Multi-coefficient Gaussian-3 method for calculating potential energy surfaces

Patton L. Fast, María Luz Sánchez, Donald G. Truhlar *

Department of Chemistry and Supercomputer Institute, University of Minnesota, Minneapolis, MN 55455-0431, USA

Abstract

We propose a multi-coefficient modification (MCG3) of the Gaussian-3 (G3) electronic structure method that is suitable for calculating continuous potential energy surfaces. We tested it for atomization energies and found that it improves the accuracy by 8% as compared to G3 and reduces the cost of single-point energy calculations by 50%. The method should be useful for calculating bond energies, potential energy surfaces, and thermochemical data of molecules. © 1999 Elsevier Science B.V. All rights reserved.

The Gaussian-3 (G3) method [1] is a recently proposed modification of the successful and widely applied Gaussian-2 (G2) method [2] for thermochemical calculations that both reduces the cost and decreases the average error. Both G3 and G2 are defined to include geometry optimization followed by calculations of electronic and vibrational energy to yield heats of formation. Our interest is more general, namely in the calculation of potential energy surfaces (heats of formation may be calculated from potential energy surfaces, but not vice versa). Thus when we refer to the G2 or G3 methods we refer to single-point energy calculations by the same procedure as used for the electronic energy part of the original G2 or G3 scheme; similar uses of G2 have been considered by previous workers [3].

The G3 method involves three essential improvements over G2, namely: (1) substitution of a polarized 6-31G basis set for the polarized 6-311G basis set in some of the steps; (2) a more balanced treatment of contracted *s* and *p* functions and valence polarization functions for first- and second-row atoms; and (3) a final-step calculation involving core correlation and core polarization functions. Improvement (1) lowers the cost more than improvements (2) and (3) raise it, and the gain in accuracy from improvements (2) and (3) outweigh the potential loss in accuracy from modification (1).

We have recently proposed a modification, called multi-coefficient G2 (MCG2) [4], of the G2 scheme that decreases the cost about 10% and decreases the average error for first-row atoms by almost a factor of 2. In the present Letter we propose a similar modification, called multi-coefficient G3 (MCG3) for the G3 method. The advantages of MCG3 as compared to G3 are reduced cost, improved accuracy, and making the energy a continuous function of geometry. The advantage of MCG3 as compared to MCG2 is that MCG3 can treat both first- and sec-

---

* Corresponding author. E-mail: truhlar@chem.umn.edu

ond-row atoms reasonably uniformly, whereas we restricted MCG2 to first-row atoms because the basis sets in the G2 method do not treat the second row uniformly with the first row.

The electronic energy (including nuclear repulsion but not vibration or rotation) in G3 is given by

$$
\begin{aligned}
E(G3) = {} & E[QCISD(T)/6\text{-}31G(d)] \\
& + \Delta E(+) + \Delta E(2df) \\
& + \Delta E(G3large) + \Delta E(SO) \\
& + \Delta E(HLC),
\end{aligned} \tag{1}
$$

where G3large is a new basis set containing core polarization functions, and the other terms are defined elsewhere [1]. The G3 method essentially provides an approximation to a QCISD(T)/G3large calculation from several smaller calculations, namely, from QCISD(T)/6-31G(d), MP4/6-31 + G(d), MP4/6-31G(2df, p), and MP2(full)/G3large energy calculations, where the notation is standard [1,5], and the frozen-core (FC) approximation is implied except where we indicate 'full'. In order to condense the notation in Eq. (2) below, we will abbreviate the first and third basis sets used in these calculations as (d) and (2df, p), respectively. For future reference we note that performing a QCISD(T) calculation with a given basis set also yields lower-level Hartree–Fock (HF), MP2, MP4SDQ, and MP4 results for that basis at no additional cost because they are all part of the overall calculation. Similarly MP4 calculations include MP4SDQ as a subset, and MP2 calculations include HF. These facts become important as we define the MCG3 method.

The MCG3 method is written as

$$
\begin{aligned}
E(MCG3) = {} & c_1 E[HF/(d)] \\
& + c_2 \Delta E[HF/MG3|(d)] \\
& + c_3 \Delta E[MP2|HF/(d)] \\
& + c_4 \Delta E[MP2|HF/MG3|(d)] \\
& + c_5 \Delta E[MP4SDQ|MP2/(d)] \\
& + c_6 \Delta E[MP4SDQ|MP2/(2df, p)|(d)] \\
& + c_7 \Delta E[MP4|MP4SDQ/(d)] \\
& + c_8 \Delta E[MP4|MP4SDQ/(2df, p)|(d)] \\
& + c_9 \Delta E[QCISD(T)|MP4/(d)] \\
& + E_{SO} + E_{CC},
\end{aligned} \tag{2}
$$

where the pipe '|' notation is defined by

$$
\Delta E(L/B2|B1) = E(L/B2) - E(L/B1), \tag{3}
$$

$$
\Delta E(L2|L1/B) = E(L2/B) - E(L1/B), \tag{4}
$$

and

$$
\begin{aligned}
\Delta E(L2|L1/B2|B1) = {} & E(L2|L1/B2) \\
& - (L2|L1/B1),
\end{aligned} \tag{5}
$$

where L1 and L2 denote levels, B1 and B2 denote basis sets, E_{SO} and E_{CC} are simple estimates (requiring negligible computational effort) described elsewhere [6,7] of the spin–orbit and core correlation energy, the coefficients $\{c_i\}$ are optimized against experimental data, and the MG3 (modified G3) basis set denotes the G3large basis set without the core polarization functions. Note that we perform an MP2/MG3 calculation instead of the MP2 (full)/G3large calculation used in the G3 method. This is because we obtain the core correlation effects by a simple estimate [7] instead of by the more expensive electronic structure calculation. Fig. 1 provides a diagram that helps one to visualize the terms in Eq. (2). The motivation for Eq. (2) is that we scale
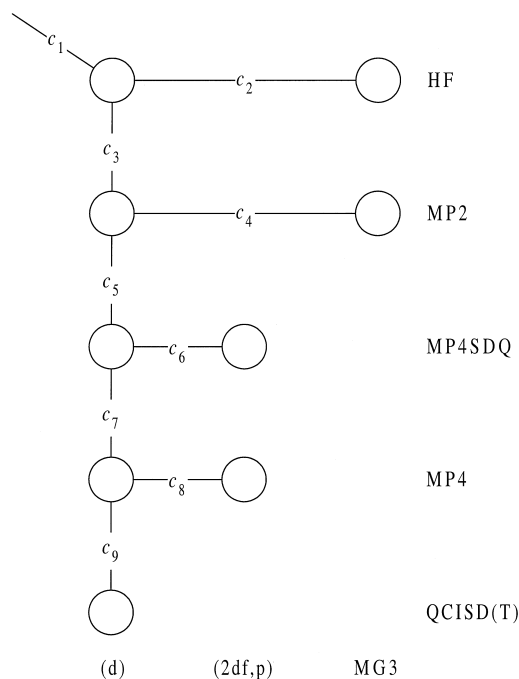

Fig. 1. Coefficient diagram for MCG3.

Table 1
Optimized coefficients

Method	c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8	c_9
MCG3	0.9936	1.1374	1.0148	1.2926	1.1685	1.2005	1.6572	1.6145	1.2206
MMCG3	1.0034	1.1215	1.0645	1.1443	1.2207	1.0666	1.4719	0.8982	1.3037

individual components of the correlation energy to make up for the lack of full configuration interaction and the incompleteness of the one-electron basis set.

All of the quantities needed in Eq. (2) are calculated as part of the G3 method, except for the MP2/MG3 calculation, which replaces the more expensive MP2(full)/G3large calculation. We also note that we do not use any calculations with the 6-31 + G(d) basis set, and this fact also decreases the overall cost of the MCG3 method. As a result, the computer time for the more expensive molecules drops by about a factor of 2.0. [1]

The training set is composed of 49 molecules for which we have recently [6] estimated the zero-point-exclusive atomization energies, D_e, on the basis of experimental heats of formation and accurate estimates of vibrational energies. This 49-molecule set is a subset of the sets against which the G2 and G3 methods were originally parameterized. For the present Letter we calculated the electronic energies of these 49 molecules by the methods used for the components of Eqs. (1) and (2). All these calculations were carried out at MP2/6-31G(d) geometries, which are available at Curtiss's website (http:// chemistry.anl.gov/compmat), and which are the same geometries used in the original G2 and G3 methods. All calculations were carried out with the GAUSSIAN94 computer program [8]. The coefficients of Eq. (2) were then determined by a least-squares fit

to the 49 experimentally based values of D_e and are given in row one of Table 1. We note that all the coefficients are positive, which implies that a 'physical' fit was achieved. In fact, all coefficients are between 0.99 and 1.66, which is very reasonable. For example, the coefficient c_4 is 1.29, which indicates that the MG3 basis captures only $1.00/1.29 = 78\%$ of the differences between the 6-31G(d) estimate of the MP2 component of the valence correlation energy and an infinite-basis estimate of that component.

From the mean errors in Table 2 we can see that the new MCG3 method reduces the mean unsigned error (MUE) and the root mean square error (RMSE) by ~ 8% and ~ 10%, respectively, as compared to G3. This increase in accuracy comes with a reduction in cost of ~ 50%.

Notice that we do not include the last term of Eq. (1), which is the higher-level correction (HLC). This term is problematic because it does not show proper dissociation. In particular, in the G3 method, it involves different constants for atoms than for molecules, and thus it cannot be used to predict continuous potential energy surfaces along bond breaking coordinates. Therefore, we do not include an HLC term in Eq. (2). For comparison with meth-

---

[1] This cost ratio was obtained on an SGI Origin 2000 as an average over the ratios for the six most expensive molecules in the G3 calculations, namely, SO_2, Si_2H_6, CH_3SH, SO, ClO, and CH_3Cl. Only the CPU time required for the single-point energy calculations is included in this estimate. The CPU time required for geometry optimizations in either method or for frequency calculations in standard G3 was omitted from the mean CPU time because the present method is designed for calculating potential energy surfaces, which requires a large number of single-point calculations.

Table 2
Mean errors[a] (kcal/mol)

Method	MSE[b]	MUE[c]	RMSE[d]
G2[e]	−0.23	1.21	1.72
PDG3	−6.98	6.98	7.63
G3[e]	−0.44	0.97	1.24
MMCG3	0.09	0.99	1.23
MCG3	0.03	0.90	1.10

[a]All mean errors are with respect to the 49 values of D_e for the training set, which is described in full in Ref. [6].
[b]Mean signed error.
[c]Mean unsigned (absolute) error.
[d]Root mean square error.
[e]Includes a higher-level correction (HLC).

ods that do not include and HLC we define proper-dissociation Gaussian-3 by

$$E(\text{PDG3}) = E(\text{G3}) - \Delta E(\text{HLC}). \qquad (6)$$

Table 2 shows that the MCG3 method has a mean unsigned error that is smaller than that for the PDG3 method by a factor of ∼ 8. This comparison is important because neither contains the HLC; therefore, this provides a comparison of the methods under the constraint that discontinuous empirical terms are not included in either.

A point of interest is to examine how well the MCG3 method works when the spin–orbit and core correlation effects are unknown or the core correlation is not well accounted for by the simple formula of Ref. [7]. To analyze this we define a 'minimal' MCG3 method, denoted MMCG3, in which the spin–orbit and core correlation terms are omitted (and the $\{c_i\}$ are re-optimized). The coefficients are given in the second row of Table 1, and the mean errors are given in Table 2. The MUE and RMSE for MMCG3 are approximately the same as for G3.

We note that the 49 molecules considered here have 113 bonds (counting a double or triple bond as one bond). Thus the mean unsigned error of 0.90 kcal/mol in atomization energies corresponds to a mean unsigned error of only 0.39 kcal/mol in bond energies.

We conclude that one can obtain higher accuracy than the G3 method for potential surfaces and thermochemistry of molecules containing first- and second-row atoms by using a method that dissociates properly (so it may be used for continuous potential energy surfaces) but which requires only about half as much computer time as G3.

Acknowledgements

References

[1] L.A. Curtiss, K. Raghavachari, P.C. Redfern, V. Rassolov, J.A. Pople, J. Chem. Phys. 109 (1998) 7764.

[2] L.A. Curtiss, K. Raghavachari, G.W. Trucks, J.A. Pople, J. Chem. Phys. 94 (1991) 7221.

[3] J.L. Durant, Am. Chem. Soc. Symp. Ser. 677 (1998) 267.

[4] P.L. Fast, M.L. Sánchez, J.C. Corchado, D.G. Truhlar, J. Chem. Phys. (in press).

[5] W.J. Hehre, L. Radom, P.v.R. Schleyer, J.A. Pople, Ab Initio Molecular Orbital Theory, Wiley, New York, 1986.

[6] P.L. Fast, J.C. Corchado, M.L. Sánchez, D.G. Truhlar, J. Phys. Chem. A 103 (1999) 3139.

[7] P.L. Fast, D.G. Truhlar, J. Phys. Chem. A 103 (1999) 3802.

[8] M.J. Frisch, G.W. Trucks, H.B. Schlegel, P.M.W. Gill, B.G. Johnson, M.A. Robb, J.R. Cheeseman, T. Keith, G.A. Petersson, J.A. Montgomery, K. Raghavachari, M.A. Al-Laham, V.G. Zakrzewski, J.V. Ortiz, J.B. Foresman, J. Cioslowski, B.B. Stefanov, A. Nanayakkara, M. Challacombe, C.Y. Peng, P.Y. Ayala, W. Chen, M.W. Wong, J.L. Andres, E.S. Replogle, R. Gomperts, R.L. Martin, D.J. Fox, J.S. Binkley, D.J. Defrees, J. Baker, J.P. Stewart, M. Head-Gordon, C. Gonzalez, J.A. Pople, Gaussian94 (Revision E.2), Gaussian, Inc., Pittsburgh, PA, 1995.