

Advances in methods and algorithms in a modern quantum chemistry program package†

Yihan Shao,^a Laszlo Fusti Molnar,^a Yousung Jung,^a Jörg Kussmann,^a Christian Ochsenfeld,^a Shawn T. Brown,^a Andrew T.B. Gilbert,^a Lyudmila V. Slipchenko,^a Sergey V. Levchenko,^a Darragh P. O'Neill,^a Robert A. DiStasio Jr,^a Rohini C. Lochan,^a Tao Wang,^a Gregory J.O. Beran,^a Nicholas A. Besley,^a John M. Herbert,^a Ching Yeh Lin,^a Troy Van Voorhis,^a Siu Hung Chien,^a Alex Sodt,^a Ryan P. Steele,^a Vitaly A. Rassolov,^a Paul E. Maslen,^a Prakashan P. Korambath,^a Ross D. Adamson,^a Brian Austin,^a Jon Baker,^a Edward F. C. Byrd,^a Holger Dachsel,^a Robert J. Doerksen,^a Andreas Dreuw,^a Barry D. Dunietz,^a Anthony D. Dutoi,^a Thomas R. Furlani,^a Steven R. Gwaltney,^a Andreas Heyden,^a So Hirata,^a Chao-Ping Hsu,^a Gary Kedziora,^a Rustam Z. Khaliulin,^a Phil Klunzinger,^a Aaron M. Lee,^a Michael S. Lee,^a WanZhen Liang,^a Itay Lotan,^a Nikhil Nair,^a Baron Peters,^a Emil I. Proynov,^a Piotr A. Pieniazek,^a Young Min Rhee,^a Jim Ritchie,^a Edina Rosta,^a C. David Sherrill,^a Andrew C. Simmonett,^a Joseph E. Subotnik,^a H. Lee Woodcock III,^a Weimin Zhang,^a Alexis T. Bell,^a Arup K. Chakraborty,^a Daniel M. Chipman,^a Frerich J. Keil,^a Arieh Warshel,^a Warren J. Hehre,^a Henry F. Schaefer III,^a Jing Kong,^a Anna I. Krylov,^a Peter M. W. Gill^a and Martin Head-Gordon^{*bc}

Received 16th December 2005, Accepted 25th April 2006

First published as an Advance Article on the web 12th June 2006

DOI: 10.1039/b517914a

Advances in theory and algorithms for electronic structure calculations must be incorporated into program packages to enable them to become routinely used by the broader chemical community. This work reviews advances made over the past five years or so that constitute the major improvements contained in a new release of the Q-Chem quantum chemistry package, together with illustrative timings and applications. Specific developments discussed include fast methods for density functional theory calculations, linear scaling evaluation of energies, NMR chemical shifts and electric properties, fast auxiliary basis function methods for correlated energies and gradients, equation-of-motion coupled cluster methods for ground and excited states, geminal wavefunctions, embedding methods and techniques for exploring potential energy surfaces.

1. Introduction

Here we describe recent developments in molecular electronic structure models and algorithms, by a large loosely connected team of scientists, who contribute to the Q-Chem program package. This introduction sets the stage for discussing these developments, by first giving a short perspective on the present status of computational quantum chemistry, and the key issues it faces today. We begin with a general overview of the goals of computational quantum chemistry and then

reprise the overall strategy that is commonly adopted at the moment to realize those goals. At this stage, it is possible to identify both current successes, and some of the challenges. These challenges cross several scientific boundaries, from electronic structure theory itself, which bridges theoretical chemistry and molecular physics, towards applied mathematics and computer science. We then turn to discussing specific developments in the body of the paper.

1.1 Context

Molecular electronic structure theory has as its goal the prediction of molecular structure and properties by solving the equations of quantum mechanics from first principles—or “*ab initio*”. This is done within the Born–Oppenheimer separation, where the electronic structure problem is to solve the time-independent Schrödinger equation for the electrons, in the field of fixed nuclear charges at positions \mathbf{R}_A , possibly with external applied electric and/or magnetic fields. The eigenvalues $E_i(\mathbf{R}_A)$ of the electronic Schrödinger equation define the potential energy surface of the ground state ($i = 0$) and the electronic excited states ($i > 0$), respectively.

^a See Acknowledgements

^b Department of Chemistry, University of California, Berkeley, CA 94720, USA

^c Chemical Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA. E-mail: mhg@cchem.berkeley.edu

† Statement on commercial interests: This paper concentrates on the scientific and technical aspects of Q-Chem 3.0. For the record we (the authors) state that Q-Chem 3.0 is distributed by Q-Chem Inc., which at the time of writing has part-owners including the corresponding author of this paper, M. Head-Gordon, and co-authors P. M. W. Gill, J. Kong, A. I. Krylov, and H. F. Schaefer, III.

“Walking” on the potential surfaces allows one to characterize stationary points as either stable structures (local minima) or transition structures (saddle points), *etc.* Derivatives of E_i with respect to parameters in the Hamiltonian, such as nuclear positions or applied fields, yield observable molecular properties. Through this approach, many of the central properties of chemistry, such as molecular structure, relative energies of different structures, and spectroscopic observables may in principle be predicted directly from quantum mechanics.

While this vision was recognized early on by pioneers such as Dirac¹ and Lewis,² to name only two, it was, and is, impossible to realize directly. Analytical solutions are not available beyond the simplest model systems, such as the hydrogen atom. Therefore one is compelled to seek numerical approximations, and it is useful to proceed in two steps. (1) Make a finite basis expansion to represent functions of one electron. The basis functions are usually hydrogen-like orbitals, composed of atom-centered fixed linear combinations of Gaussian functions.³ Plane waves are a commonly used alternative in solid state physics and for treating condensed matter. (2) Make a well-defined approximation for the interactions between electrons, which is called the electron correlation model. The use of a complete (and thus infinite-dimensional) one-electron basis is unfeasible: in practice finite basis sets must be employed. Exact solution of the time-independent Schrödinger equation in a finite basis, called full configuration interaction (FCI), is also unfeasible. It is an exponentially difficult problem as a function of the number of basis functions N and electrons n , since ${}^N C_n$ possible configurations enter the wavefunction. FCI calculations are used as benchmarks, and while the largest ones reported involve literally billions of terms, they are on molecules that have no more than 10 or 15 electrons.⁴

To make practical progress, we are then faced with the necessity of making well-defined approximations for both the basis expansion, and the treatment of electron–electron correlations. Well-defined approximations, for our purposes, are those that make no reference to features of specific molecules beyond what is contained in the molecular Hamiltonian: the nuclear positions and charges, the total charge and spin state. Other properties of the atoms may be used to define the basis set in terms of atomic orbitals. In this approach, the basis set and correlation treatment define what John Pople termed a theoretical model chemistry⁵—a complete computational prescription for simulating chemistry, which will yield predictions for observables that can be compared against experiment. After validation against known data, the use of such a model for predictive purposes becomes possible. After all, the input to the calculations is just the nuclear positions and charges, the overall charge and spin multiplicity of the molecule, and the basis set and correlation model. From this elementary input one can obtain as output the model potential energy surfaces $E_i(\mathbf{R}_i)$ from first principles—an approximate realization of the vision of Dirac.

Space limits preclude us from digging into the standard model chemistries at this stage—there are textbooks,^{6–9} as well as some general reviews^{10–12} that play this role—and we will provide short overviews later in the paper in the context of specific recent developments. Suffice to say, there are two main

branches to the family tree of quantum chemistry—one based on Kohn–Sham density functional theory (DFT) where the central variable is the electron density, $\rho(\mathbf{r})$,^{13,14} and the other based on modeling the wavefunction.^{6,9} One important point is that these theoretical chemical models are computationally tractable on modern computers for molecules (and even extended systems) of chemical interest. This is partly due to developments in the theoretical models and algorithms (our main topic). It is also partly due to the pace of developments in computing technology. Computer processing power has been doubling roughly every two years or so since the 1970’s in accord with Moore’s (empirical) law, and the cost of acquiring this power has plummeted. Another important point is that results of useful accuracy are obtained with these standard models. For example, predictions of molecular structure are good to roughly 0.01 Å, qualitatively useful predictions of vibrational, NMR and electronic spectra are possible, and it is possible to obtain reasonably reliable results for the energy changes associated with chemical reactions. As a consequence, computational quantum chemistry has changed over the past twenty years from a specialist research area to a branch of chemistry that is usable by chemists at large as a supplement to (and even occasionally as a replacement for) experiments.

With this transition, program packages that implement *ab initio* theoretical model chemistries have become widely used by a world-wide community of tens of thousands of chemists. The first programs were produced not long after the first computers were available, spear-headed by visionaries such as Frank Boys.¹⁵ The first usable programs appeared as early as 1970, with the introduction of the Gaussian 70 program, the creation of John Pople, and his research students and postdocs at the time. Since then, there has been steady growth in terms of both improved and new programs, including over ten commercial programs, and at least the same number of public domain programs, many of which have been described in scientific reviews and overviews (such as ref. 16–19). This growth has been driven by continued developments in theory and algorithms, synergistically coupled to developments in general purpose computing hardware. Broad acceptance of the usefulness of standard theoretical chemical models, and the pervasiveness of the software tools has taken chemical research to the stage where a substantial fraction of experimental papers employ electronic structure calculations in addition to synthesis and measurement.

This is the context in which the Q-Chem program package exists. It is one of the commercial quantum chemistry codes, where end-users purchase licenses for executable versions of either Q-Chem itself or for the Spartan modeling package, which includes Q-Chem as the “back-end” for computational quantum chemistry calculations. Q-Chem originated as a splinter group from the Gaussian program collaboration centered around John Pople’s group, starting in 1993. Key technical progress in the previous major release was described in an earlier article,²⁰ with a separate report on parallel capabilities.²¹ Since that time much scientific progress has occurred as discussed below, but there has also been a great loss. John Pople, 1998 Nobel Laureate for his work on quantum chemistry, was an active scientific contributor to Q-Chem and a member of Q-Chem’s Board of Directors, since

January 2, 1999, when his six-year no-compete obligation to Gaussian expired. John passed away in March 2004 after a brief illness at the age of 78, and he is sorely missed. We would like to dedicate this article to his memory.^{22–26} Within Q-Chem, John pressed his career-long vision of new capabilities from a scientist's perspective, a focus on software robustness from a user's perspective, as well as strict financial discipline from a director's perspective. He was our harshest critic and truest friend.

1.2 Challenges

While the impact of *ab initio* quantum chemistry is already great, there are a number of key areas where further progress is needed. We identify the ones that primarily concern us in this section. They provide good motivation for a number of the new developments discussed later in the paper, and present a context in which the success of those developments can be measured. For further perspective, we also touch on some of the other frontiers of electronic structure theory that are not specifically addressed in our new developments.

Efficient algorithms. Accepting the viability of the standard theoretical chemical models for many problems of interest, the first issue one faces is the question of the performance and scaling of the computational algorithms used to implement those models. Even in areas that seem as well-defined and mature as computational matrix algebra,²⁷ significant progress continues to occur in the performance of basic kernels such as matrix diagonalization. This type of progress does not alter the scaling of the diagonalization problem with size of the matrix, but instead reduces the *prefactor* that multiplies the cubic size scaling of the cost. Computational quantum chemistry is presently at a far less evolved stage, where the numerical kernels are still being actively developed, and in some cases even replaced or redefined by new formulations.

Accordingly, advancing algorithmic efficiency is one of the main themes of this paper. We must admit that assessing improvements in efficiency and performance is a potentially troublesome issue, because it raises the related problem of the comparative performance of different software packages. It is difficult to establish a fair comparison between different codes due to the fact that options controlling precision and cutoffs do not always map directly, and also because markedly different results can be obtained on different computer architectures. Also, some codes such as the Gaussian series explicitly forbid publication or sharing of comparative timings as part of their license agreement. For reasons such as these where performance is a key issue, we have elected to simply compare against the previous major release of Q-Chem²⁰ in all results discussed here—in other words we focus solely on the extent of internal algorithmic and coding improvements. Numerous examples are discussed later, including new two-electron integral algorithms, and the use of auxiliary basis expansions. Other examples emerge in the context of other frontiers, such as dealing with the size-scaling of calculations, and enabling higher accuracy calculations.

Scaling with system size. We have already discussed the fact that brute force solution of the Schrödinger equation has

computational costs that increase exponentially with the number of electrons, n , and basis functions N , as the number of terms in the wavefunction is N^{C_n} . The standard theoretical chemical models, by contrast, exhibit only polynomial cost increases with molecular size, such as N^3 , N^4 ,...etc. The self-consistent field (SCF) molecular orbital models discussed in section 2, Kohn–Sham density functional theory (DFT), and Hartree–Fock (HF), are in principle N^3 as a function of molecular size, but can scale N^4 (due to 4-center 2-electron integral evaluation) for very small systems. Wavefunction-based treatments of electron correlations scale with higher polynomial powers of molecular size, typically N^5 , N^6 and N^7 for the widely used second order Møller–Plesset (MP2) and coupled cluster (CC) models discussed in sections 3 and 4.

These scaling relations for computational cost as a function of molecular size are usually based on the assumption that the matrix element manipulations involve dense linear algebra. Fortunately this does not have to be true for large molecules where sparsity in the matrix elements and wavefunction unknowns can be exploited to make such estimates overly pessimistic. For example, there are $O(N^4)$ 4-center 2-electron integrals, formally, which gives rise to N^4 scaling for SCF calculations. However, for large molecules only $O(N^2)$ of them are numerically significant,²⁸ if localized Gaussian basis functions are used. Further algorithmic developments of increasing sophistication can even replace two-electron integrals corresponding to long-range interactions by multipole-based representations, which opens the way for linear scaling in steps that involve 2-electron integral processing for treating Coulomb interactions. This is a particular example but, as emphasized by Kohn,²⁹ linear scaling is in principle possible in general because of what he termed the near-sightedness of electrons—their interactions with distant electrons are effectively screened. Locality in real space can be exploited in calculations that employ either real-space representations of the wavefunction or density, or local basis functions. Thus the long-range goal for electronic structure calculations on nanoscale systems is to achieve linear scaling with molecular size.

However, true linear scaling is a goal that cannot always be fully realized in practice because the length scales on which this short-sightedness occurs often cannot be routinely reached yet, particularly for more sophisticated theoretical model chemistries. Locality of electronic structure, as measured for instance by the “size” of a localized orbital is on the order of ten or so atoms in a line for insulators.³⁰ Such a size is measured to a high precision of about 1 ppm, as opposed to cartoon representations, which are more typically localized to within two or three atoms. It means a natural “linear scaling regime” would correspond to thousands of atoms in three dimensions, well beyond the realm of the routine at present. See section 2.2, 2.3 and 2.5 for what is possible at present with self-consistent field methods. It is therefore also useful to seek faster algorithms which become effective at smaller system sizes, perhaps through making additional approximations which are nonetheless not linear scaling. This is a theme that pervades much recent work in quantum chemistry, and shows up in many of the specific topics discussed later on, such as approximations to MP2 theory in section 3, and fast methods for strong correlations in Section 5. Approaches that subdivide

the system into a primary region that is modeled at a relatively high level and a secondary region that serves as an environment for the primary region represent another approach to the problem of making calculations on very large systems feasible in the short run, and a variety of these approaches are discussed in section 6.

Higher accuracy calculations. Quantum chemical approximations can be viewed as points on a two-dimensional (2-D) graph where one axis is the extent of completeness of the one-particle basis set and the other axis is the extent of completeness of the electron correlation treatment. High accuracy calculations are ones where both of these approximations are carried to a sufficiently sophisticated level that one approaches convergence in the resulting computed observables. For many years, the target in electronic structure theory has been an overall accuracy of approximately $1\text{--}2\text{ kcal mol}^{-1}$ in thermochemistry, a standard which has come to be called “chemical accuracy”. Methods such as the G3 approach (a composite of several separate calculations) achieve this accuracy, albeit at great computational cost.³¹ On-going developments yield even higher accuracy at even greater computational cost.^{32,33}

Another recurring theme of this paper are new developments that are designed to permit higher accuracy calculations at reduced computational cost. We foreshadow a few of them now. Dual basis approaches to economically approaching the basis set limit in DFT, HF and MP2 calculations are discussed in Sections 2 and 3. Auxiliary basis sets reduce the growth of computational cost with size of atomic orbital basis and are discussed in the context of MP2 theory in Section 3. The use of non-standard reference configurations such as high spin states to permit more accurate coupled cluster calculations is discussed in Section 4.

There are also many other approaches to the problem of high accuracy which lie beyond our present scope. One important class of examples are the Quantum Monte Carlo methods,³⁴ which avoid the need for virtual orbitals, in exchange for an inherently stochastic character. Another class of examples are the so-called r12 methods³⁵ which introduce an explicit dependence on inter-electronic distances into the wavefunction. Researchers with a focus on small molecules and connections with spectroscopy are now starting to focus on the challenge of achieving what they describe as “spectroscopic accuracy”.

Strongly correlated systems. The most widely used standard theoretical models (HF, DFT, MP2, CCSD *etc.*) are all based on the assumption that a single electron configuration is a qualitatively correct picture of electronic structure. This is usually a good approximation for stable molecules, because the electrons pair into bonds. However, it sometimes breaks down for reactive molecules (or molecules in reactive configurations, such as transition structures), where the energy gap between the nominal ground state configuration, and excited configurations becomes small, and they mix together strongly. This happens in singlet diradicaloid molecules in organic chemistry, in certain types of low-spin antiferromagnetically coupled transition metal systems in inorganic chemistry, and

in many transition structures (*e.g.* of diradicaloid type). It remains a frontier of electronic structure theory to develop tractable approximations for the more complex wavefunctions in this category, because of their potential multireference character.

In particular, what should the form of the strongly correlated wavefunction be? One widely used approach is to employ “complete active space” (CAS) methods³⁶ to form the exact (full CI) wavefunction in a small set of optimized orbitals that are important or “active”, and then optionally correct by perturbation theory. But the cost of a CAS calculation is exponential just like FCI, and therefore molecule-specific or even geometry-specific choices of the active space must be made to keep it small enough to be feasible. One promising way around this problem are novel approximations to FCI based on the density matrix renormalization group approach, although this work is still at an early stage.^{37,38}

Another option is to approximate the wavefunction but then allow a larger range of active orbitals (either all valence orbitals or all orbitals). Of course, incomplete approximations to the wavefunction in a full active space may not be sufficiently accurate. Research continues on true multireference methods,³⁹ which we do not consider in this paper, and also on simpler models that can capture these strong correlations, which we do discuss. Coupled cluster methods are one such possibility, as well as theories based on geminals, and approaches that blend the two, such as coupled cluster perfect and imperfect pairing. These approaches are discussed in Sections 4 and 5. One can even imagine entirely new approaches to electronic structure that may offer future advantages such as intracule-based models which are foreshadowed in Section 9.

Excited states. Obtaining reliable descriptions of the electronic excited states (the higher eigenvalues E_i , $i > 0$) presents a challenge somewhat similar to the treatment of highly correlated systems. There is a great diversity of excited states—some involve one-electron valence excitations from the ground state, such as an $n \rightarrow \pi^*$ state in carbonyl groups, while others involve excitations into very diffuse Rydberg orbitals, such as lower excited states of atoms and saturated molecules, some involve charge transfer from one region to another, some involve significant contributions from promotions of 2 (or even more) electrons, such as the so-called dark states of polyenes, *etc.* Thus it is difficult to build economical and yet accurate theoretical chemical models that can describe this diversity. Most popular at present are approaches based on time-dependent extensions to density functional theory (TDDFT), although this still suffers from significant limitations.^{40,41} Coupled cluster theory provides an alternative framework that is useful for smaller molecules, as discussed in Section 4. CAS-based methods are another alternative, although the selection of the active space is a tremendous challenge.

High-dimensional potential energy surfaces. The realm of routine applicability of quantum chemical methods ranges from small molecules to systems of several hundred atoms, depending upon the computational resources available and the

level of theory chosen. As studies on larger molecules become more feasible, new challenges arise in the exploration of their potential energy surfaces, which become increasingly high-dimensional. Isolating and characterizing local minima and the saddle points that connect them becomes increasingly difficult, and indeed less meaningful, as their number grows exponentially with molecular size. This issue demands increasing convergence with the methods of molecular dynamics and statistical mechanical simulations in general. Some aspects of this issue are discussed in Section 7, including direct dynamics, and methods for locating reaction paths.

Relativistic quantum chemistry. Largely, the quantum chemical methods discussed in this paper are non-relativistic in nature—with the relativistic character of inner shell electrons in heavier elements being “hidden” inside effective core potentials. However, the adequate treatment of all electrons in molecules that contain heavy elements is another challenge for the field, where significant progress is being made.^{42,43}

1.3 Outline

The first several sections of this paper focus on summarizing new developments in the available model chemistries in Q-Chem, starting with self-consistent field density functional theories (DFT)⁴⁴ and Hartree–Fock (HF) methods (Section 2). This is followed in Section 3 by the least expensive corrections to the HF wavefunction, namely second-order perturbation theory (MP2), as well as modifications to it. Even here new models and methods are continuing to arise, with new strengths. We then turn to more advanced coupled cluster methods in Section 4, with particular emphasis on the new spin-flip approach and its implications. This is followed by a discussion of novel inexpensive treatments of highly correlated systems in Section 5. These methods model the so-called non-dynamical (or bond-breaking) correlations, and then correct perturbatively for the remaining dynamical (or atomic-like) correlations.

The following several sections build upon the basic theoretical model chemistries in several different directions. The first is the definition of hybrid model chemistries, in which part of a system is described at a higher level of sophistication and a larger part (the environment) is described at a lower level, as in methods that mix quantum mechanics with molecular mechanics (QM/MM) or describe solvation using continuum or reaction field solvent models. The next section concerns molecular properties, which are generally given as responses of the molecular energy to perturbations, such as applying magnetic fields (NMR properties), electric fields, or displacing nuclei (computing vibrational frequencies). The focus here is on new linear scaling NMR algorithms, which open up a new size scale for applications. After this, we turn to issues concerned with walking on potential energy surfaces, and discuss efficient location of transition structures, reaction paths, and the execution of direct dynamics using quantum chemistry. Finally we discuss capabilities related to the analysis of electronic structure calculations.

2. Fast density functional theory and Hartree–Fock calculations

Self-consistent field (SCF) methods are the cornerstone of modern quantum chemistry. They include the HF model and Kohn–Sham DFT methods. They describe electronic structure in terms of molecular orbitals with electron–electron correlations either averaged out (as in the HF model), or included implicitly through functionals as in DFT. Present-day DFT methods in particular, often offer the best compromise between accuracy and feasibility for chemical applications to medium and larger sized molecules. This began with development of gradient-corrected functionals in the late 1980’s, and was cemented with the ground-breaking development of hybrid functionals such as B3LYP by Becke in 1993.^{45,46} Hybrid functionals include a component of exact (Hartree–Fock) exchange. Functionals that do not include exact exchange such as gradient-corrected, or local density functionals, offer some advantage in computational efficiency as the exact exchange terms are computationally demanding. However they are generally slightly less accurate.

Intense effort continues on the design of new density functionals, with numerous proposals appearing each year. Some of these new ones including tau-containing functionals such as BMK,⁴⁷ and EDF2 which is optimized for vibrational frequency predictions⁴⁸ are included in Q-Chem, in addition to more standard ones, and users can also input their own functionals. Many of the new functionals offer advantages in niche applications, particularly if they have been empirically parameterized for such applications. However, none that we are aware of appear to be broadly better than B3LYP and related functionals for general use. Finding functionals that improve over B3LYP without destroying the simplicity that makes DFT so computationally efficient has simply proven to be a most difficult task. There are, however, genuine physical weaknesses associated with present-day functionals, such as the incomplete cancellation of Coulomb self-interaction by the exchange functional⁴⁹ and lack of long-range van der Waals interactions in the correlation functional,⁵⁰ that provide ample motivation for continuing effort in this area in the future.

In the absence of usable breakthroughs on these difficult issues, our main focus on new developments in DFT calculations is therefore on algorithms to perform the calculations, rather than on new functionals, as we discuss below.

2.1 Improved 2-electron integral evaluation

Evaluating 4-center two-electron integrals over Gaussian atomic orbital basis functions is the fundamental step in the assembly of the mean-field Hamiltonian in any SCF calculation which involves exact exchange (e.g. B3LYP, or HF itself), and has thus been a perennial research topic in quantum chemistry. Q-Chem combines the Head-Gordon-Pople (HGP) method,⁵¹ the COLD prism method⁵² and the Rys polynomial method^{53,54} to evaluate two-electron integrals, derivatives and second derivatives.⁵⁵ All 2-electron 4-center integrals are classified according to their angular momentum types and degrees of contraction. For each type of integral, the program chooses the available algorithm with lowest cost. In practice, the HGP method is chosen for most integral classes

Table 1 CPU timings (in seconds) for evaluation of the gradient and Hessian terms for closed shell HF calculations on the AZT molecule^a

Basis set	AIX			Linux		
	Old	New	New/old	Old	New	New/old
Gradient evaluation: 2-electron integral derivative term						
3-21G	34 s	20 s	0.58	25 s	14 s	0.56
6-31G**	259 s	147 s	0.57	212 s	120 s	0.57
DZ	128 s	118 s	0.92	72 s	62 s	0.86
cc-pVDZ	398 s	274 s	0.69	308 s	185 s	0.60
Hessian evaluation: 2-electron integral second derivative term						
3-21G	294 s	136 s	0.46	238 s	100 s	0.42
6-31G**	2520 s	976 s	0.39	2065 s	828 s	0.40
DZ	631 s	332 s	0.53	600 s	230 s	0.38
cc-pVDZ	3202 s	1192 s	0.37	2715 s	866 s	0.32

^a The AIX timings were obtained on a single processor of an IBM RS/6000 workstation running the AIX4 operating system, and the Linux timings were obtained on a single 2 GHz processor of an AMD Opteron cluster where the Q-Chem executable was compiled with Intel 32-bit compilers.

in a gradient or hessian calculation, and thus it usually dominates the total CPU time. As angular momentum (and derivative order) increases, the memory required per shell quartet grows strongly and we observe corresponding performance degradation with the existing algorithm that processes batches of shell quartets in vector-like fashion. Therefore we have re-written our HGP codes for the evaluation of the two-electron integrals, integral derivative terms in the gradient evaluation, and the corresponding second derivative terms in the Hessian evaluation, fundamentally reordering the loops. Our approach aims to improve code efficiency by reducing cache misses rather than by explicitly reducing floating point operation counts. Some timing results from a Hartree-Fock calculation on the AZT molecule are shown below, showing significant efficiency improvements over our previous implementation (Table 1).

2.2 Fast Coulomb matrix evaluation

For DFT calculations without exact exchange (*e.g.* BLYP or EDF1), the 4-center 2-electron integrals are not necessarily required, since it is possible to formulate alternative approaches that evaluate their contribution to the Coulomb terms in the mean field Hamiltonian (the Coulomb matrix). In the last decade, dramatic progress has been made to improve the efficiency of Coulomb calculations, including techniques such as the Continuous Fast Multipole Method (CFMM)^{56–59} and the J-Engine^{60–62} which were first implemented in Q-Chem. An important recent development is the advent of the Fourier Transform Coulomb (FTC) method (ref. 63–65) and we have recently developed an improved version of this approach for evaluating Coulomb matrices and forces.⁶⁶ It replaces the least efficient part of the previous Coulomb methods with an accurate numerical integration scheme that scales as $O(N^2)$ instead of $O(N^4)$ with the basis size. The result is a smaller coefficient of linear scaling with respect to molecular size, which speeds up the Coulomb calculation by several times over the most efficient existing Coulomb code, which was the combination of CFMM and the J-engine.

Table 2 Efficiency improvements^a in the linear scaling Coulomb technique by using the Fourier transform Coulomb method, based on calculations on alanine oligomers of length 5, 10, or 15 amino acids

Basis set	5(old)	5(ftc ^b)	10(old)	10(ftc ^b)	15(old)	15(ftc ^b)
6-31G(df,pd)	39	27	145	98	263	169
6-31G+(df,pd)	109	46	736	160	1240	362
cc-pvdz	56	21	177	79	369	143
cc-pvtz	361	85	1165	280	2482	551

^a All calculations have been performed on a single processor of a 2.0 GHz Opteron cluster. The timings are in minutes and include the total costs of the Coulomb part for all SCF iterations and the force calculations. A SAD (superposition of atomic density) guess and 10^{-12} integral threshold were used. The SCF convergence threshold was 10^{-7} for the maximum DIIS error. ^b $3.8 a_0^{-1}$ grid density has been used in the plane wave part of the calculations.

The practical speed-ups obtained using the FTC technique strongly depend on the basis function pair density *i.e.* the number of near field integrals which have to be evaluated analytically. This quantity depends on the density of the system in question and the basis set that is used. Table 2 depicts the effect of adding FTC to the Coulomb calculation in Q-Chem 3.0 for an alanine test series using various basis sets. The 5-mer, 10-mer and 15-mer correspond to 21, 41, and 61 non-hydrogen atoms, respectively. The speed-ups from using FTC are moderate when relatively small Pople-type basis sets are used without diffuse functions. However, the computational penalty usually associated with adding diffuse and polarization functions to the basis set is greatly reduced by the FTC due to its better $O(N^2)$ scaling with the number of basis functions per atom. This is particularly exciting because it is for large extended basis sets that the existing linear scaling methods such as CFMM are least effective (for given molecule size). A good example is the 6-31G+(df,pd) basis calculations where, due to the role of the diffuse functions, the linear scaling region has not been reached yet by 20 heavy atoms (an Alanine₅ molecule). The computational costs of the Coulomb part using our existing algorithms went up from 109 to over 700 min. by increasing the system size from 21 to 41 heavy atoms (Alanine₅ to Alanine₁₀) showing that the linear scaling algorithms are not in their asymptotic regime. By contrast, the computational cost for Alanine₁₀ with this basis is only 160 min. with the FTC technique. A systematic study⁶⁶ has been published recently exploring the role of the J-matrix and CFMM techniques when the FTC is used. The importance of using all three methods in concert has been clearly shown. The CFMM, J-matrix engine and FTC techniques together offer the best performance for the Coulomb part of DFT calculations today, making DFT more affordable for very large systems with high-quality basis sets.

2.3 Exchange–correlation quadrature evaluation

Together with Coulomb and exchange matrix evaluation, the other part of a DFT calculation with substantial computational cost is the numerical integration of exchange–correlation (XC) terms. Most quantum chemistry codes today, including Q-Chem, employ pruned atom-centered grids of the Becke type.⁶⁷ Many years ago, the Pople group introduced

the first standard quadrature grid,⁶⁸ SG-1, for molecular density functional calculations and this has subsequently found widespread use in both the Q-Chem and Gaussian packages. However, continued improvements (CFMM, FTC, *etc*) in Coulomb technology have made it increasingly important to devise a smaller grid suitable for preliminary investigations on large systems. For this purpose, Q-Chem 3.0 contains a highly optimized new grid, SG-0,⁶⁹ which is roughly half the size and accuracy of SG-1, and roughly twice as fast. In the design of SG-0, the atomic grids for each of the elements up to chlorine (except helium and neon) were individually optimized and particular attention was paid to the elements (H, C, N and O) that are most important in organic chemistry.

To accelerate XC quadrature on these standard grids, we have devised and implemented a scheme called the incremental DFT method (IncDFT)⁷⁰ which exploits the small differences between the XC matrix on successive SCF cycles for efficiency gains. Unlike the case for the corresponding HF method,⁷¹ the XC functional is not linear with respect to the density, and our scheme is formulated in such a way that the problem of using functionals is circumvented and proper mathematical consistency is maintained. As convergence is approached in the iterative solution to the KS-DFT equations, an increasing fraction of the grid points associated with the numerical quadrature can be screened out due to the decreasing change in the density. In conjunction with a variable threshold, the IncDFT method yields up to 45% saving in the time needed for the integration procedure with negligible loss in accuracy. Further developments that make use of a uniform grid in conjunction with the atom-centered grids offer the hope of additional speedups in the future,⁷² without the loss of translational invariance normally associated with uniform grids.

2.4 Dual basis self-consistent field calculations

While self-consistent field (SCF) methods are in principle the most efficient first principles electronic structure methods, they are still expensive in large basis sets, particularly when exact exchange is required. In fact, the MP2 correlation correction (performed with the efficient methods described in the following section) can be substantially cheaper than the pre-requisite SCF calculation (see for example the calculations discussed in Section 3). One way of attacking this problem is *via* so-called dual basis methods,^{73–76} where the SCF calculation is performed in a small “primary” basis, followed by a one-step correction for basis set extension effects that is performed in a large “secondary” basis. If the primary and secondary basis sets are chosen in an appropriate manner, there is very modest loss of accuracy relative to performing a full SCF calculation in the large basis set, and the SCF calculation can be accelerated by over an order of magnitude (*i.e.* by a factor that can approach the typical number of iterations in the SCF scheme).⁷⁶ Of particular note for MP2 calculations in large basis sets is the fact that we have formulated suitable subsets of the large cc-pVTZ and cc-pVQZ basis sets^{77,78} that can be used very effectively in conjunction with the efficient resolution of the identity (RI) MP2 methods described in the following section. Together they reduce the cost of large basis

MP2 calculations such as cc-pVQZ by roughly an order of magnitude.

2.5 Linear scaling diagonalization replacements

For large molecules, the initial computational bottlenecks are associated with assembly of the effective one-particle Hamiltonian (Fock matrix) including the Coulomb and exchange–correlation terms discussed above, and the exact exchange terms that are treated *via* the linear exchange (LinK) method.⁷⁹ Given the resulting linear scaling of Fock matrix assembly, the cubic scaling diagonalization of this matrix then eventually becomes rate-determining.⁸⁰ A multitude of methods have been proposed to exploit natural sparsity^{30,81} to achieve linear scaling replacements for this diagonalization step (for reviews see ref. 82 and 83). We use a blocking strategy^{84,85} to avoid substantial overhead in manipulating sparse matrices, and have implemented several effective linear scaling approaches, including canonical purification,⁸⁶ so-called curvy steps,^{87,88} and Chebyshev-based⁸⁹ fast summation methods.^{90,91} These algorithms become effective for systems with one-dimensional connectivity that are in the 100 atom regime. Their effectiveness is least for true 3-dimensional systems, and also diminishes for large basis sets (where the dual basis strategy discussed above is a viable alternative).

3. Auxiliary basis expansions for second order perturbation theory calculations

Second order Møller-Plesset theory (MP2) is a simple wavefunction based method to correct HF theory for the effects of electron correlation.⁶ MP2 yields reliable predictions of structure, vibrational frequencies, and relative conformational energies of closed shell organic molecules,⁹² although it can perform poorly for radicals.⁹³ Relative to DFT methods, MP2 has the advantage that it naturally includes long-range dispersion interactions, which are absent in standard density functionals. For the broad class of molecules where MP2 works well, the main limitation is that traditionally MP2 calculations are significantly more expensive than the SCF calculation that precedes it. It is a particularly exciting development that this is no longer necessarily the case, thanks to the development of auxiliary basis methods.^{94,95}

Auxiliary basis expansions replace 4-center 2-electron integrals by variationally optimal⁹⁶ linear combinations of 3- and 2-center 2-electron integrals. This is often referred to as a “resolution of the identity” (RI) or “density fitting” (DF) approximation. It does not change the scaling of integral evaluation with respect to the size of the molecule, since there are still nominally a quadratic number of numerically significant integrals to evaluate and process. However it greatly reduces the quadratic pre-factor, which now only grows with the cube of the number of functions per atom, rather than the 4th power. This greatly facilitates development of efficient algorithms, and facilitates more accurate calculations using larger basis sets.

With the development of standardized MP2 auxiliary basis sets by the Karlsruhe group^{94,95} MP2 methods based on the RI approximation^{97,98} are ripe for widespread adoption in quantum chemistry. The loss of accuracy associated with using

optimized auxiliary basis expansions in MP2 calculations is very small. For example, on MP2 calculations of atomization energies for the 148 neutral molecules of the G2 data set, the RMS and maximum errors are only 0.10 and 0.27 kcal mol⁻¹, respectively, using the PVDZ basis. This is insignificant relative to the errors in MP2 itself, which are about 50 times larger.

3.1 RI-MP2 energies and gradients

We have developed efficient algorithms for RI-MP2 energies and gradients—the latter differing significantly from the original literature^{94,98} to eliminate data transfer bottlenecks. In addition to restricted and unrestricted calculations, restricted open shell MP2^{99,100} is also supported. Some sample timings are shown in Table 3 for conventional and RI-MP2 energy evaluation with the cc-pVDZ basis, and in Table 4 for gradient evaluation with the 6-31G** basis. Q-Chem exploits the shared s and p exponents of the 6-31G** basis to perform substantially faster 4 center 2-electron integral and integral derivative evaluation than is possible in basis sets like cc-pVDZ which lack this constraint. Comparing conventional MP2 and RI-MP2 shows that the RI approximation yields very substantial speedups. For energy calculations (Table 3) on systems in the 20 heavy atom regime (tetrapeptides), the RI-MP2 correlation treatment is inexpensive relative to the preceding HF calculation, in contrast to the situation with conventional algorithms. For gradient calculations, the RI-MP2 gradient code is about twice as fast, overall, as a well-optimized conventional MP2 gradient code¹⁰¹ for a basis of this modest size. Larger speedups are obtained with larger basis sets, due to the n^3 scaling with number of basis functions per atom, n , as opposed to n^4 for conventional MP2. The overall RI-MP2 gradient cost is less than 3 times the cost of an SCF gradient for the same molecule in the 20 heavy atom regime, by contrast with the conventional MP2 gradient evaluation which is several times more computationally expensive. As molecule size grows into the 40 heavy atom regime (as exemplified in the tables by octapeptides) and then the 80 heavy atom (hexadecapeptide) regime, the RI-MP2 steps that are 5th order in molecule size eventually dominate the HF steps and one must resort to lower scaling alternatives to MP2

Table 3 CPU timings (in seconds on a 2 GHz AMD Opteron processor) for energy evaluation with the cc-pVDZ basis on two conformations (one globular and thus 3-D, and one that is extended and thus 1-D) of an alanine tetrapeptide (390 basis functions), an alanine octapeptide (770 basis functions), and a hexadecapeptide (1530 basis functions). The frozen core approximation was used in all post-HF calculations. A threshold of 10^{-12} was used for integral evaluation, the SCF convergence criterion was 10^{-8} , and no symmetry was employed. The correlation energy calculations do not include the HF timing, so that the relative size of the two can be compared

	HF	MP2	RI-MP2	TRIM	RI-TRIM	SOS-MP2
Tetra3D	2340	3087	245	2694	211	282
Octa3D	15729	47480	5392	34510	2486	3394
Hexadeca3D	71178	—	145600	373900	33320	48630
Tetra1D	1735	2143	248	2041	212	280
Octa1D	7592	24240	5171	16920	2305	3290
Hexadeca1D	41355	—	143100	194900	31880	47040

theory in order to keep computational costs manageable, as discussed in the following sections.

3.2 Triatomics-in-molecules local correlation (TRIM-MP2) energies

Local correlation methods^{102,103} introduce spatial cutoffs to reduce the scaling of computational cost with molecular size. The TRIM model¹⁰⁴ is one such approach, which uses (redundant) atom-labeled functions¹⁰⁵ to span the occupied and virtual sub-spaces, and then retains only excitations where at least one electron fluctuates from an occupied to virtual orbital on the same atom. A new RI-based algorithm for the TRIM-MP2 energy¹⁰⁶ greatly accelerates these local correlation calculations (*e.g.* compare TRIM and RI-TRIM in Table 3 above), which scale with the 4th power of molecule size. In addition, relative to the parent RI-MP2 method itself, RI-TRIM-MP2 accelerates the electron correlation part of the calculation by a factor proportional to the number of atoms in the molecule, regardless of the shape of the molecule. At the 80 heavy atom level (*e.g.* the hexadecapeptide), for instance, the speedup is approximately a factor of 4, as is evident in Table 3. The TRIM model can also be applied to the scaled opposite spin models discussed below. A principal advantage of TRIM relative to other local correlation models such as the seminal Pulay–Saebø model^{102,103} is the fact that potential energy surfaces in the TRIM model are continuous.

3.3 4th order scaled opposite spin (SOS-MP2) energies and gradients

In a statistical sense, the accuracy of relative energies from MP2 calculations can be significantly improved by semi-empirically scaling the opposite-spin and same-spin correlation components with separate scaling factors, as shown by Grimme.¹⁰⁷ Results of similar quality can be obtained by just scaling the opposite spin correlation (by 1.3),¹⁰⁸ and completely omitting same-spin correlation, which is the scaled opposite spin MP2 (SOS-MP2) method. Furthermore the SOS-MP2 energy can be evaluated using the RI approximation together with a Laplace transform technique,¹⁰⁹ in an effort that scales only with the 4th power of molecular size, as opposed to 5th order for MP2 or RI-MP2. Efficient algorithms for the SOS-MP2 energy¹⁰⁸ and analytical gradient are available in Q-Chem 3.0, and offer advantages in speed over MP2 for larger molecules, as well as statistically significant improvements in accuracy. Finally, the SOS-MP2 method does systematically underestimate long-range dispersion (for which the appropriate scaling factor is 2 rather than 1.3) but this can be accounted for by making the scaling factor distance-dependent, which is done in the modified opposite spin variant (MOS-MP2) that has recently been proposed and tested.¹¹⁰ The MOS-MP2 energy and analytical gradient are also available in Q-Chem 3.0 at a cost that is essentially identical with SOS-MP2. Indeed we have chosen to use SOS-MP2 timings in Table 3 above for energies, and MOS-MP2 timings in Table 4 above for gradients. These timings show that the 4th order implementation of SOS-MP2 and MOS-MP2 yields substantial speedups over RI-MP2 for molecules in the 40 heavy atom regime and larger. It is also possible to customize the scale

Table 4 Total CPU timings (seconds on a 2GHz AMD Opteron processor) for force evaluation with the 6-31G** basis (using the on two conformations (one globular and thus 3-D, and one that is extended and thus 1-D) of an alanine tetrapeptide, octapeptide, and hexadecapeptide. All tolerances and settings were the same as given in Table 3. These timings include *all* steps necessary to evaluate the force

	HF	MP2	RI-MP2	MOS-MP2
Tetra3D	1704	9160	4470	4141
Octa3D	9596	121473	46953	30853
Hexadeca3D	46 305	—	781151	227251
Tetra1D	1253	6559	3400	3046
Octa1D	5672	68989	34389	18689
Hexadeca1D	29098	—	721879	164979

factors for particular applications, such as weak interactions, if required.

3.4 Lower scaling SOS-MP2 energies

The 4th order scaling SOS-MP2 energy evaluation can be evaluated in reduced cost for larger systems by exploiting natural sparsity,^{111,112} *via* a newly developed local algorithm. Within the RI formulation, this algorithm separates intrinsically long-range terms, which are contained in 2-center integrals, from intrinsically short-range terms, which comprise 3-center expansion coefficients. The former are treated exactly without cutoffs, while sparsity is exploited in the latter, so that the results can reproduce the original theory to a specified accuracy. Scaling is reduced to linear in the rate-determining step (which is short-range) while other (formerly smaller) terms are still quadratic or cubic. As an example of the speedups that are achievable, while retaining microHartree accuracy, we summarize calculations on two (one extended and the other globular) conformations of hexadeca-alanine peptides, $\text{CH}_3\text{CO}[\text{NHCHCH}_3\text{CO}]_{15}\text{NHCH}_3$ using the VDZ(d) basis set. For the linear conformer, the local SOS-MP2 code (LSOS-MP2) shows a 3.3 times speed-up in CPU time (2.3 speedup in elapsed time) compared to SOS-MP2. For the globular conformer, LSOS-MP2 yields a speedup of 1.5 times for CPU time (1.1 times elapsed). This is expected as the onset of the linear scaling electronic methods for dense 3-D systems usually occurs at quite large molecules. Nonetheless, the fact that LSOS-MP2 yields sizable CPU savings for a globular conformer with only 80 heavy atoms is encouraging.

4. Coupled cluster methods for ground and excited states

Coupled cluster (CC) methods^{113,114} are amongst the most powerful approaches available for treating electron correlations, and thus systematically improve upon lower level wavefunction methods such as Hartree–Fock and MP2 theory, as well as present-day DFT methods. Of course this power comes at a much increased computational price—the scaling of CC calculations with system size starts at 6th order for methods that include single and double excitations, and is 7th order or higher if triples are considered.

Q-Chem's CC capabilities include the standard singles and doubles CC method (CCSD)¹¹⁵ plus non-iterative triples

corrections *via* the CCSD(T)¹¹⁶ and CCSD(2)¹¹⁷ approaches. Great flexibility is possible in the choice of reference orbitals—including the ability to use Kohn–Sham orbitals¹¹⁸ which can be very useful when the HF orbitals are poor. The energy and gradient can also be evaluated with optimized orbital (OO) approaches including optimized doubles (OD)¹¹⁹ and quadratic CCD (QCCD)^{120,121}. An excited state treatment based on OD is available,¹²² as are (T) and (2) corrections.¹²³ OD and QCCD can also be evaluated within active spaces, defining the VOD¹²⁴ and VQCCD¹²¹ methods, for which (2) corrections are available.^{125,126} Additionally, the analytical gradient of the CCSD energy is available.

Equation-of-motion (EOM) CCSD methods expand the range of chemical problems that can be treated by CC methods. Specifically, EOM-CC methods enable accurate calculations of electronically excited states (EOM-EE-CCSD);¹²⁷ ground and excited states of diradicals and triradicals *via* the spin-flip (SF) approach (EOM-SF-CCSD);^{128,129} and ionization potentials and electron attachment energies as well as robust treatment of problematic doublet radicals (EOM-IP/EA-CCSD,^{130–132} and EOM-DIP-CCSD¹³³). With both closed and open-shell references (RHF/UHF/ROHF), frozen core/virtual options, full use of molecular point group symmetry, analytic gradients for RHF and UHF references,¹³⁴ and properties calculations (permanent and transition dipole moments, $\langle S^2 \rangle$, $\langle R^2 \rangle$, *etc*), we believe this is the most complete EOM-CCSD capability available.

Conceptually, EOM is very similar to configuration interaction (CI): target EOM states are found by diagonalizing the so-called similarity transformed Hamiltonian; $\bar{H} = e^{-T} H e^T$ to obtain eigenvalues, E , and eigenvectors, R , according to $\bar{H}R = ER$. In the EOM-CCSD models, T and R are truncated at single and double excitations, and the amplitudes T satisfy the usual CCSD equations for a reference $|\Phi_0\rangle$, which are solved first. The computational scaling of EOM-CCSD and CISD methods is identical, *i.e.*, N^6 , however EOM-CCSD is numerically superior to CISD because correlation effects are “folded into” the transformed Hamiltonian, and because EOM-CCSD is rigorously size-extensive. With 6th order scaling, our current implementation of the EOM-XX-CCSD methods enables calculations on medium-size molecules, *e.g.*, up to 10–14 heavy atoms.

By combining different types of excitation operators and references $|\Phi_0\rangle$, different groups of target states can be accessed. For example, electronically excited states can be described when the reference $|\Phi_0\rangle$ corresponds to the ground state, and the operators R conserve the number of electrons and total spin.¹²⁷ In the ionized/electron attached EOM models,^{130,131} the operators R are not electron conserving—from a closed shell reference these models can accurately treat ground and excited states of doublet radicals and some other open-shell systems, whose theoretical treatment is often plagued by symmetry breaking. Finally, the EOM-SF method^{128,129} in which the excitation operators include a spin-flip, allows one to access diradicals (from a triplet reference), triradicals (from a quartet reference), and bond-breaking.

There is a growing number of applications that employ the EOM-SF-CCSD with success for state-of-the-art studies of diradicals,^{135–138} and also triradicals,^{136,139} illustrating the

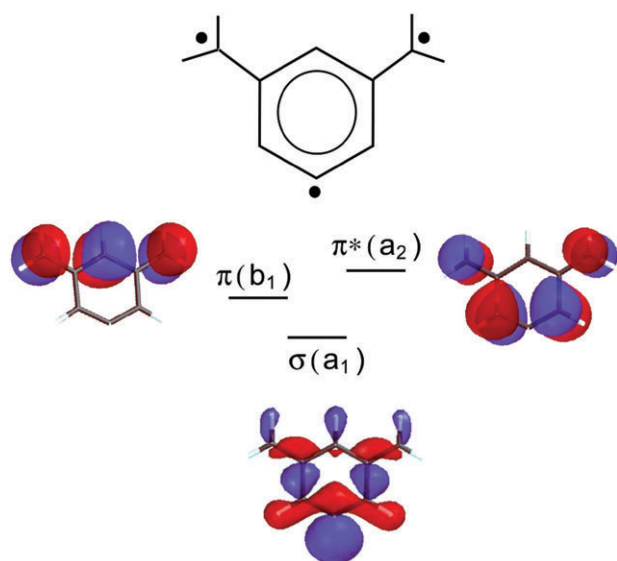


Fig. 1 The structure and key molecular orbitals of the DMX tri-radical.¹⁴⁰

increased range of systems that be treated relative to CCSD for ground states. One particularly intriguing example is the so-called DMX triradical,¹⁴⁰ which is shown in Fig. 1. The low-lying states of DMX are summarized in Fig. 2, which shows that contrary to Hund's rules, this triradical has an open shell doublet ground state.

In some applications, restricting the excitation operator to singles and doubles is insufficient, and to treat such cases, Q-Chem includes the EOM-CC(2,3) approach,^{141,142} where R is instead truncated at triple excitations. The reference state excitation operator T , however, is still taken from CCSD. The

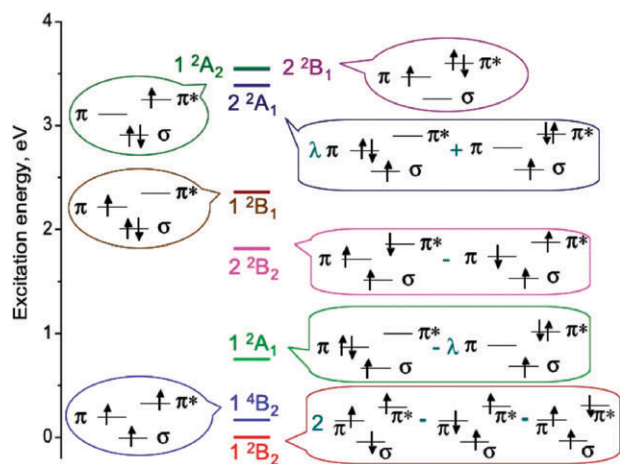


Fig. 2 Low lying electronic states of DMX as calculated by the EOM-SF-CCSD method. Note that the closed-shell doublet (1^2B_2 state) in which electrons are distributed in accordance with Aufbau principle is 2.5 eV above the ground state. Another likely candidate for the ground state, the quartet 1^4B_2 state, is 0.16 eV above the open-shell doublet, thus violating Hund's rule. The multistate nature and simultaneous treatment of dynamical and nondynamical correlation by EOM-SF-CCSD were instrumental in characterizing electronic structure of this triradical.

inclusion of triple excitations is necessary for achieving chemical accuracy (1 kcal mol^{-1}) for ground state properties. Triples are even more important for excited states. In particular, triple excitations are crucial for doubly excited states,¹⁴¹ excited states of some radicals and SF calculations (diradicals, triradicals, bond-breaking) when the open-shell reference state is heavily spin-contaminated. The accuracy of the EOM-CC(2,3) method closely follows full EOM-CCSDT [which can also be called EOM-CC(3,3)], but its computational cost is less, though it still scales as N^8 , and so is limited to relatively small systems. For medium size molecules, EOM-CC(2,3) calculations can be performed by using the active space variant of the method, or within energy additivity schemes where the effects of triple excitations are evaluated in a small basis set.

5. Simplified methods for strong electron correlations

5.1 Singlet strongly orthogonal geminal (SSG) methods

Computational models that use a single reference wavefunction describe molecules in terms of independent electrons interacting *via* averaged Coulomb and exchange fields. It is natural to improve this description by using correlated electron pairs, or geminals,¹⁴³ as building blocks of molecular wavefunctions. In the singlet strongly orthogonal geminal (SSG) model,¹⁴⁴ the wavefunction is taken as an antisymmetrized product of different geminals, one for each electron pair. Strong orthogonality means that each orthogonal molecular orbital participates in only one geminal, and the number of molecular orbitals contributing to each geminal is an adjustable parameter chosen to minimize the total energy. Open shell orbitals remain uncorrelated. Both spin-restricted (RSSG) and spin-unrestricted (USSG) versions are available. A Mulliken population analysis is performed for each geminal, which enables assignment of geminals as core electron pairs, chemical bonds, and lone electron pairs.

The presence of any but the leading geminal coefficient with a large absolute value (for instance, 0.1 is often used for the definition of "large") indicates the importance of electron correlation in that geminal, and thus in the molecule. As an example, consider a calculation on ScH with the 6-31G basis set at the experimental bond distance of 1.776 Å. In its singlet ground state the molecule has 11 geminals (22 electrons). Nine of them form core electrons on Sc. The second highest geminal (10) has $E = -1.342609 \text{ a.u.}$ and its 3 leading geminal expansion coefficients are 0.991, -0.126 and -0.036 . Mulliken population analysis shows that geminal 10 is delocalized between Sc and H, indicating a bond. It is moderately correlated, as indicated by its second expansion coefficient of magnitude 0.126. Geminal 11 has $E = -0.757086 \text{ a.u.}$ and its 5 leading expansion coefficients are 0.961, -0.174 , -0.169 , -0.124 , -0.032 . It is localized on Sc, and represents the pair of 4s electrons and their partial excitation into 3d orbitals. The presence of three large secondary expansion coefficients shows that this electron correlation effect indeed involves an expanded valence shell—something that cannot be described within active spaces that provide one correlating orbital per

nominally occupied orbital, such as the perfect pairing model¹⁴⁵ discussed below.

The description of molecular electronic structure can be improved by a perturbative description of the missing intergeminal correlation effects, taking the SSG model as the reference. We have implemented an Epstein–Nesbet (EN) form of perturbation theory¹⁴⁶ that permits a balanced description of one- and two-electron contributions to excitation energies from the SSG reference. This form of perturbation theory is especially accurate for calculation of weak intermolecular forces. Carried to second order, this defines the SSG(EN2) model. All perturbative contributions to the SSG(EN2) energy (second-order Epstein–Nesbet perturbation theory from the SSG wavefunction) are analyzed in terms of the largest numerators, smallest denominators, and total energy contributions by the type of excitation to permit a careful assessment of the quality of SSG reference wavefunction. Formally, the SSG(EN2) correction can be applied both to RSSG and USSG wavefunctions. Experience shows that molecules with broken or nearly broken bonds may have divergent RSSG(EN2) corrections. By contrast, USSG(EN2) theory seems to be well balanced, with the largest perturbative corrections to the wavefunction rarely exceeding 0.1 in magnitude. It is therefore the recommended form.

5.2 Coupled cluster perfect pairing methods

Closely related to the geminal model discussed above, but with a slightly different set of strengths and weaknesses, are coupled cluster perfect pairing methods. These are simple approximations to the Schrödinger equation in the valence active space composed of 2 orbitals to represent each electron pair, one nominally occupied (bonding or lone pair in character) and the other nominally empty, or correlating (it is typically antibonding in character). This is often called the perfect pairing active space,¹⁴⁵ and it is chemically appealing because it allows each individual bond to be correlated with the single antibonding orbital necessary for it to break correctly. The orbitals are to be optimized to minimize the energy, in addition to solving for the coefficients of all excited configurations that contribute to the wavefunction.

The exact quantum chemistry in this (or any other) active space is given by a complete active space SCF (CASSCF) calculation,³⁶ whose exponential cost growth with molecule size makes it prohibitive for systems with more than about 14 active orbitals. One well-defined coupled cluster approximation to CASSCF is to include only double substitutions in the valence space,¹⁴⁷ whose orbitals are optimized. In the framework of conventional CC theory, this defines the valence optimized doubles (VOD) model,¹²⁴ which scales as N^6 . This method is available in Q-Chem but its 6th order cost makes it prohibitive to apply to large systems. Fortunately it is possible to further approximate VOD (in the perfect pairing space) quite faithfully at dramatically reduced computational cost.

This can be accomplished by using the coupled cluster perfect pairing (PP),^{148–150} imperfect pairing (IP)^{149,151} and restricted pairing (RP)¹⁵² models. These methods are local approximations to VOD that include only a linear and quadratic number of double substitution amplitudes respectively.

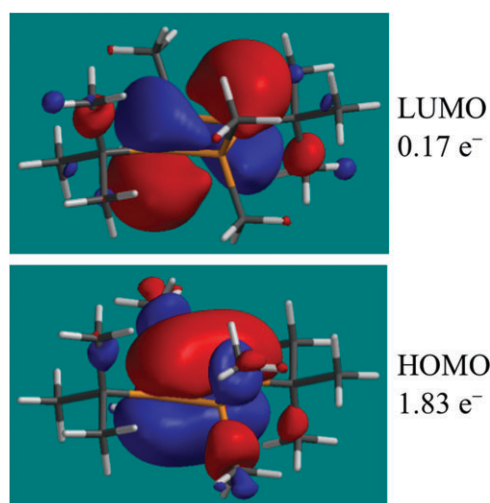


Fig. 3 Diradical character in a recently synthesized¹⁵⁹ 4-membered BPBP ring system. The lower panel shows the highest occupied orbital (HOMO), which has an occupation number of 1.83 electrons based on PP calculations correlating all 94 valence electrons.^{155,156} It is nominally bonding (through space) between the 2 boron atoms, but also exhibits significant (through bond) neighboring group interactions. The upper panel shows the corresponding lowest unoccupied orbital (LUMO), which is calculated to have occupation number 0.17, corresponding to 17% diradicaloid character.

We have implemented PP, IP and RP with a new resolution of the identity (RI) algorithm that makes them computationally very efficient. The PP model is available with restricted and unrestricted orbitals,¹⁵⁰ while the IP and RP models are at present only available with restricted orbitals. They can be applied to systems with more than 100 active electrons, and both energies and analytical gradients are available. This opens up chemical applications on molecules that are many times larger than could possibly be treated by any other active space approach. While still more expensive than DFT or HF, these methods are qualitatively correct for many highly correlated systems that SCF cannot treat reliably.

One such application is the study of singlet molecules that may have diradicaloid character, the extent of which can be quantified by calculating the occupation number of the most strongly occupied anti-bonding orbital. In an idealized singlet diradical this would be one, whereas in an idealized closed shell molecule, it would be zero, thus defining a convenient (though not unique^{153,154}) scale (0 to 100%) on which to compare different molecules. A number of interesting studies of this type have been reported recently^{155–158} using Q-Chem's new implementation. An example of PP calculations on a diradical of experimental interest¹⁵⁹ is shown in Fig. 3, which establishes a diradical character of 17%, which, while far from an idealized singlet diradical (100%), is still more than twice as high as typical stable closed shell molecules. This type of information cannot be extracted from SCF calculations.

The PP and IP models are potential replacements for HF theory as a zero order description of electronic structure and can be used as a starting point for perturbation theory.¹²⁵ The leading (second order) correction to the PP model, termed PP(2), has been formulated and efficiently implemented for restricted and unrestricted orbitals (energy only). PP(2)

improves upon many of the worst failures of MP2 theory (to which it is analogous), such as for open shell radicals. PP(2) is implemented using a resolution of the identity (RI) approach to keep the computational cost manageable. This cost scales in the same 5th order way with molecular size as RI-MP2, but with a pre-factor that is about 5 times larger.

These simplified active space coupled cluster methods are powerful and open up classes of applications that are either not computationally feasible or not sufficiently reliable with existing theoretical models. However, they should not be applied blindly without an awareness of the limitations associated with the approximations involved (limited number of doubles amplitudes in a limited perfect pairing active space). One weakness of the PP and IP models (and to some extent also the SSG model discussed above) is occasional symmetry-breaking artifacts that are a consequence of the limited number of retained pair correlation amplitudes. For example, benzene in the PP approximation prefers D_{3h} symmetry over D_{6h} by 3 kcal mol⁻¹ (with a 2° distortion), while in IP, this difference is reduced to 0.5 kcal mol⁻¹ and less than 1°. ¹⁵¹ Likewise, the allyl radical breaks symmetry in the PP model, ¹⁵⁰ although to a lesser extent than in restricted open shell HF. Another weakness is the limitation to the perfect pairing active space, which is not necessarily appropriate for molecules with expanded valence shells, such as the ScH example discussed in the previous section, although it works well for most organic molecules.

6. Composite model chemistries

Well-defined theoretical model chemistries such as those discussed in the preceding 4 sections generally involve only two approximations—a choice of atomic orbital basis set, and a choice of electron correlation treatment. If one agrees to abandon general models and make them instead system-specific, it is also possible to greatly add to the flexibility of the theoretical model by permitting different parts of the system to be described differently. This could be as simple as using different basis functions on atoms of the same type according to their location, or it could be more drastic in order to make modeling of much larger systems tractable. We discuss Q-Chem 3.0's new capabilities of this type below.

6.1 Hybrid quantum mechanics—molecular mechanics (QM/MM) methods

Many interesting systems, particularly in the condensed phase, are still well beyond the capacity of rigorous quantum chemistry methods, even with the new progress described in previous sections. To provide a tool for studying such systems, the popular QM/MM strategy was developed by Warshel and Levitt, ¹⁶⁰ Singh and Kollman, ¹⁶¹ Bash *et al.*, ¹⁶² Maseras and Morokuma, ¹⁶³ and many others (see ref. 164–166 for reviews).

Within QM/MM methods, a system of interest is divided into a central “core” region, which is going to be treated with rigorous quantum chemistry methods, and its surrounding “environment”, which is going to be treated with molecular mechanics (MM). In Q-Chem 3.0, there are four different types of QM/MM interfaces available, with others under development.

(1) A fixed point charge model. If the “environment” is rather rigid, then we can approximate it as a collection of point charges at fixed positions. The QM “core” region is subjected to the electrostatic potential and the Lennard-Jones potential from the “environment” atoms. ¹⁶⁷ In this model, ¹⁶⁸ we adopted the charge values and the Lennard-Jones parameters from the AMBER force field, ¹⁶⁹ and re-optimized the Lennard-Jones parameters for the QM atoms.

(2) A two-layer (ONIOM) model. The QM “core” region is saturated with a hydrogen atom for each covalent bond at the interface. The total energy of the system is defined additively as ¹⁶³

$$E_{\text{QM/MM}} = E(\text{core})_{\text{QM}} - E(\text{core})_{\text{MM}} + E(\text{core} + \text{environment})_{\text{MM}}$$

Q-Chem 3.0 has the ability to use its geometry optimization capabilities (discussed in the following section) with the hybrid energy defined in this way.

(3) The MOLARIS simulation package. ¹⁷⁰ Q-Chem 3.0 includes the MOLARIS package which has an extensive range of condensed phase QM/MM dynamical simulation capabilities. These include calculations of solvation energies, ¹⁷¹ calculations of electronic spectra in solutions, and, with application of proper constraints, calculations of chemical reactions in solutions. It is also possible to perform QM/MM calculations of chemical reactions in enzyme active sites, ¹⁷² if the system has been appropriately defined according to the MOLARIS manual.

(4) A Q-Chem/CHARMM interface. This new software interface fully supports Hartree-Fock (HF), Density Functional Theory (DFT), and post-HF methods (MP2, RI-MP2, CCSD) for the QM region. The standard single-link atom approach (SLA) and the exclusion of QM/MM electrostatic interactions of the MM host group (EXGR) ¹⁷³ are supported. The current interface communicates between CHARMM and Q-Chem *via* file based information sharing, which allows each program to maintain its independence—the CHARMM program must be obtained separately.

6.2 Continuum solvation models

Continuum solvation ^{174,175} models the molecular solute in quantum mechanical detail, while the surrounding solvent is treated as a classical polarizable continuum. We have implemented both the simple spherical cavity Onsager reaction field model and the more sophisticated Langevin dipoles model ¹⁷⁶ of aqueous solvation that naturally includes dielectric saturation effects. The spherical cavity model has been extended to include analytical SCF gradients, as well as to include higher order multipoles, *via* the Kirkwood treatment, and also to treat solvent with dissolved salts, *via* the Debye-Hückel approach. In addition, Q-Chem 3.0 also contains an additional polarizable continuum solvation model developed by Chipman. ^{177–179} This model defines the dielectric cavity as an isodensity contour, and solves the surface and simulation of volume polarization for electrostatics (SS(V)PE) equations that take careful account of electrostatic effects associated with solute charge outside the cavity. This model is available

for self-consistent reaction field energy evaluation with HF and DFT calculations.

7. Exploring potential energy surfaces

Theoretical chemical models such as the types discussed in the preceding sections yield an implicit representation of the electronic potential energy surface, as a function of the nuclear positions. With the addition of analytic gradients with respect to nuclear positions (available for many methods) and, when necessary, analytical second derivatives (available only for SCF methods, and single excitation CI for excited states¹⁸⁰), it is possible to “walk” on these surfaces to find either local minima, saddle points, or reaction paths, and an extensive literature has developed on this issue (*e.g.* ref. 181) since Pulay originated the ability to compute analytical gradients.¹⁸² Q-Chem incorporates Baker’s OPTIMIZE package,¹⁸³ including new enhancements¹⁸⁴ to its automatically generated internal coordinates.^{185,186} New capabilities are discussed below; others are under development.^{187,188}

7.1 Transition structure searching

Once a good approximation to the minimum energy pathway is obtained, *e.g.* with the help of an interpolation algorithm such as the growing string method, local surface walking algorithms can be used to determine the exact location of the saddle point. Baker’s¹⁸³ partitioned rational function optimization (P-RFO) method,¹⁸⁹ which utilizes an approximate or exact Hessian, has proven to be a very powerful method for this purpose. Our implementation has been made more flexible by the ability to compute analytical frequencies in DFT and HF optimizations at prescribed intervals. An alternative approach is the dimer method,¹⁹⁰ which is a mode following algorithm that utilizes only the curvature along one direction in configuration space (rather than the full Hessian) and requires only gradient evaluations. It is especially applicable for large systems where a full Hessian calculation is very time consuming, or for saddle point searches where the eigenvector of the lowest Hessian eigenvalue of the starting configuration does not correspond to the reaction coordinate. A recent modification of this method has been developed^{191,192} to significantly reduce the influence of numerical noise in the forces on the performance of the dimer algorithm, and to significantly reduce its computational cost.

7.2 Reaction path finding

One of the difficulties associated with a saddle point search is to obtain a good initial guess for the starting configuration that can later be used for a local surface walking algorithm. This difficulty becomes especially relevant for large systems, where the search space dimensionality is high and the initial starting configuration is often far away from the final saddle point. A very promising method for finding a good guess for the saddle point configuration and the minimum energy pathway connecting reactant and product states are interpolation algorithms. For example, the nudged elastic band method,^{193,194} and the string method¹⁹⁵ start from a certain initial reaction pathway connecting the reactant and the product state, and then optimize in discretized path space towards the

minimum energy pathway. The highest energy point on the resulting approximation to the minimum energy pathway becomes a good initial guess for the saddle point configuration that can subsequently be used with any local surface walking algorithm. Inevitably, the performance of an interpolation method relies heavily on the choice of the initial reaction pathway, and a poorly chosen initial pathway can cause slow convergence, or convergence to an incorrect pathway. The growing string method¹⁹⁶ offers an elegant solution to this problem, in which two string fragments (one from the reactant and the other from the product state) are grown until the two fragments join. It represents a valuable, although computationally intensive, tool for exploring challenging reaction paths and their associated transition structures.

7.3 Direct dynamics

Traditionally quantum chemistry has focused on characterizing stationary points on potential surfaces. For large systems or even not-so-large molecules where the potential surfaces are soft enough so that thermal energy allows the crossing of many barriers, this static approach becomes less useful and dynamics are required. Classical trajectories on the Born–Oppenheimer potential energy surface are the simplest useful treatment of molecular dynamics. For systems of just a few atoms, or larger molecules when no bond-making or breaking occurs, molecular dynamics can often be well-described using empirical potentials. For systems not in these categories, it is desirable to perform molecular dynamics directly from forces computed *via* quantum chemistry,^{197,198} which is usually called direct Born–Oppenheimer molecular dynamics (BOMD). We have implemented direct dynamics for self-consistent field (SCF) wavefunctions (*i.e.* density functional theory or Hartree–Fock theory), together with an efficient Fock matrix extrapolation procedure¹⁹⁹ based on the suggestion of Pulay and Fogarasi.²⁰⁰ This procedure greatly reduces the number of SCF iterations required to converge at each time step, even when moderately tight tolerances are applied, as necessary to obtain satisfactory energy conservation. In many cases only 2–4 iterations are required, which significantly improves computation time for BOMD relative to not applying extrapolation.

It is also possible to perform extended Lagrangian *ab initio* molecular dynamics (ELMD), as pioneered by Car and Parrinello.²⁰¹ This reduces the number of SCF steps per time step to 1, at the cost of requiring a generally substantially shorter time step in order to yield good approximations to the BO trajectories. The SCF wavefunction is propagated *via* a fictitious dynamics that relies on the large separation of time-scales between electronic and nuclear motions. The version we employ²⁰² is distinguished by the fact that no constraints are employed, as the variables that parameterize the fictitious electronic dynamics are the generators of unitary rotation matrices, and therefore automatically preserve orthogonality. These variables are also special in that they correspond to geodesics on this manifold, the great circle steps that are the shortest distance between any two valid points²⁰³ (a robust SCF procedure is also based on the same variables^{204,205}). The relative efficiency of BOMD and ELMD depends on the

system under consideration, and is a topic that continues to be debated in the literature.

8. Molecular properties

8.1 Linear scaling evaluation of NMR chemical shifts

The importance of nuclear magnetic resonance (NMR) spectroscopy for modern chemistry and biochemistry cannot be overestimated. Despite tremendous progress in experimental techniques, the understanding and reliable assignment of observed experimental spectra often remains a highly difficult task, so that quantum chemical methods can be extremely useful both in the solution and the solid state (*e.g.* ref. 206–210 and references therein). The cost for the computation of NMR chemical shifts within even the simplest quantum chemical methods such as Hartree–Fock (HF) or density functional theory (DFT) approximations increases conventionally with the third power of the molecular size, M , *i.e.* $O(M^3)$. Therefore, the computation of NMR chemical shieldings has so far been limited to molecular systems in the order of 100 atoms without molecular symmetry. For larger systems it is crucial to reduce the increase of the computational effort to linear, which has been recently achieved by Kussmann and Ochsenfeld.²⁰⁶ In this way, the computation of NMR chemical shifts becomes possible at both the HF and DFT levels for molecular systems with 1000 atoms and more, while the accuracy and reliability of these traditional methods is fully preserved.

In our formulation we use gauge-including atomic orbitals (GIAO),^{211,212} which have proven to be particularly successful.²¹³ For example, for many molecular systems the HF (GIAO-HF) approach provides typically an accuracy of 0.2–0.4 ppm for the computation of proton chemical shifts (*e.g.* ref. 206–210). NMR chemical shifts are calculated as second derivatives of the energy with respect to the external magnetic field \mathbf{B} and the nuclear magnetic moment m_N of a nucleus N . For the computation of the NMR shielding tensor it is necessary to solve for the response of the one-particle density matrix with respect to the magnetic field, so that the solution of the coupled perturbed SCF (CPSCF) equations either within the HF or the DFT approach is required. These equations can be solved within a density matrix-based formalism for the first time with only linear-scaling effort for molecular systems with a non-vanishing HOMO-LUMO gap.²⁰⁶ The solution is even simpler in DFT approaches without explicit exchange, since present density functionals are local and not dependent on the magnetic field. The present implementation of NMR shieldings in Q-Chem employs the LinK (linear exchange K) method^{79,214} for the formation of exchange contributions. Since the derivative of the density matrix with respect to the magnetic field is skew-symmetric, its Coulomb-type contractions vanish. For the remaining Coulomb-type matrices the CFMM method^{59,215} is adapted. In addition, a multitude of different approaches for the solution of the CPSCF equations can be selected within Q-Chem.

Perhaps the largest molecular system for which NMR shieldings have been computed to date contained 1003 atoms and 8593 basis functions (GIAO-HF/6-31G*) without molecular symmetry.²⁰⁶

8.2 Linear scaling computation of electric properties

The search for new optical devices is a major field of materials sciences. Here, polarizabilities and hyperpolarizabilities provide particularly important information on molecular systems. In order to predict optical properties for larger molecular systems, Kussmann and Ochsenfeld recently developed linear-scaling methods in order to compute the following optical properties: static and dynamical polarizabilities, first static hyperpolarizability, second harmonic generation, the electro-optical Pockels effect, and optical rectification. These optical properties can be computed for the first time using linear-scaling methods (LinK/CFMM) for all integral contractions. Although the present implementation available in Q-Chem still uses MO-based time-dependent SCF (TDSCF) equations both at the HF and DFT level, the prefactor of this $O(M^3)$ scaling step is smaller. Here, all derivatives are computed analytically. The third-order properties can be computed both using a second-order TDSCF scheme or by use of Wigner's $(2n + 1)$ rule.^{216,217}

8.3 Harmonic and anharmonic vibrational frequencies

Computing the vibrational spectrum of a molecule has become routine, where the computed spectrum is usually based on a potential energy surface (PES) that is assumed to be locally quadratic about the equilibrium geometry. To obtain an accurate harmonic spectrum at minimum computational cost, a new density functional, empirical density functional II (EDF2), was constructed.⁴⁸ This functional is a combination of existing functionals with the mixing parameters obtained by minimizing the error between the computed and experimentally obtained harmonic frequencies. This functional gives an accurate description of the harmonic potential curvature. With the cc-pVTZ basis set and ECP-containing basis sets for heavier elements, the RMS error of the harmonic frequencies for 315 molecules is 34 cm^{−1} whilst the error for the most popular density functional, B3LYP, is 38 cm^{−1}. Also, we examined thermochemical properties which are related to the depth of the potential surface. We found for the G2 dataset, EDF2 with either 6-31+G* or aug-cc-pVTZ performs better than B3LYP in most cases. For atomization energies, ionization potentials, electron affinities and proton affinities, the RMS errors for EDF2/6-31+G* are 6.17, 4.77, 4.38 and 6.38 kcal mol^{−1} respectively whilst for B3LYP/6-31+G*, they are 8.10, 5.05, 4.40, and 5.51 kcal mol^{−1}. With the larger aug-cc-pVTZ basis set, the RMS errors for EDF2 are 5.94, 4.96, 2.97 and 2.03 kcal mol^{−1}, whilst for B3LYP the errors are 6.12, 5.28, 3.80, and 1.91 kcal mol^{−1}.

The harmonic approximation gives rise to an error that should be accounted for before meaningful comparisons can be made with the anharmonic energies obtained experimentally. Anharmonic effects include Fermi/Darling-Dennison type resonances, doublet splitting, overtones and anharmonic shifts, all of which are observed in the experimental spectrum. Unfortunately, going beyond the harmonic approximation requires computing third, fourth and even higher derivatives which is extremely expensive, and most algorithms require a computational effort that scales as N_{atoms}^6 or worse. Several new methods have been implemented within Q-Chem 3.0 that

are able to compute the anharmonic correction more efficiently and more accurately than existing algorithms. Two nuclear vibration theories, vibrational configuration interaction (VCI) and vibrational perturbation theory (VPT), have been implemented.

Furthermore, a new approach that applies first-order perturbation theory with a transition-optimized shifted Hermite (TOSH) function has been implemented. This introduces a shift, σ , along the mass-weighted normal coordinate. This moves the center of the harmonic wavefunction away from the energy minimum and gives a more appropriate wavefunction for describing the distribution of nuclei on the anharmonic PES. The key difference between TOSH and other anharmonic methods that consider experimental or effective geometries is that when one applies TOSH the derivatives are still calculated at the bottom of the potential energy well, whilst other approaches calculate the derivatives at the non-stationary point. The idea of using this shifted-Hermite function may be simple, but the effect on the anharmonic correction can be dramatic. For example, most σ values are smaller than 0.7 au, but the RMS error for TOSH on 22 polyatomic molecules is 27.03 cm⁻¹, while for VPT2 it is 43.46 cm⁻¹ and for VCI (using up to four quanta) the RMS error is 30.33 cm⁻¹. Comparing TOSH with a VCI calculation using up to four quanta gives a better anharmonic frequency, furthermore, the relative timing is 500 times faster. In this new version of Q-Chem, we also implemented a new finite difference method to obtain 3rd and 4th derivatives using information from energy, gradient, and hessian (EGH), which is faster than using any single order of derivative. The speed is 1.25, 2 and 1.67 times faster than using only energies, gradients, or Hessians, respectively. Also, we examine the derivatives and compute only those that are important, which saves 54 N_{atoms}^4 for a four-mode coupling representation, and 36 N_{atoms}^2 for a two-mode coupling representations. These numerical speed-up techniques along with TOSH allow anharmonic corrections to be computed on larger systems than is otherwise possible.

8.4 Spin densities at nuclei

Gaussian basis sets violate nuclear cusp conditions.²¹⁸ This may lead to large errors in wavefunctions at nuclei, which particularly affects spin density calculations.²¹⁹ This problem can be alleviated by using an averaging operator that computes the wavefunction density based on constraints that the wavefunction must satisfy near a Coulomb singularity.^{220,221} The derivation of these operators is based on the hypervirial theorem.²²⁰ Application to molecular spin densities for spin-polarized²²¹ and DFT²²² wavefunctions show considerable improvement over the traditional delta function operator. One of the simplest forms of such operators is based on the Gaussian weight function $\exp[-(Z/r_0)^2(\mathbf{r} - \mathbf{R})^2]$ sampling the vicinity of a nucleus of charge Z at \mathbf{R} . The parameter r_0 has to be small enough to neglect two-electron contributions of the order $O(r_0^4)$ but large enough for meaningful averaging. A range of values between 0.15 to 0.3 a.u. has been shown to be adequate, with the final answer being relatively insensitive to the exact choice of r_0 .^{220,221} The averaging operators are implemented for single determinant Hartree–Fock and DFT,

and the correlated SSG wavefunctions. Spin and charge densities are printed for all nuclei in a molecule, including ghost atoms.

9. Analysis tools

9.1 Atomic charges and population analysis

Over the years, numerous models have been proposed for the extraction of atomic charges from the results of molecular orbital calculations. One of the earliest of these is the Mulliken partition but, in spite of its resilient popularity, it is one of the least theoretically satisfactory schemes failing, in general, to reproduce even the dipole moment of the electron density from which it is derived. A more satisfactory approach is the widely used natural bond order (NBO) analysis^{223,224} program, which yields charges that are more stable with respect to changes of basis than Mulliken charges, as well as the ability to analyze bonding in terms of localized orbital contributions (see also the following section). Q-Chem includes an interface to the latest version of the NBO program.

We have also proposed a new method for extracting atomic charges—multipole-derived atomic charges²²⁵—that overcomes some of the difficulties traditionally associated with deriving charges from electrostatic potentials. By construction, the multipole-derived atomic charges exactly reproduce as many as possible of the low-order multipole moments of a system. Unlike some other approaches, they are well-defined, do not require sampling of the electrostatic potential around the molecule, and can be generated with little computational effort. Nevertheless, because of the Legendre expansion, they provide an excellent model of that potential at points outside the van der Waals surface of the system.

9.2 Localized orbitals

The molecular orbitals (MOs) that emerge from a HF or DFT calculation have the property that they not only make the energy stationary but they are also the eigenvectors of the effective one-particle Hamiltonian. Occupied MOs can thus be mixed together without changing the energy. This mixing can be done to extremize measures of locality such as those proposed by Boys,²²⁶ Edmiston–Ruedenberg (ER)²²⁷ and Pipek–Mezey,²²⁸ which can be very helpful in understanding chemical bonding in large molecules. Q-Chem 3.0 contains an efficient new algorithm²²⁹ designed to find localized occupied orbitals. To make the ER orbitals computationally tractable to optimize, auxiliary basis expansions are employed to make the computation of electron-repulsion matrix elements relatively inexpensive. We also note that geminal, PP, IP and RP calculations (see the earlier section on coupled cluster perfect pairing methods) lead naturally to localized orbitals by energy minimization and those localized orbitals can also be visualized and used for interpretive purposes.

9.3 Intracules

The electron density $\rho(\mathbf{r})$ gives the probability density for finding an electron at the point \mathbf{r} and the momentum density $\pi(\mathbf{p})$ gives the analogous probability for finding an electron with momentum \mathbf{p} . Although these distributions are

important, they do not provide any information about the *relative* coordinates of electrons and therefore cannot easily illuminate intrinsically two-electron phenomena such as electron correlation. Instead, one must turn to analogous two-electron distributions,²³⁰ which are usually called intracules.

The most familiar two-electron distribution is the position intracule $P(u)$, which gives the probability density for the interelectronic coordinate $u = |\mathbf{r}_1 - \mathbf{r}_2|$. The analogous distribution in momentum space is the momentum intracule $M(v)$ which is the probability density for $v = |\mathbf{p}_1 - \mathbf{p}_2|$, the analogous interelectronic coordinate in momentum space.²³¹ A new function, the Wigner intracule $W(u, v)$, gives the joint probability density for finding two electrons separated by a distance u and moving with relative momentum v . Actually, the Wigner intracule is not a proper probability distribution because it sometimes takes negative values but, because it possesses many of the properties of a proper distribution, it is often termed a “quasi-probability”. Notwithstanding this weakness, it appears to be a promising starting point for a fundamentally new way of thinking about electron correlation,²³² which could have exciting implications for the development of new theoretical models in the future, as well as being useful for interpretive purposes at present. The position, momentum and Wigner intracules²³³ can be computed in Q-Chem 3.0, on a user-specified grid of u and/or v values.

10. Graphical user interfaces

Since this article is on advances in electronic structure methods and algorithms in the Q-Chem program, which forms the so-called computational “back-end” to a user-interface, we have chosen to discuss the “front-end” interfaces only briefly, and as the final topic. Q-Chem, and other electronic structure codes can be run as text-based programs, but this is awkward particularly for large systems. Several graphical alternatives to text-based input and output are available. In particular, Q-Chem 3.0 is seamlessly incorporated as the electronic structure back-end of the Spartan user-interface and program (see <http://www.wavefun.com> for details) beginning with the Spartan 06 version. The public domain version of WebMO (see <http://www.webmo.net> for details) is included in the Q-Chem 3.0 distribution, and we also provide scripts that support the Molden program (<http://www.cmbi.ru.nl/molden>), which permits visualization of orbitals as well as structures.

11. Conclusions

This decade continues to be a time of rapid progress for computational quantum chemistry. In this paper, we have first discussed briefly the state of the field, and our view of the main challenges for the future—the quest for optimally efficient algorithms, for reduced scaling of computational cost with molecular size, for higher accuracy and highly correlated systems, excited states, and more. The body of the paper has then described recent developments that address many aspects of these issues. These developments define the main new features of version 3.0 of the Q-Chem software package that we collaborate on (see <http://www.q-chem.com> for further information). We have described advances such as an

intelligent Coulomb matrix algorithm, dual basis methods, auxiliary basis expansions, new 4th order MP2-like methods, spin-flip coupled cluster capabilities, geminal and perfect pairing coupled cluster models, linear scaling NMR algorithms, intracule capabilities, *etc.* We discussed their performance and chemical implications, and, where appropriate, their limitations. These developments, as discussed throughout the paper, are a mixture of our own original contributions, and our implementations of what we view as compelling ideas from other groups, as will be obvious from the citations.

Stepping back from the details, we hope it is clear to the non-specialist as well as the specialist that the fundamental algorithms of electronic structure theory, and even the theoretical models that these algorithms implement are in a state of continuing flux. On the one hand this testifies to the health of electronic structure theory as a research area. It is simply not yet a mature area in the sense of numerical linear algebra, or even computational fluid dynamics, for comparison. We have too many fundamental issues that remain at best partially solved. On the other hand the continuing progress also hints at the fact that the breadth of chemical (and other) applications of these models and algorithms is expanding at a rate that is extremely rapid. We have tried to illustrate this by presenting comparative timings between our new and old versions where such comparisons are possible. We have also discussed examples of chemical applications that are made possible by the new theoretical chemical models—for instance diradicals, triradicals, very large molecules, *etc.*

This scientific progress synergizes with the continuing exponential improvements in computing platforms on which quantum chemistry calculations are performed. These exponential improvements up until recently have been driven by increasing clock speeds for serial computation. In 2000, the typical desktop computer was driven by a processor that was clocked at about 700 MHz and could perform typically one add or multiply per cycle. Now clock speeds are typically 2.5 GHz, and it is common to be able to perform 2 or even 4 floating point operations per cycle, giving at least an order of magnitude more computational capability. Looking to the future, it appears that this progress will continue at nearly the same rate but driven largely by multiple processors on a given chip rather than advances in clock speeds.

The content of this paper, and the associated program package, offers in a sense a snap-shot of the efforts of our large and loosely connected development team to push the state of the art in molecular electronic structure theory. We hope that the paper conveys a little of the excitement that we feel about progress in computational quantum chemistry, as well as some of the physical and computational ideas that underlie new capabilities, and some of the applications possibilities. We hope it also conveys (correctly) the implication that progress is on-going. In due course, we can look forward to seeing many of the developments discussed here being surpassed by future advances in our field.

Acknowledgements

Q-Chem’s technical developments reported in this paper have been partly supported by the following SBIR (Small Business

Innovation Research) grants from the NIH (National Institutes of Health): GM062053, GM064295, GM065617, GM066484, GM067335, GM069255, GM073408. The academic contributors also gratefully acknowledge partial support of this work from numerous separate grants from government research agencies including NSF, DOE, NIH, and EPSRC. Scientists who contributed only to one of our earliest releases (version 1.0, 1.1 and 1.2) do not explicitly appear as authors of this paper, but their earlier contributions still play a significant role in many of Q-Chem's capabilities, so we would also like to acknowledge their valuable past efforts.

Yihan Shao,^{ab} Laszlo Fusti Molnar,^a Yousung Jung,^{bc} Jörg Kussmann,^d Christian Ochsenfeld,^d Shawn T. Brown,^{ae} Andrew T. B. Gilbert,^f Lyudmila V. Slipchenko,^{gh} Sergey V. Levchenko,^{gj} Darragh P. O'Neill,^f Robert A. DiStasio Jr.,^b Rohini C. Lochan,^{bc} Tao Wang,^{gk} Gregory J. O. Beran,^{bci} Nicholas A. Besley,^j John M. Herbert,^b Ching Yeh Lin,^f Troy Van Voorhis,^{bck} Siu Hung Chien,^f Alex Sodt,^{bc} Ryan P. Steele,^{bc} Vitaly A. Rassolov,^l Paul E. Maslen,^m Prakashan P. Korambath,^{am} Ross D. Adamson,^a Brian Austin,^{bc} Jon Baker,^o Edward F. C. Byrd,^{bcp} Holger Dachsel,^{aq} Robert J. Doerksen,^{bcr} Andreas Dreuw,^{bcs} Barry D. Dunietz,^{bct} Anthony D. Dutoi,^{bc} Thomas R. Furlani,^u Steven R. Gwaltney,^{bcv} Andreas Heyden,^w So Hirata,^{bex} Chao-Ping Hsu,^y Gary Kedziora,^a Rustam Z. Khallilulin,^{bc} Phil Klunzinger,^z Aaron M. Lee,^A Michael S. Lee,^{bcB} WanZhen Liang,^{bC} Itay Lotan,^D Nikhil Nair,^A Baron Peters,^{Ei} Emil I. Proynov,^a Piotr A. Pieniazek,^g Young Min Rhee,^{bc} Jim Ritchie,^u Edina Rosta,^g C. David Sherrill,^{bcf} Andrew C. Simmonett,^j Joseph E. Subotnik,^{bc} H. Lee Woodcock III,^G Weimin Zhang,^a Alexis T. Bell,^E Arup K. Chakraborty,^{ikE} Daniel M. Chipman,^H Frerich J. Keil,^w Arie Warshel,^g Warren J. Hehre,^z Henry F. Schaefer III,^I Jing Kong,^a Anna I. Krylov,^g Peter M. W. Gill^f

^a Q-Chem Inc., The Design Center, Suite 690, 5001 Baum Blvd, Pittsburgh, PA 15213, USA

^b Department of Chemistry, University of California, Berkeley, CA 94720, USA

^c Chemical Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

^d Institut für Physikalische und Theoretische Chemie, Universität Tübingen, Auf der Morgenstelle 8, D-72076 Tübingen, Germany

^e Present address: Pittsburgh Supercomputing Center, 300 S. Craig Street Pittsburgh, PA 15213, USA

^f Research School of Chemistry, Australian National University, Canberra ACT 0200 Australia

^g Department of Chemistry, University of Southern California, Los Angeles, CA 90089-0482, USA

^h Present address: Department of Chemistry, Iowa State University, Ames, IA 50011, USA

ⁱ Present address: Department of Chemical Engineering, Massachusetts Institute of Technology, 77 Massachusetts Ave., Cambridge, MA 02139-4307, USA

^j School of Chemistry, University of Nottingham, University Park, Nottingham NG7 2RD, UK

^k Present address: Department of Chemistry, Massachusetts Institute of Technology, 77 Massachusetts Ave., Cambridge, MA 02139-4307, USA

^l Department of Chemistry and Biochemistry, University of South Carolina, 631 Sumter Street, Columbia, SC 29208, USA

^m Department of Chemistry, Rutgers University, 315 Penn Street, Camden, NJ 08102, USA

ⁿ Present address: Academic Technology Services, University of California, Los Angeles, Mail Code 155705, Los Angeles, CA 90095-1557, USA

^o Department of Chemistry and Biochemistry, University of Arkansas, Fayetteville, AR 72701, USA

^p Present address: AMSRD-ARL-WM-BD, Aberdeen Proving Ground, MD 21005-5069, USA

^q Present address: John von Neumann Institute for Computing, Central Institute for Applied Mathematics Research Centre Jülich, 52425 Jülich, Germany

^r Present address: Department of Medicinal Chemistry, University of Mississippi, P.O. Box 1848, University, MS 38677, USA

^s Present address: Institut für Physikalische und Theoretische Chemie Johann Wolfgang Goethe-Universität, Marie Curie-Str. 11, 60439 Frankfurt am Main, Germany

^t Chemistry Department, University of Michigan, 930 N. University, Ann Arbor, MI 48109-1055, USA

^u Center for Computational Research, SUNY University at Buffalo, Buffalo, NY 14260-1800, USA

^v Present address: Department of Chemistry, and ERC Center for Computational Sciences, Mississippi State University, Mississippi State, MS 39762, USA

^w Department of Chemical Engineering, Hamburg University of Technology, D-21073 Hamburg, Germany

^x Present address: Department of Chemistry, University of Florida, Gainesville, FL 32611, USA

^y Institute of Chemistry, Academia Sinica, Nankang, Taipei 115 Taiwan

^z Wavefunction Inc, 8401 Von Karman Avenue, Suite 370, Irvine, CA 92612 USA

^A Department of Chemistry, University of Cambridge, Lensfield Rd, Cambridge, CB2 1EW, UK

^B Present address: Department of Cell Biology and Biochemistry, US Army Medical Research Institute of Infectious Diseases, Fort Detrick, MD 21702, USA

^C Present address: Hefei National Laboratory for Physical Science at Microscale and Department of Chemical Physics, University of Science and Technology of China, Hefei, 230026, P. R. China

^D Department of Bioengineering, University of California, Berkeley, CA 94720, USA

^E Department of Chemical Engineering, University of California, Berkeley, CA 94720, USA

^F Present address: Department of Chemistry & Biochemistry, Georgia Institute of Technology, 770 State Street, Atlanta, GA 30332-0400, USA

^G Laboratory of Computational Biology, National Heart, Lung and Blood Institute, National Institutes of Health, Bethesda, MD 20892, USA

^H Radiation Laboratory University of Notre Dame, Notre Dame, IN 46556, USA

^I Center for Computational Chemistry, and Department of Chemistry, 1004 Cedar St., The University of Georgia, Athens, GA 30602-2556, USA

^J Present address: Department of Chemistry, University of Pennsylvania, 231 South 34th Street, Philadelphia, PA 19104-6323, USA

^K Present address: Shanghai Supercomputer Center, Shanghai 201203, P.R. China

References

- 1 P. A. M. Dirac, *Proc. R. Soc. London, Ser. A*, 1929, **123**, 714.
- 2 G. N. Lewis, *J. Chem. Phys.*, 1933, **1**, 17.
- 3 I. Shavitt, *Isr. J. Chem.*, 1993, **33**, 357.
- 4 L. Thogersen and J. Olsen, *Chem. Phys. Lett.*, 2004, **393**, 36.
- 5 J. A. Pople, *Rev. Mod. Phys.*, 1999, **71**, 1267.
- 6 A. Szabo and N. S. Ostlund, *Modern quantum chemistry: Introduction to advanced electronic structure theory*, Dover, New York, 1996.
- 7 F. Jensen, *Introduction to computational chemistry*, Wiley, Chichester, 1999.
- 8 M. Springborg, *Methods of electronic structure calculations*, Wiley, Chichester, 2000.
- 9 T. Helgaker, P. Jorgensen and J. Olsen, *Molecular electronic structure theory*, Wiley, Chichester, 2000.
- 10 J. Simons, *J. Phys. Chem.*, 1991, **95**, 1017.
- 11 M. Head-Gordon, *J. Phys. Chem.*, 1996, **100**, 13213.
- 12 R. A. Friesner, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 6648.
- 13 R. G. Parr and W. Yang, *Density-functional theory of atoms and molecules*, Oxford, New York, 1989.
- 14 W. Koch and M. C. Holthausen, *A Chemist's Guide to Density Functional Theory*, Wiley-VCH, Weinheim, 2001.
- 15 S. F. Boys, G. B. Cook, C. M. Reeves and I. Shavitt, *Nature*, 1956, **178**, 1207.
- 16 R. Ahlrichs, M. Bar, M. Haser, H. Horn and C. Kolmel, *Chem. Phys. Lett.*, 1989, **162**, 165.
- 17 M. W. Schmidt, K. K. Baldridge, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. J. Su, T. L. Windus, M. Dupuis and J. A. Montgomery, *J. Comput. Chem.*, 1993, **14**, 1347.
- 18 R. A. Kendall, E. Apra, D. E. Bernholdt, E. J. Bylaska, M. Dupuis, G. I. Fann, R. J. Harrison, J. L. Ju, J. A. Nichols, J. Nieplocha, T. P. Straatsma, T. L. Windus and A. T. Wong, *Comput. Phys. Commun.*, 2000, **128**, 260.
- 19 M. F. Guest, I. J. Bush, H. J. J. Van Dam, P. Sherwood, J. M. H. Thomas, J. H. Van Lenthe, R. W. A. Havenith and J. Kendrick, *Mol. Phys.*, 2005, **103**, 719.
- 20 J. Kong, C. A. White, A. I. Krylov, D. Sherrill, R. D. Adamson, T. R. Furlani, M. S. Lee, A. M. Lee, S. R. Gwaltney, T. R. Adams, C. Ochsenfeld, A. T. B. Gilbert, G. S. Kedziora, V. A. Rassolov, D. R. Maurice, N. Nair, Y. H. Shao, N. A. Besley, P. E. Maslen, J. P. Dombroski, H. Daschel, W. M. Zhang, P. P. Korambath, J. Baker, E. F. C. Byrd, T. Van Voorhis, M. Oumi, S. Hirata, C. P. Hsu, N. Ishikawa, J. Florian, A. Warshel, B. G. Johnson, P. M. W. Gill, M. Head-Gordon and J. A. Pople, *J. Comput. Chem.*, 2000, **21**, 1532.
- 21 T. R. Furlani, J. Kong and P. M. W. Gill, *Comput. Phys. Commun.*, 2000, **128**, 170.
- 22 A. D. Buckingham, *J. Phys. Chem.*, 1990, **94**, 5431.
- 23 G. Segal, *J. Phys. Chem.*, 1990, **94**, 5436.
- 24 L. Radom, *J. Phys. Chem.*, 1990, **94**, 5439.
- 25 L. Radom, *Nature*, 2004, **428**, 816.
- 26 P. M. W. Gill, M. S. Gordon, M. Head-Gordon and L. Radom, *J. Chem. Phys.*, 2004, **120**, 9445.
- 27 J. W. Demmel, *Applied numerical linear algebra*, SIAM, Philadelphia, 1997.
- 28 M. Haser and R. Ahlrichs, *J. Comput. Chem.*, 1989, **10**, 104.
- 29 W. Kohn, *Phys. Rev. Lett.*, 1996, **76**, 3168.
- 30 P. E. Maslen, C. Ochsenfeld, C. A. White, M. S. Lee and M. Head-Gordon, *J. Phys. Chem. A*, 1998, **102**, 2215.
- 31 L. A. Curtiss, K. Raghavachari, P. C. Redfern, V. Rassolov and J. A. Pople, *J. Chem. Phys.*, 1998, **109**, 7764.
- 32 A. D. Boese, M. Oren, O. Atasoylu, J. M. L. Martin, M. Kallay and J. Gauss, *J. Chem. Phys.*, 2004, **120**, 4129.
- 33 A. Tajti, P. G. Szalay, A. G. Csaszar, M. Kallay, J. Gauss, E. F. Valeev, B. A. Flowers, J. Vazquez and J. F. Stanton, *J. Chem. Phys.*, 2004, **121**, 11599.
- 34 W. M. C. Foulkes, L. Mitás, R. J. Needs and G. Rajagopal, *Rev. Mod. Phys.*, 2001, **73**, 33.
- 35 W. Klopper and C. C. M. Samson, *J. Chem. Phys.*, 2002, **116**, 6397.
- 36 P. E. M. Siegbahn, J. Almlöf, A. Heiberg and B. O. Roos, *J. Chem. Phys.*, 1981, **74**, 2384.
- 37 G. K. L. Chan and M. Head-Gordon, *J. Chem. Phys.*, 2002, **116**, 4462.
- 38 U. Schollwock, *Rev. Mod. Phys.*, 2005, **77**, 259.
- 39 K. Hirao, *Recent Advances in Computational Chemistry, Vol. 4, Recent Advances in Multireference Methods*, World Scientific, Singapore, 1999.
- 40 K. Burke, J. Werschnik and E. K. U. Gross, *J. Chem. Phys.*, 2005, **123**, 062 206.
- 41 A. Dreuw and M. Head-Gordon, *Chem. Rev.*, 2005, **105**, 4009.
- 42 H. M. Quiney, H. Skaane and I. P. Grant, *Adv. Quantum Chem.*, 1999, **32**, 1.
- 43 T. Nakajima and K. Hirao, *Monatsh. Chem.*, 2005, **136**, 965.
- 44 W. Kohn, A. D. Becke and R. G. Parr, *J. Phys. Chem.*, 1996, **100**, 12974.
- 45 A. D. Becke, *J. Chem. Phys.*, 1993, **98**, 5648.
- 46 R. H. Hertwig and W. Koch, *Chem. Phys. Lett.*, 1997, **268**, 345.
- 47 A. D. Boese and J. M. L. Martin, *J. Chem. Phys.*, 2004, **121**, 3405.
- 48 C. Y. Lin, M. W. George and P. M. W. Gill, *Aust. J. Chem.*, 2004, **57**, 365.
- 49 J. Grafenstein, E. Kraka and D. Cremer, *J. Chem. Phys.*, 2004, **120**, 524.
- 50 S. Kristyan and P. Pulay, *Chem. Phys. Lett.*, 1994, **229**, 175.
- 51 M. Head-Gordon and J. A. Pople, *J. Chem. Phys.*, 1988, **89**, 5777.
- 52 T. R. Adams, R. D. Adamson and P. M. W. Gill, *J. Chem. Phys.*, 1997, **107**, 124.
- 53 M. Dupuis, J. Rys and H. F. King, *J. Chem. Phys.*, 1976, **65**, 111.
- 54 J. Rys, M. Dupuis and H. F. King, *J. Comput. Chem.*, 1983, **4**, 154.
- 55 P. M. W. Gill, *Adv. Quantum Chem.*, 1994, **25**, 141.
- 56 C. A. White, B. G. Johnson, P. M. W. Gill and M. Head-Gordon, *Chem. Phys. Lett.*, 1994, **230**, 8.
- 57 C. A. White and M. Head-Gordon, *J. Chem. Phys.*, 1996, **105**, 5061.
- 58 C. A. White and M. Head-Gordon, *Chem. Phys. Lett.*, 1996, **257**, 647.
- 59 C. A. White, B. G. Johnson, P. M. W. Gill and M. Head-Gordon, *Chem. Phys. Lett.*, 1996, **253**, 268.
- 60 C. A. White and M. Head-Gordon, *J. Chem. Phys.*, 1996, **104**, 2620.
- 61 Y. H. Shao and M. Head-Gordon, *Chem. Phys. Lett.*, 2000, **323**, 425.
- 62 Y. H. Shao, C. A. White and M. Head-Gordon, *J. Chem. Phys.*, 2001, **114**, 6572.
- 63 L. Fusti-Molnar and P. Pulay, *J. Chem. Phys.*, 2002, **117**, 7827.
- 64 L. Fusti-Molnar and P. Pulay, *J. Chem. Phys.*, 2002, **116**, 7795.
- 65 L. Fusti-Molnar, *J. Chem. Phys.*, 2003, **119**, 11080.
- 66 L. Fusti-Molnar and J. Kong, *J. Chem. Phys.*, 2005, **122**, 074108.
- 67 A. D. Becke, *J. Chem. Phys.*, 1988, **88**, 2547.
- 68 P. M. W. Gill, B. G. Johnson and J. A. Pople, *Chem. Phys. Lett.*, 1993, **209**, 506.
- 69 S. H. Chien and P. M. W. Gill, *J. Comput. Chem.*, 2006, **27**, 730.
- 70 S. T. Brown and J. Kong, *Chem. Phys. Lett.*, 2005, **408**, 395.
- 71 J. Almlöf, K. Faegri and K. Korsell, *J. Comput. Chem.*, 1982, **3**, 385.
- 72 S. T. Brown, L. Fusti-Molnar and J. Kong, *Chem. Phys. Lett.*, 2006, **418**, 490.
- 73 R. Jurgens-Lutovsky and J. Almlöf, *Chem. Phys. Lett.*, 1991, **178**, 451.
- 74 M. S. Lee and M. Head-Gordon, *Comput. Chem.*, 2000, **24**, 295.
- 75 K. Wolinski and P. Pulay, *J. Chem. Phys.*, 2003, **118**, 9497.
- 76 W. Z. Liang and M. Head-Gordon, *J. Phys. Chem. A*, 2004, **08**, 3206.
- 77 T. H. Dunning, *J. Chem. Phys.*, 1989, **90**, 1007.
- 78 T. H. Dunning, K. A. Peterson and A. K. Wilson, *J. Chem. Phys.*, 2001, **114**, 9244.

- 79 C. Ochsenfeld, C. A. White and M. Head-Gordon, *J. Chem. Phys.*, 1998, **109**, 1663.
- 80 G. E. Scuseria, *J. Phys. Chem. A*, 1999, **103**, 4782.
- 81 R. Baer and M. Head-Gordon, *Phys. Rev. Lett.*, 1997, **79**, 3962.
- 82 S. Goedecker, *Rev. Mod. Phys.*, 1999, **71**, 1085.
- 83 D. R. Bowler, T. Miyazaki and M. J. Gillan, *J. Phys.: Condens. Matter*, 2002, **14**, 2781.
- 84 M. Challacombe, *Comput. Phys. Commun.*, 2000, **128**, 93.
- 85 C. Saravanan, Y. Shao, R. Baer, P. N. Ross and M. Head-Gordon, *J. Comput. Chem.*, 2003, **24**, 618.
- 86 A. H. R. Palser and D. E. Manolopoulos, *Phys. Rev. B: Condens. Matter*, 1998, **58**, 12704.
- 87 Y. Shao, C. Saravanan, M. Head-Gordon and C. A. White, *J. Chem. Phys.*, 2003, **118**, 6144.
- 88 M. Head-Gordon, Y. H. Shao, C. Saravanan and C. A. White, *Mol. Phys.*, 2003, **101**, 37.
- 89 R. Baer and M. Head-Gordon, *J. Chem. Phys.*, 1997, **107**, 10003.
- 90 W. Z. Liang, C. Saravanan, Y. H. Shao, R. Baer, A. T. Bell and M. Head-Gordon, *J. Chem. Phys.*, 2003, **119**, 4117.
- 91 W. Z. Liang, R. Baer, C. Saravanan, Y. H. Shao, A. T. Bell and M. Head-Gordon, *J. Comput. Phys.*, 2004, **194**, 575.
- 92 T. Helgaker, J. Gauss, P. Jorgensen and J. Olsen, *J. Chem. Phys.*, 1997, **106**, 6430.
- 93 E. F. C. Byrd, C. D. Sherrill and M. Head-Gordon, *J. Phys. Chem. A*, 2001, **105**, 9736.
- 94 F. Weigend, M. Haser, H. Patzelt and R. Ahlrichs, *Chem. Phys. Lett.*, 1998, **294**, 143.
- 95 F. Weigend, A. Kohn and C. Hattig, *J. Chem. Phys.*, 2002, **116**, 3175.
- 96 B. I. Dunlap, *Phys. Chem. Chem. Phys.*, 2000, **2**, 2113.
- 97 M. Feyereisen, G. Fitzgerald and A. Komornicki, *Chem. Phys. Lett.*, 1993, **208**, 359.
- 98 F. Weigend and M. Haser, *Theor. Chem. Acc.*, 1997, **97**, 331.
- 99 P. J. Knowles, J. S. Andrews, R. D. Amos, N. C. Handy and J. A. Pople, *Chem. Phys. Lett.*, 1991, **186**, 130.
- 100 D. J. Tozer, N. C. Handy, R. D. Amos, J. A. Pople, R. H. Nobes, Y. M. Xie and H. F. Schaefer, *Mol. Phys.*, 1993, **79**, 777.
- 101 M. Head-Gordon, *Mol. Phys.*, 1999, **96**, 673.
- 102 P. Pulay, *Chem. Phys. Lett.*, 1983, **100**, 151.
- 103 S. Saebø and P. Pulay, *Annu. Rev. Phys. Chem.*, 1993, **44**, 213.
- 104 M. S. Lee, P. E. Maslen and M. Head-Gordon, *J. Chem. Phys.*, 2000, **112**, 3592.
- 105 M. S. Lee and M. Head-Gordon, *Int. J. Quantum Chem.*, 2000, **76**, 169.
- 106 R. A. DiStasio, Y. S. Jung and M. Head-Gordon, *J. Chem. Theor. Comput.*, 2005, **1**, 862.
- 107 S. Grimme, *J. Chem. Phys.*, 2003, **118**, 9095.
- 108 Y. S. Jung, R. C. Lochan, A. D.UTOI and M. Head-Gordon, *J. Chem. Phys.*, 2004, **121**, 9793.
- 109 J. Almlöf, *Chem. Phys. Lett.*, 1991, **181**, 319.
- 110 R. C. Lochan, Y. Jung and M. Head-Gordon, *J. Phys. Chem. A*, 2005, **109**, 7598.
- 111 Y. Jung, A. Sodt, P. M. W. Gill and M. Head-Gordon, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 6692.
- 112 P. M. W. Gill, A. T. B. Gilbert, S. W. Taylor, G. Friesecke and M. Head-Gordon, *J. Chem. Phys.*, 2005, **123**, 061101.
- 113 R. J. Bartlett and J. F. Stanton, in *Reviews in Computational Chemistry, Volume V*, ed. K. B. Lipkowitz and D. B. Boyd, VCH Publishers Inc, New York, 1994, p. 65.
- 114 T. D. Crawford and H. F. Schaefer, *Rev. Comput. Chem.*, 2000, **14**, 33.
- 115 G. D. Purvis and R. J. Bartlett, *J. Chem. Phys.*, 1982, **76**, 1910.
- 116 K. Raghavachari, G. W. Trucks, J. A. Pople and M. Head-Gordon, *Chem. Phys. Lett.*, 1989, **157**, 479.
- 117 S. R. Gwaltney and M. Head-Gordon, *J. Chem. Phys.*, 2001, **115**, 2014.
- 118 G. J. O. Beran, S. R. Gwaltney and M. Head-Gordon, *Phys. Chem. Chem. Phys.*, 2003, **5**, 2488.
- 119 C. D. Sherrill, A. I. Krylov, E. F. C. Byrd and M. Head-Gordon, *J. Chem. Phys.*, 1998, **109**, 4171.
- 120 T. Van Voorhis and M. Head-Gordon, *Chem. Phys. Lett.*, 2000, **330**, 585.
- 121 E. F. C. Byrd, T. Van Voorhis and M. Head-Gordon, *J. Phys. Chem. B*, 2002, **106**, 8070.
- 122 A. I. Krylov, C. D. Sherrill and M. Head-Gordon, *J. Chem. Phys.*, 2000, **113**, 6509.
- 123 S. R. Gwaltney and M. Head-Gordon, *Chem. Phys. Lett.*, 2000, **323**, 21.
- 124 A. I. Krylov, C. D. Sherrill, E. F. C. Byrd and M. Head-Gordon, *J. Chem. Phys.*, 1998, **109**, 10669.
- 125 S. R. Gwaltney, C. D. Sherrill, M. Head-Gordon and A. I. Krylov, *J. Chem. Phys.*, 2000, **113**, 3548.
- 126 S. R. Gwaltney, E. F. C. Byrd, T. Van Voorhis and M. Head-Gordon, *Chem. Phys. Lett.*, 2002, **353**, 359.
- 127 J. F. Stanton and R. J. Bartlett, *J. Chem. Phys.*, 1993, **98**, 7029.
- 128 A. I. Krylov, *Chem. Phys. Lett.*, 2001, **338**, 375.
- 129 S. V. Levchenko and A. I. Krylov, *J. Chem. Phys.*, 2004, **120**, 175.
- 130 J. F. Stanton and J. Gauss, *J. Chem. Phys.*, 1994, **101**, 8938.
- 131 M. Nooijen and R. J. Bartlett, *J. Chem. Phys.*, 1995, **102**, 3629.
- 132 D. Sinha, S. K. Mukhopadhyay, R. Chaudhuri and D. Mukherjee, *Chem. Phys. Lett.*, 1989, **154**, 544.
- 133 M. Wladyslawski and M. Nooijen, *Low-Lying Potential Energy Surfaces, ACS Symposium Series*, 2002, **828**, 65.
- 134 S. V. Levchenko, T. Wang and A. I. Krylov, *J. Chem. Phys.*, 2005, **122**, 224106.
- 135 L. V. Slipchenko and A. I. Krylov, *J. Chem. Phys.*, 2002, **117**, 4694.
- 136 L. V. Slipchenko and A. I. Krylov, *J. Chem. Phys.*, 2003, **118**, 6874.
- 137 T. E. Munsch, L. V. Slipchenko, A. I. Krylov and P. G. Wenthold, *J. Org. Chem.*, 2004, **69**, 5735.
- 138 T. Wang and A. I. Krylov, *J. Chem. Phys.*, 2005, **123**, 104304.
- 139 A. M. C. Cristian, Y. Shao and A. I. Krylov, *J. Phys. Chem. A*, 2004, **108**, 6581.
- 140 L. V. Slipchenko, T. E. Munsch, P. G. Wenthold and A. I. Krylov, *Angew. Chem., Int. Ed.*, 2004, **43**, 742.
- 141 S. Hirata, M. Nooijen and R. J. Bartlett, *Chem. Phys. Lett.*, 2000, **326**, 255.
- 142 L. V. Slipchenko and A. I. Krylov, *J. Chem. Phys.*, 2005, **123**, 084107.
- 143 P. R. Surjan, *Top. Curr. Chem.*, 1999, **203**, 63.
- 144 V. A. Rassolov, *J. Chem. Phys.*, 2002, **117**, 5978.
- 145 W. A. Goddard and L. B. Harding, *Annu. Rev. Phys. Chem.*, 1978, **29**, 363.
- 146 V. A. Rassolov, F. Xu and S. Garashchuk, *J. Chem. Phys.*, 2004, **120**, 10385.
- 147 T. Van Voorhis and M. Head-Gordon, *J. Chem. Phys.*, 2000, **113**, 8873.
- 148 J. Cullen, *Chem. Phys.*, 1996, **202**, 217.
- 149 T. Van Voorhis and M. Head-Gordon, *J. Chem. Phys.*, 2002, **117**, 9190.
- 150 G. J. O. Beran, B. Austin, A. Sodt and M. Head-Gordon, *J. Phys. Chem. A*, 2005, **109**, 9183.
- 151 T. Van Voorhis and M. Head-Gordon, *Chem. Phys. Lett.*, 2000, **317**, 575.
- 152 T. Van Voorhis and M. Head-Gordon, *J. Chem. Phys.*, 2000, **115**, 7814.
- 153 M. Head-Gordon, *Chem. Phys. Lett.*, 2003, **372**, 508.
- 154 A. D.UTOI, Y. S. Jung and M. Head-Gordon, *J. Phys. Chem. A*, 2004, **108**, 10270.
- 155 Y. S. Jung and M. Head-Gordon, *ChemPhysChem*, 2004, **4**, 522.
- 156 Y. S. Jung and M. Head-Gordon, *J. Phys. Chem. A*, 2003, **107**, 7475.
- 157 Y. S. Jung and M. Head-Gordon, *Phys. Chem. Chem. Phys.*, 2004, **6**, 2008.
- 158 Y. S. Jung, T. Heine, P. V. R. Schleyer and M. Head-Gordon, *J. Am. Chem. Soc.*, 2004, **126**, 3132.
- 159 D. Scheschkewitz, H. Amii, H. Gornitzka, W. W. Schoeller, D. Bourissou and G. Bertrand, *Science*, 2002, **295**, 1880.
- 160 A. Warshel and M. Levitt, *J. Mol. Biol.*, 1976, **103**, 227.
- 161 U. C. Singh and P. A. Kollman, *J. Comput. Chem.*, 1986, **7**, 718.
- 162 P. A. Bash, M. J. Field and M. Karplus, *J. Am. Chem. Soc.*, 1987, **109**, 8092.
- 163 F. Maseras and K. Morokuma, *J. Comput. Chem.*, 1995, **16**, 1170.
- 164 N. Reuter, A. Dejaegere, B. Maigret and M. Karplus, *J. Phys. Chem. A*, 2000, **104**, 1720.
- 165 A. Shurki and A. Warshel, *Adv. Protein Chem.*, 2000, **66**, 249.
- 166 R. A. Friesner and V. Guallar, *Annu. Rev. Phys. Chem.*, 2005, **56**, 389.

- 167 M. Freindorf and J. L. Gao, *J. Comput. Chem.*, 1996, **17**, 386.
- 168 M. Freindorf, Y. H. Shao, T. R. Furlani and J. Kong, *J. Comput. Chem.*, 2005, **26**, 1270.
- 169 D. A. Pearlman, D. A. Case, J. W. Caldwell, W. S. Ross, T. E. Cheatham, S. Debolt, D. Ferguson, G. Seibel and P. Kollman, *Comput. Phys. Commun.*, 1995, **91**, 1.
- 170 F. S. Lee, Z. T. Chu and A. Warshel, *J. Comput. Chem.*, 1993, **14**, 161.
- 171 V. Luzhkov and A. Warshel, *J. Am. Chem. Soc.*, 1991, **113**, 4491.
- 172 E. Rosta, M. Klahn and A. Warshel, *J. Phys. Chem. B*, 2006, **110**, 2934.
- 173 D. Das, K. P. Eurenium, E. M. Billings, P. Sherwood, D. C. Chatfield, M. Hodoscek and B. R. Brooks, *J. Chem. Phys.*, 2002, **117**, 10534.
- 174 J. Tomasi and M. Persico, *Chem. Rev.*, 1994, **94**, 2027.
- 175 C. J. Cramer and D. G. Truhlar, *Chem. Rev.*, 1999, **99**, 2161.
- 176 J. Florian and A. Warshel, *J. Phys. Chem. B*, 1997, **101**, 5583.
- 177 D. M. Chipman, *J. Chem. Phys.*, 2000, **112**, 5558.
- 178 D. M. Chipman, *Theor. Chem. Acc.*, 2002, **107**, 80.
- 179 D. M. Chipman and M. Dupuis, *Theor. Chem. Acc.*, 2002, **107**, 90.
- 180 D. Maurice and M. Head-Gordon, *Mol. Phys.*, 1999, **96**, 1533.
- 181 H. B. Schlegel, *J. Comput. Chem.*, 2003, **24**, 1514.
- 182 P. Pulay, *Mol. Phys.*, 1969, **17**, 197.
- 183 J. Baker, *J. Comput. Chem.*, 1986, **7**, 385.
- 184 P. E. Maslen, *J. Chem. Phys.*, 2005, **122**, 014104.
- 185 J. Baker, A. Kessi and B. Delley, *J. Chem. Phys.*, 1996, **105**, 192.
- 186 J. Baker, *J. Comput. Chem.*, 1997, **18**, 1079.
- 187 R. Chang, P. A. Barile and P. E. Maslen, *J. Chem. Phys.*, 2004, **120**, 8379.
- 188 P. E. Maslen, *Mol. Phys.*, 2005, **103**, 841.
- 189 A. Banerjee, N. Adams, J. Simons and R. Shepard, *J. Phys. Chem.*, 1985, **89**, 52.
- 190 G. Henkelman and H. Jonsson, *J. Chem. Phys.*, 1999, **111**, 7010.
- 191 A. Heyden, B. Peters, A. T. Bell and F. J. Keil, *J. Phys. Chem. B*, 2005, **109**, 1857.
- 192 A. Heyden, A. T. Bell and F. J. Keil, *J. Chem. Phys.*, 2005, **123**, 224101.
- 193 G. Mills and H. Jonsson, *Phys. Rev. Lett.*, 1994, **72**, 1124.
- 194 G. Henkelman and H. Jonsson, *J. Chem. Phys.*, 2000, **113**, 9978.
- 195 W. N. E. W. Q. Ren and E. Vanden-Eijnden, *Phys. Rev. B: Condens. Matter*, 2002, **66**, 052301.
- 196 B. Peters, A. Heyden, A. T. Bell and A. Chakraborty, *J. Chem. Phys.*, 2004, **120**, 7877.
- 197 T. Helgaker, E. Uggerud and H. J. A. Jensen, *Chem. Phys. Lett.*, 1990, **173**, 145.
- 198 M. S. Gordon, G. Chaban and T. Taketsugu, *J. Phys. Chem.*, 1996, **100**, 11512.
- 199 J. M. Herbert and M. Head-Gordon, *Phys. Chem. Chem. Phys.*, 2005, **7**, 3269.
- 200 P. Pulay and G. Fogarasi, *Chem. Phys. Lett.*, 2004, **386**, 272.
- 201 R. Car and M. Parrinello, *Phys. Rev. Lett.*, 1985, **55**, 2471.
- 202 J. M. Herbert and M. Head-Gordon, *J. Chem. Phys.*, 2004, **121**, 11542.
- 203 A. Edelman, T. A. Arias and S. T. Smith, *SIAM J. Mater. Anal. Appl.*, 1998, **20**, 303.
- 204 T. Van Voorhis and M. Head-Gordon, *Mol. Phys.*, 2002, **100**, 1713.
- 205 B. Dunietz, T. Van Voorhis and M. Head-Gordon, *J. Theor. Comput. Chem.*, 2002, **1**, 255.
- 206 C. Ochsenfeld, J. Kussmann and F. Koziol, *Angew. Chem., Int. Ed.*, 2004, **43**, 4485.
- 207 C. Ochsenfeld, *Phys. Chem. Chem. Phys.*, 2002, **2**, 2153.
- 208 C. Ochsenfeld, S. P. Brown, I. Schnell, J. Gauss and H. W. Spiess, *J. Am. Chem. Soc.*, 2001, **123**, 2597.
- 209 S. P. Brown, T. Schaller, U. P. Seelbach, F. Koziol, C. Ochsenfeld, F. G. Klarner and H. W. Spiess, *Angew. Chem., Int. Ed.*, 2001, **40**, 717.
- 210 C. Ochsenfeld, F. Koziol, S. P. Brown, T. Schaller, U. P. Seelbach and F. G. Klarner, *Solid State Nucl. Magn. Reson.*, 2002, **22**, 128.
- 211 R. Ditchfield, *Mol. Phys.*, 1974, **27**, 789.
- 212 K. Wolinski, J. F. Hinton and P. Pulay, *J. Am. Chem. Soc.*, 1990, **112**, 8251.
- 213 J. Gauss, *Ber. Bunsen-Ges. Phys. Chem.*, 1995, **99**, 1001.
- 214 C. Ochsenfeld, *Chem. Phys. Lett.*, 2000, **327**, 216.
- 215 C. A. White and M. Head-Gordon, *J. Chem. Phys.*, 1994, **101**, 6593.
- 216 S. P. Karna and M. Dupuis, *J. Comput. Chem.*, 1991, **12**, 487.
- 217 H. Sekino and R. J. Bartlett, *J. Chem. Phys.*, 1986, **85**, 976.
- 218 V. A. Rassolov and D. M. Chipman, *J. Chem. Phys.*, 1996, **104**, 9908.
- 219 D. M. Chipman, *Theor. Chim. Acta*, 1989, **76**, 73.
- 220 V. A. Rassolov and D. M. Chipman, *J. Chem. Phys.*, 1996, **105**, 1470.
- 221 V. A. Rassolov and D. M. Chipman, *J. Chem. Phys.*, 1996, **105**, 1479.
- 222 B. Wang, J. Baker and P. Pulay, *Phys. Chem. Chem. Phys.*, 2000, **2**, 2131.
- 223 A. E. Reed, R. B. Weinstock and F. Weinhold, *J. Chem. Phys.*, 1985, **83**, 735.
- 224 A. E. Reed, L. A. Curtiss and F. Weinhold, *Chem. Rev.*, 1988, **88**, 899.
- 225 A. C. Simmonett, A. T. B. Gilbert and P. M. W. Gill, *Mol. Phys.*, 2005, **103**, 2789.
- 226 S. F. Boys, *Rev. Mod. Phys.*, 1960, **32**, 296.
- 227 C. Edmiston and K. Ruedenberg, *Rev. Mod. Phys.*, 1963, **35**, 457.
- 228 J. Pipek and P. G. Mezey, *J. Chem. Phys.*, 1989, **90**, 4916.
- 229 J. E. Subotnik, Y. H. Shao, W. Z. Liang and M. Head-Gordon, *J. Chem. Phys.*, 2004, **121**, 9220.
- 230 P. M. W. Gill, D. P. O'Neill and N. A. Besley, *Theor. Chem. Acc.*, 2003, **109**, 241.
- 231 N. A. Besley, A. M. Lee and P. M. W. Gill, *Mol. Phys.*, 2002, **100**, 1763.
- 232 P. M. W. Gill, D. L. Crittenden, D. P. O'Neill and N. A. Besley, *Phys. Chem. Chem. Phys.*, 2006, **8**, 15.
- 233 N. A. Besley, D. P. O'Neill and P. M. W. Gill, *J. Chem. Phys.*, 2003, **118**, 2033.