N. Leifer, S. Greenbaum, M. Smart, R. Bugga, M. Blanco, S. R. Narayanana, *J. Electrochem. Soc.* **2007**, *154*, A929.

[241] V. S. Bryantsev, M. Blanco, *J. Phys. Chem. Lett.* **2011**, *2*, 379.

[242] V. S. Bryantsev, V. Giordani, W. Walker, M. Blanco, S. Zecevic, K. Sasaki, J. Uddin, D. Addison, G. V. Chase, *J. Phys. Chem. A* **2011**, *115*, 12399.

[243] V. S. Bryantsev, F. Faglioni, *J. Phys. Chem. A* **2012**, *116*, 7128.

[244] W.-Q. Deng, X. Xu, W. A. Goddard, *Nano Lett.* **2004**, *4*, 2331.

[245] Y. S. Park, J. R. Widawsky, M. Kamenetska, M. L. Steigerwald, M. S. Hybertsen, C. Nuckolls, L. Venkataraman, *J. Am. Chem. Soc.* **2009**, *131*, 10820.

[246] S. Acharya, H. Song, J. Lee, P. S. Kwon, J.-P. Lee, G. Yogendranath, Y. H. Kim, Y. H. Jang, T. Lee, S. Samal, J.-S. Lee, *Org. Electron.* **2009**, *10*, 85.

[247] A. C. Fahrenbach, J. C. Barnes, H. Lia, D. Bentez, A. N. Basuray, L. Fang, C.-H. Sue, G. Barin, S. K. Dey, W. A. Goddard, III, J. F. Stoddart, *Proc. Nat. Acad. Sci. USA* **2011**, *108*, 20416.

[248] D. Walter, D. Neuhauser. R. Baer, *Chem. Phys.* **2004**, *299*, 139.

[249] T. Wang, R. A. Friesner, *J. Phys. Chem. C* **2009**, *113*, 2553.

[250] A. Li, S.-H. Wen, J.-L. Song, W.-Q. Deng, *Org. Electron.* **2009**, *10*, 1054.

[251] X.-K. Chen, L.-Y. Zou, J.-X. Fan, S.-F. Zhang, A.-M. Ren, *Org. Electron.* **2012**, *13*, 2832.

[252] J. C. Deaton, S. C. Switalski, D. Y. Kondakov, R. H. Young, T. D. Pawlik, D. J. Giesen, S. B. Harkins, A. J. M. Miller, S. F. Mickenberg, J. C. Peters, *J. Am. Chem. Soc.* **2010**, *132*, 9499.

[253] M. E. Kondakova, T. D. Pawlik, R. H. Young, D. J. Giesen, D. Y. Kondakov, C. T. Brown, J. C. Deaton, J. R. Lenhard, K. P. Klubek, *J. Appl. Phys.* **2008**, *104*, 094501.

[254] X. Ren, D. J. Giesen, M. Rajeswaran, M. Madaras, *Organometallics* **2009**, *28*, 6079.

[255] X. Wang, J. Li, M. E. Thompson, J. I. Zink, *J. Phys. Chem. A* **2007**, *111*, 3256.

[256] T. Sajoto, P. I. Djurovich, A. B. Tamayo, J. Oxgaard, W. A. Goddard, III, M. E. Thompson, *J. Am. Chem. Soc.* **2009**, *131*, 9813.

[257] W. H. Howie, F. Claeyssens, H. Miura, L. M. Peter, *J. Am. Chem. Soc.* **2008**, *130*, 1367.

[258] E. L. Tae, S. H. Lee, J. K. Lee, S. S. Yoo, E. J. Kang, K. B. Yoon, *J. Phys. Chem. B* **2005**, *109*, 22513.

[259] J. L. Banks, H. S. Beard, Y. Cao, A. E. Cho, W. Damm, R. Farid, A. K. Felts, T. A. Halgren, D. T. Mainz, J. R. Maple, R. Murphy, D. M. Philipp, M. P. Repasky, L. Y. Zhang, B. J. Berne, R. A. Friesner, E. Gallicchio, R. M. Levy, *J. Comput. Chem.* **2005**, *26*, 1752.

[260] A. D. Bochevarov, J. Li, W. J. Song, S. J. Lippard, R. A. Friesner, *J. Am. Chem. Soc.* **2011**, *133*, 7384.

[261] J. Y. Kravitz, V. L. Pecoraro, H. A. Carlson, *J. Chem. Theory Comput.* **2005**, *1*, 1265.

[262] M. Wirstam, S. J. Lippard, R. A. Friesner, *J. Am. Chem. Soc.* **2003**, *125*, 3980.

[263] A. Ganguly, P. C. Bevilacqua, S. Hammes-Schiffer, *J. Phys. Chem. Lett.* **2011**, *2*, 2906.

[264] I. Tvaroška, S. Kozmon, M. Wimmerová, J. Koča, *J. Am. Chem. Soc.* **2012**, *134*, 15563.

[265] L. Tian, R. A. Friesner, *J. Chem. Theory Comput.* **2009**, *5*, 1421.

[266] J. L. Klepeis, L.-K. Larsen, R. O. Dror, D. E. Shaw, *Curr. Opt. Struct. Biol.* **2009**, *19*, 120.

[267] S. J. Marrink, A. H. de Vries, D. P. Tieleman, *Biochim. Biophys. Acta* 2009, **1788**, 149.

[268] D. E. Shaw, P. Maragakis, K. Lindorff-Larsen, S. Piana, R. O. Dror, M. P. Eastwood, J. A. Bank, J. M. Jumper, J. K. Salmon, Y. Shan, W. Wriggers, *Science* **2010**, *330*, 341.

[269] J. B. Klauda, N. Kucerka, B. R. Brooks, R. W. Pastor, J. F. Nagle, *Biophys. J.* **2006**, *90*, 2796.

[270] E. Harder, A. D. MacKerell, Jr., B. Roux, *J. Am. Chem. Soc.* **2009**, *131*, 2760.

[271] D. B. Kitchen, H. Decornez, J. R. Furr, J. Bajorath, *Nature Rev. Drug Discov.* **2004**, *3*, 935.

[272] C. J. Margulis, H. A. Stern, B. J. Berne, *J. Phys. Chem. B* **2002**, *106*, 12017.

[273] E. Harder, D. E. Walters, Y. D. Bodnar, R. S. Faibish, B. Roux, *J. Phys. Chem. B* **2009**, *113*, 10177.

[274] M. Levitt, *Nat. Struct. Bio.* **2001**, *8*, 392.

[275] A. D. MacKerell, Jr., D. Bashford, M. Bellott, R. L. Dunbrack, Jr., J. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, I. Reiher, W. E., B. Roux, M. Schlenkrich, J. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, M. Karplus, *J. Phys. Chem B* **1998**, *102*, 3586.

[276] W. L. Jorgensen, T.-J. Rives, *J. Am. Chem. Soc.* **1998**, *110*, 1657.

[277] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, P. A. Kollman, *J. Am. Chem. Soc.* **1995**, *117*, 5179.

[278] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, C. Simmerling, *Proteins* **2006**, *65*, 712.

[279] M. Feig, A. D. MacKerell, Jr., C. L. Brooks, III, *J. Phys. Chem. B* **2002**, *107*, 2831.

[280] T. D. Rasmussen, P. Ren, J. W. Ponder, F. Jensen, *Int. J. Quantum Chem.* **2007**, *107*, 1390.

[281] G. A. Kaminski, R. A. Friesner, T.-J. Rives, W. L. Jorgensen, *J. Phys. Chem. B* **2001**, *105*, 6474.

[282] M. P. Jacobson, G. A. Kaminski, R. A. Friesner, C. S. Rapp, *J. Phys. Chem. B* **2002**, *106*, 11673.

[283] A. D. MacKerell, Jr., M. Feig, C. L. Brooks, III, *J. Comput. Chem.* **2004**, *25*, 1400.

[284] G. A. Kaminski, H. A. Stern, B. J. Berne, R. A. Friesner, Y. X. Cao, R. B. Murphy, R. Zhou, T. A. Halgren, *J. Comput. Chem.* **2002**, *23*, 1515.

[285] D. Shivakumar, E. Harder, W. Damm, R. A. Friesner, W. Sherman, *J. Chem. Theory Comput.* **2012**, *8*, 2553.

[286] L. Wang, B. J. Berne, R. A. Friesner, *Proc. Nat. Acad. Sci. USA* **2012**, *109*, 1937.

[287] L. Wang, R. A. Friesner, B. J. Berne, *J. Phys. Chem. B* **2011**, *115*, 9431.

[288] A. C. T. van Duin, S. Dasgupta, F. Lorant, W. A. Goddard, III, *J. Phys. Chem. A* **2001**, *105*, 9396.

[289] K. D. Nielson, A. C. T. van Duin, J. Oxgaard, W.-Q. Deng, W. A. Goddard, III, *J. Phys. Chem. A* **2005**, *109*, 493.

[290] K. Chenoweth, A. C. T. van Duin, W. A. Goddard III, *J. Phys. Chem. A* **2008**, *112*, 1040.

[291] A. Strachan, A. C. T. van Duin, D. Chakraborty, S. Dasgupta, W. A. Goddard, III, *Phys. Rev. Lett.* **2003**, *91*, 098301.

[292] A. Bagri, C. Mattevi, M. Acik, Y. J. Chabal, M. Chhowalla, V. B. Shenoy, *Nature Chem.* **2010**, *2*, 581.