LA-UR- 96-0244

Title: New Approaches for Quantum Chemical
Calculations on Very Large Molecules

RECEIVED
APR 01 1996
OSTI

Author(s): Richard L. Martin, T-12
Thomas V. Russo, T-12
P. Jeffrey Hay, T-12

Submitted to: DOE Office of Scientific and Technical
Information (OSTI)

Los Alamos
NATIONAL LABORATORY

Los Alamos National Laboratory, an affirmative action/equal opportunity employer, is operated by the University of California for the U.S. Department of Energy under contract W-7405-ENG-36. By acceptance of this article, the publisher recognizes that the U.S. Government retains a nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or to allow others to do so, for U.S. Government purposes. The Los Alamos National Laboratory requests that the publisher identify this article as work performed under the auspices of the U.S. Department of Energy.

Form No. 836 R5
ST 2629 1091

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

MASTER

DISCLAIMER

**Portions of this document may be illegible
in electronic image products. Images are
produced from the best available original
document.**

New Approaches for Quantum Chemical Calculations on Very Large Molecules

Richard L. Martin*, Thomas V. Russo, and P. Jeffrey Hay

Abstract

This is the final report for a three-year, Laboratory-Directed Research and Development (LDRD) project at the Los Alamos National Laboratory (LANL). *Ab initio* quantum chemistry has proven to be a valuable tool for understanding the electronic structure of small molecules containing 5-50 atoms. Calculations on larger systems have been prohibitive because of the forbidding scaling properties of the methods used. In the simplest traditional many-body method, second-order perturbation theory, the work scales as the fifth power of N , where N is proportional to the number of electrons in the system. New methods for the many-body problem, which scale as the third power of N and are well suited to utilize parallel computation, have been studied. The combination of these two developments enable the treatment of molecules in the range of 50-500 atoms with first principles methods.

1. Background and Research Objectives

Ab initio quantum chemistry has proven to be an extremely valuable tool for understanding the electronic structure of molecules. Specific molecular properties such as the equilibrium geometry, frequencies and intensities of infrared and Raman active vibrational modes, excitation energies and intensities of transitions to excited states, and properties of transition states in chemical reactions can be probed with these techniques. For certain properties, such as equilibrium geometries, sufficient accuracy can be achieved at the Hartree-Fock (HF) level of approximation. Other properties require more accurate treatments that incorporate the effects of electron correlation using many-body techniques. While such calculations have proven quite useful for small molecules containing less than approximately 10 atoms, many chemically interesting species lie in the 50-500 atom regime.

The reason that these techniques are limited to smaller systems is that the computational effort involved in these methods increases very rapidly with the size of the system. The work in HF calculations scales as N^4 , where N is proportional to the number of electrons in the system. The simplest traditional quantum chemistry many-body method,

*Principal investigator, email: rlm@lanl.gov

second-order perturbation theory (MP2), scales as N^5 . This project was initiated to examine many-body correlation methods that scale as N^3 . With these new algorithms and access to massively parallel computer architectures such as the CM-5 computer, we hoped to study molecules at least an order of magnitude larger than those addressed by existing methods.

2. Importance to LANL's Science and Technology Base and National R&D Needs

The ability to treat molecules containing 50-500 atoms with quantum chemistry techniques would open new vistas for theoretical chemistry. Large organic molecules of pharmaceutical interest and novel species such as fullerenes would be amenable to study. Inorganic complexes containing transition metals and actinides could be treated using realistic ligands. Important problems such as the chemistry at active sites in enzymes and the photochemistry involved in photosynthesis of chlorophyll could be studied. Catalytic reactions of hydrocarbons on zeolites and other supports of interest to the petrochemical and chemical industries could be examined. This work complements ongoing LANL programs in computational chemistry and enhances the capability of the Laboratory in the areas of computational and quantum chemistry.

3. Scientific Approach and Results

This project had a very specific goal. The methods used in quantum chemistry that are capable of producing accurate bond energies must include the correlations among the motions of the electrons. Methods that allow these correlations are limited to relatively small molecules (approximately ten atoms) because of the way in which they scale with the number of atoms. This project had the specific goal of investigating new methods that can be applied to larger molecules.

The research was originally based on two considerations. First, an idea that allowed second-order many-body perturbation theory to be applied with an effort scaling as N^4 as opposed to N^5 . Second, the availability of a new machine, the massively parallel CM-5 computer. At the completion of this project, accurate calculations on molecules containing approximately 50 atoms were being performed, but using neither the original idea nor the CM-5 computer. In this section we briefly review the progress, the problems, and the path taken in this research.

In the initial stages of this effort the "auxiliary basis function" method was employed. The initial results were encouraging, and aspects of the approach have been pursued by other groups. However, this many-body perturbation theory approach was abandoned after the appearance of a paper by Johnson, Gill and Pople [1] on "gradient-corrected" density functional theory (GCDFT). While density functional theory (DFT), particularly the local density approximation (LDA), is widely used in solid-state physics, it was not generally accepted in the theoretical chemistry community because of its well-known problems in overestimating bond energies. For example, when Johnson, *et al.* used the LDA to predict the heats of atomization of 32 small molecules of first-row elements, the mean deviation between theory and experiment was 36 kcal/mole, with the LDA consistently overbinding the molecules. By way of contrast, the error with MP2 was 22 kcal/mole, and it consistently underbound the molecules. The exciting aspect of this study, however, was the performance of the newly-developed GCDFT. These functions include not only the density, but the gradient of the density at each point in space in order to estimate the correlation energy. A particular form of gradient-corrected theory, B-LYP, was in error for this set of molecules by only 1.0 kcal/mole. This exceptional performance of GCDFT relative to MP2, along with its favorable scaling properties, led us to focus on it for the remainder of this project.

We have written a GCDFT code called Truchas [2,3]. It is approximately 150,000 lines long, about 25,000 of which are devoted to the execution of GCDFT. Since the study of Johnson, *et al.* was limited to small molecules of first-row elements, and we were primarily interested in inorganic species, we began systematic studies of this theory for transition metal complexes.

In order to describe transition metal complexes reliably, a theoretical method must be able to accurately reproduce the atomic state splittings. We therefore first examined GCDFT for a simple, but notoriously difficult, problem: the ionization potentials and excitation energies of the first-transition-series atoms. In preparation for this study, we developed a spin-restricted version of DFT [4]. The B-LYP method reproduced the ionization potentials in this series with a mean deviation of 0.15 eV [4]. This can be compared with a mean deviation using MP2 of 0.4 eV. The errors in the excitation energies were somewhat larger, but still preferable to MP2.

The next step in our research was to examine the performance of GCDFT on the bond energies and geometries of small transition-metal complexes. This required the implementation of analytic techniques to evaluate the force on the atoms [5]. With this new capability we first examined a series of transition metal fluorides (ScF₃, TiF₄, VF₅) for which accurate bond energies were available from experiment. An additional incentive was that the next member of this series, CrF₆, was the subject of considerable controversy in the literature as to whether or

not it had been prepared in the gas phase. Figure 1 shows the results of Hartree-Fock theory, the LDA, and the B-LYP approximation with respect to the average bond energy [5]. The LDA overestimates the bond energy by about 30 kcal/mole per bond. The B-LYP approximation reduces this error to about 8 kcal/mole per bond. Most importantly, a "hybrid" approach that couples GCDFT with a component of the full non-local exchange interaction, gave the results denoted as Becke3LYP in the figure. The results are within 1-2 kcal/mole per bond of experimental values across the series. This level of accuracy for a transition metal complex is remarkable. The bond distances are also within 0.01Å of experimental values. The hybrid method was used to settle the controversy regarding the existence and vibrational spectrum of CrF₆. Additional studies on metal carbonyls [6] show similar accuracy. These hybrid methods are very promising.

Given this extremely encouraging agreement with experiment, we next addressed the issues that needed to be resolved in order to apply these techniques to the remainder of the periodic table: systems containing second- and third-row transition metals, lanthanides and actinides. These systems demand that relativistic effects be treated properly, and this is usually done via the use of effective core potentials which replace the "chemically inert" core electrons with an effective potential. New effective potentials appropriate for DFT calculations were developed and benchmarked for their accuracy [7]. In addition, we discovered that the existing effective core potentials could be used in these calculations with little loss of accuracy as long as some care was taken to redefine the basis sets used in the calculations. These effective core potentials have been used in a study [8] of the oxidation of butene by Cp*ReO₃, a prototypical reaction of importance in homogenous catalysis.

The second aspect of this project, implementation of these methods on the CM-5 computer, was somewhat frustrating. While this machine has been useful for several important computational problems, such as global climate models and molecular dynamics, the requirements of electronic structure theory demand a more diverse set of mathematical operations. Efficient implementation on the CM-5 computer requires rewriting the code in a language, CM-FORTRAN, specific to the CM-5 computer. Over a period of one year, three separate algorithms were written and tested, but none of them performed satisfactorily. After talking with several other groups who had experienced similar problems with this machine, we abandoned the CM-5 computer and began an implementation aimed at a network of workstations using a message passing language called TCGMSG. This code runs in parallel on our HP workstations and on a cluster of eight IBM computers. We have benchmarked the way in which the program scales with the number of atoms, improved communications and load balancing, and reduced memory requirements for each node. A new approach to formation of the Coulomb matrix, which combines a numerical solution of the Poisson

equation with analytic integration techniques, was stimulated by our experience with the CM-5 computer. The new approach runs approximately an order of magnitude faster than the conventional approach. Using this technique [9], a calculation on a 61-atom Fe-porphyrin takes about an hour on the IBM cluster. This is at the lower limit of our originally stated goal. Given the fact that the IBM cluster has 8 nodes, as opposed to the 1024 of the CM-5 computer, we are pleased with this achievement.

Supplemental work completed on this project included (1) the development of analytic second derivatives of effective core potentials needed for evaluation of vibrational frequencies in transition metal complexes [10]; (2) the implementation of electric field gradient integrals which were used [11] to resolve an issue in the nuclear quadrupole resonance spectrum of La_2CuO_4 ; (3) the coupling of GCDFT with dielectric continuum models of solvation [12] as shown in Figure 2; (4) an investigation of the thermochemistries predicted by GCDFT for small metal clusters [13]; and (5) a study of the scaling of the method on a series of Fe-porphyrins [9].

In summary, we have found that GCDFT, and in particular the hybrid methods that incorporate a portion of the full non-local exchange interaction, give remarkably accurate treatments of molecular electronic structure. The methods scale favorably, and a parallel implementation of this theory capable of accurately treating systems with about 50 atoms is in routine use. We suspect that these methods, their extensions and variants, will be the workhorse of electronic structure theory for the next decade. Of course, with the capability to examine larger systems comes new problems. For example, it is known that the techniques used to determine optimum geometries for small molecules do not work so well for larger ones.

References

- [1] B.G. Johnson, P.M.W. Gill and J.A. Pople, *J. Chem. Phys.*, **98**, 5612 (1993).
- [2] R.L. Martin, T.V. Russo, B.H. Lengsfeld III, P.W. Saxe, G.J. Tawa, M. Page, P.J. Hay and A.K. Rappe, "Truchas," copyright 1995.
- [3] R.L. Martin, T.V. Russo, B.H. Lengsfeld III, P.W. Saxe, G.J. Tawa, M. Page, P.J. Hay and A.K. Rappe, "Truchas User's Guide," LAUR 95-1629, (1995).
- [4] Thomas V. Russo, Richard L. Martin and P. Jeffrey Hay, *J. Chem. Phys.*, **101**, 7729 (1994).
- [5] Thomas V. Russo, Richard L. Martin and P. Jeffrey Hay, *J. Chem. Phys.*, **102**, 8023 (1995).
- [6] Thomas V. Russo, Richard L. Martin and P. Jeffrey Hay, unpublished.
- [7] Thomas V. Russo, Richard L. Martin and P. Jeffrey Hay, *J. Phys. Chem.*, **99**, 17095 (1995).
- [8] Thomas V. Russo, Richard L. Martin and P. Jeffrey Hay, to be published.
- [9] Richard L. Martin and T.V. Russo, to be published.
- [10] T. V. Russo, R.L. Martin, P.J. Hay, and A.K. Rappe, *J. Chem. Phys.*, **102**, 9315 (1995).

- [11] R. L. Martin, Phys. Rev. Lett., **75**, 744(1995).
- [12] G. J. Tawa, R.L. Martin, L.R. Pratt and T.V. Russo, LAUR 94-3125, J. Phys. Chem., in press.
- [13] C.S. Yeh, Y.G. Byun, S. Afzaal, S.Z. Kan, S. Lee, B.S. Frieser and P.J. Hay, J.Am. Chem. Soc., **117**, 4042 (1995).

Calculated Metal-Fluorine Bond Energies from HF and DFT Approaches

Figure 1. The average bond energy in a series of metal fluorides as predicted by several methods compared with experimental data. HF denotes Hartree-Fock, S-VWN the local density approximation, B-LYP a gradient-corrected theory, and Becke3LYP a hybrid functional which incorporates a portion of the full non-local exchange.

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

Figure 2. The free energies of solvation for a number of small organic molecules compared with the values predicted using a dielectric continuum model of solvation and gradient-corrected density functional theory.