

New efficient multi-level electronic structure methods (MLSE n) for atomization energies and reaction energy barriers

Tsung-Hui Li, Chun-Hao Mou, Wei-Ping Hu *

Department of Chemistry and Biochemistry, National Chung Cheng University, Chia-Yi 621, Taiwan

Received 21 May 2004; in final form 18 June 2004

Available online 25 September 2004

Abstract

We have developed a new set of multi-level methods with scaled energies (MLSE n) for calculating the atomization energies and reaction energy barriers. The MP2 and QCISD(T) theories with correlation-consistent basis sets were used to obtain various energy components, and simple scaling schemes were used to optimize the scaling factors against a database of 109 atomization energies. The resulting methods were also tested on 21 energies of reactions, and 42 barrier heights. The most accurate methods give mean unsigned errors of 0.78 kcal/mol for atomization energies (MLSE4) and 0.61 kcal/mol for barrier heights (MLSE1).

© 2004 Elsevier B.V. All rights reserved.

1. Introduction

Achieving chemical accuracy (errors of 1–2 kcal/mol) in energies for a variety of molecules has long been the goal for the development of quantum chemical methods. However, it is evident that with a single theoretical level this kind of accuracy usually requires very high-level theories, such as MP4(SDTQ), QCISD(T) and CCSD(T), with very large basis sets, such as G3Large [1], aug-cc-pVTZ, and aug-cc-pVQZ [2–4]. The costs of these methods are prohibitively high except for very small molecules. Various more economical multi-level methods that have been developed in the past two decades, notably the Gaussian- n [1,5,6] and CBS methods [7], offered alternative approaches to close in on this goal. In most of these methods, a series of affordable energy calculations with a range of correlation levels and basis-set sizes are performed on a molecular geometry obtained from a lower but usually reliable theoretical level. Then various energy components from these calculations (and sometimes other empirically derived

high-level correction terms) are combined or extrapolated to account for the incomplete treatment of the correlation energies and the incompleteness of the basis-set sizes. While the accuracy of these multi-level methods has improved noticeably over the years, it was recently recognized that by using scaled energy components in the multi-level methods [8] higher accuracy and sometimes higher efficiencies can be achieved. In these so-called multi-coefficient methods, such as the more recent G3S [9], G3SX [10], MCCM/3 suite, and G3S/3 [11] methods, all the scaling factors for various energy components were optimized against databases of experimentally derived or very high-level theoretical energies. One additional important feature for the G3S and G3SX method is that the high-level correction (HLC) terms, which may cause discontinuity on the potential energy surfaces (PES), are no longer used and thus the methods are now suitable for obtaining PES data for chemical dynamics study.

In terms of applicability, two approaches of optimizing the scaling factors in these multi-coefficient methods are often used. The first is to optimize the coefficients against a database of a single type of energy, usually the atomization energy (AE). For example, in the original parameterization of the MCG3/2 [12] and

* Corresponding author. Fax: 886 5 272 1040.

E-mail address: chewph@ccu.edu.tw (W.-P. Hu).

MC-QCISD/2 [13] methods, the database used was an 82-molecule data set for AEs. The other approach is to optimize the coefficients against databases of several types of energies simultaneously. For example, the parameterization of the G3S was based on the G2/97 test set which contains 148 enthalpies of formation, 88 ionization potentials (IPs), 58 electron affinities (EAs) and 8 proton affinities, while the parameterization of the MCCM/3 suite was based on a training set (Database/3) of 109 AEs, 13 IPs, 13 EAs and 44 barrier heights. The advantage of the latter approach apparently is the wider applicability of the resulting methods. However, systematic errors for different types of energy calculation are usually different in nature. This means higher-quality basis sets and more sophisticated correlation correction schemes are required to achieve high accuracies for all types of energies. Furthermore, there is no unique way to balance the obtained accuracies for different types of energies. The optimized parameters and the performance may be sensitive to the future extension of the training set, especially for the types of energies that have fewer data in the training set. For the study of thermochemistry or reaction dynamics of neutral systems, the methods derived from first approach using an extended database of AEs are perhaps more adequate due to better performance/cost ratios. In this Letter, we present a simple and efficient multi-level scheme with scaled energies for calculating the AEs and reaction barrier heights for neutral systems. We used the correlation consistent basis sets by Dunning and coworkers [2–4] to obtain the required energy components. The correlation-consistent basis sets have been used in the earlier development of multi-coefficient methods (MCCM, MCSAC) [8] but not used in the more recent development probably due to the higher calculation cost. Instead, the most recent multi-coefficient methods such as MC-QCISD, G3S and MCG3 all used the Pople-type basis sets. In this Letter, we would like to show that it is not necessarily the case, and our new methods compete favorably in cost and efficiency with other recent multi-coefficient methods in predicting the AEs and barrier heights.

2. Method

We compute the electronic energies by the four related methods, MLSE n ($n = 1-4$), which are described below

$$\begin{aligned} E(\text{MLSE}_n) &= C_{\text{HF}} \times E(\text{HF/cc-pVDZ}) \\ &+ C_{\text{AHF}} \times [E(\text{HF/cc-pVTZ}) - E(\text{HF/cc-pVDZ})] \\ &+ C_{\text{E2}} \times [E(\text{MP2/cc-pVDZ}) - E(\text{HF/cc-pVDZ})] \\ &+ C_{\text{E34}} \times [E(\text{MP4SDQ/cc-pVDZ}) \end{aligned}$$

$$\begin{aligned} &- E(\text{MP2/cc-pVDZ})] \\ &+ C_{\text{QCI}} \times [E(\text{QCISD(T)/cc-pVDZ}) \\ &- E(\text{MP4SDQ/cc-pVDZ})] \\ &+ C_{\text{B}} \times \gamma_{\text{E2}} \times \{ [E(\text{MP2/cc-pVTZ}) \\ &- E(\text{HF/cc-pVTZ})] - [E(\text{MP2/cc-pVDZ}) \\ &- E(\text{HF/cc-pVDZ})] \} \\ &+ C_{+} \times [E(\text{MP2/aug-cc-pVDZ}) \\ &- E(\text{MP2/cc-pVDZ})] \end{aligned} \quad (1)$$

where $E(\text{Theory/Basis Set})$ denotes the single-point Born–Oppenheimer energy calculated with the combination of the particular theory and basis set. In the MLSE1 and MLSE2 methods the factor γ_{E2} is set to unity, while in MLSE3 and MLSE4, γ_{E2} is set to $[E(\text{MP2/cc-pVTZ}) - E(\text{HF/cc-pVTZ})]/[E(\text{MP2/cc-pVDZ}) - E(\text{HF/cc-pVDZ})]$, i.e., the ratio of the second-order energy corrections calculated by MP2/cc-pVTZ and MP2/cc-pVDZ methods. The function of this factor is to put even more weights on the contribution of basis-set extension for cases where the second-order energy corrections are more sensitive to the basis-set sizes. In the MLSE2 and MLSE4 methods, separate C_{E2} were used for closed-shell and open-shell species. It is noted that the use of γ_{E2} makes it very difficult to program the energy gradients for MLSE3 and MLSE4 methods, and the use of separate C_{E2} may give discontinuous PES upon bond dissociation. The treatment of the core-correlation energies and spin–orbital coupling is implicit in the current study. The training set used is the database developed by Truhlar and Lynch [11] which contains 109 zero-point exclusive AEs. Three sets of molecular structures which were optimized using MP2/cc-pVDZ, QCISD/MG3 [11], and B3LYP/6-31 + G(d,p) methods were used. The coefficients were determined to minimize the mean unsigned error (MUE) of the AEs. The resulting methods were then tested on the 42 classical barrier heights and 21 energies of reactions in a recent database (BH42/03) also developed by Truhlar and coworkers [14]. For these energy calculations, molecular structures (for the reactants, products, and transition states) optimized at MP2/cc-pVDZ and QCISD/MG3 levels were used. It is noted that the performance of these methods on the barrier heights and energies of reactions was *not* used to optimize the coefficients in the MLSE n methods.

3. Results

Optimized coefficients were obtained for each MLSE n method based on each geometry set. The resulting methods based on the MP2/cc-pVDZ, QCISD/MG3, and B3LYP/6-31 + G(d,p) geometries are conveniently labeled as MLSE n -A, MLSE n -B, and MLSE n -C, respectively. The MUEs on the AEs by the MLSE n methods are shown in Table 1. The G3S/3 and

Table 1
Mean unsigned errors on the atomization energies (kcal/mol)

	A	B	C
MLSE1	1.33	1.20	1.15
MLSE2	1.20	1.05	1.04
MLSE3	1.04	0.91	0.84
MLSE4	0.93	0.81	0.78
MCG3/3	1.04		
G3S/3	0.94		

MCG3/3 results [11] are also shown for comparison. Most of the MLSE n methods give very good accuracy on the AEs, and the most accurate MLSE4 methods give MUEs of only 0.78–0.93 kcal/mol, slightly better than the highly accurate G3S/3 method. The performance of the methods depends noticeably on the geometry sets used in the optimization of the coefficients. The difference is as large as 0.2 kcal/mol between MLSE n -A and MLSE n -C. It is interesting that MLSE n -C is slightly better than MLSE n -B. By using separate C_{E2} for open- and closed-shell species, the MUEs of MLSE2 is 0.11–0.15 kcal/mol lower than those of MLSE1. Introducing the γ_{E2} factor in Eq. (1) makes the MUEs of MLSE3 \sim 0.3 kcal/mol lower than those of MLSE1, which is very significant. The MLSE4 methods, which combines the features of MLSE2 and MLSE3, further lower the MUEs of MLSE3 by 0.06–0.11 kcal/mol. Table 2 shows the MUEs in calculating the 42 barrier heights and 21 energies of reactions. Since B3LYP is not a good theory to calculate the transition-state geometry of these hydrogen abstraction reactions, only MP2/cc-pVDZ and QCISD/MG3 structures were used in the comparison. The MLSE n -C coefficients, however, can still be applied here. Even though the current methods were not optimized against these barrier heights, the performance is still reasonably good with MUEs of \sim 1 kcal/mol. The MLSE1 methods have the lowest MUEs (0.61–0.85 kcal/mol) on the barrier heights, comparable to the G3S/3 result of 0.64 kcal/mol. It is also interesting to note that the combination of MP2/cc-pVDZ structures with MLSE n -C coefficients seems to give noticeably better results. However, we caution that with the databases

Table 2
Mean unsigned errors on the barrier heights and reaction energies

	Barrier heights			Energies of reactions		
	A	B	C ^a	A	B	C ^a
MLSE1	0.85	0.82	0.61 (0.79)	1.19	1.12	0.81 (0.82)
MLSE2	1.23	1.19	0.86 (0.99)	1.25	1.16	0.90 (1.00)
MLSE3	0.93	0.94	0.90 (1.07)	0.80	0.81	0.59 (0.65)
MLSE4	1.10	1.02	0.82 (1.03)	0.84	0.82	0.67 (0.75)
MCG3/3	0.90			0.82		
G3S/3	0.64			0.50		

^a Using the MP2/cc-pVDZ and QCISD/MG3 geometries with MLSE n -C coefficients. Data in parentheses are results obtained using QCISD/MG3 geometries.

used in the current study, the uncertainties on the barrier heights are probably significantly larger than those on the AEs, and thus the performance on the barrier heights calculation might not be as conclusive. The performance on the energies of reactions is also impressive, with MUEs of 0.59–0.81 kcal/mol by the MLSE3 methods and 0.67–0.84 kcal/mol by the MLSE4 methods. In comparison, the MCG3/3 and G3S/3 methods give MUEs of 0.82 and 0.50 kcal/mol, respectively, for the same set of energies. The optimized coefficients for the MLSE4 methods are listed in Table 3. Interestingly, the optimized C_+ are negative in values from -0.15 to -0.19 . The role of the C_+ term in the current methods is probably to compensate the over-correction by other basis-set extension or energy amplification terms. The values of other coefficients are all within the reasonable range. The relative computational costs for energies and gradients of MLSE n , MCG3/3, and G3S/3 are compared in Table 4. The costs are obtained by the CPU time for seven relatively large molecules in the database. The costs for energies of the MLSE n methods are \sim 70% and \sim 20% of the MCG3/3 and G3S/3 methods, respectively. The cost for gradients of the MLSE1 method is approximately the same as the MCG3/3 method and is an order of magnitude lower than the G3S/3 method. Thus, it seems that the new MLSE n , especially MLSE4, are efficient and accurate methods to calculate thermochemical data. The MLSE1 method which provides continuous PES and well-behaved gradients can be readily used for dynamics calculation. For cases where only energies are needed and the closed- or open-shell characters do not change during the courses

Table 3
Optimized coefficients in the MLSE4

	A	B	C
C_{HF}	0.99635	0.99729	0.99646
C_{AHF}	1.60715	1.48675	1.54872
C_{E2} (closed shell)	1.08275	1.07167	1.07506
C_{E2} (open shell)	1.08929	1.07780	1.07974
C_{E34}	0.99454	0.95478	0.99170
C_{QCI}	1.21432	1.28970	1.20217
C_B	1.13393	1.10267	1.15608
C_+	-0.19240	-0.15171	-0.14979

Table 4
Computational costs^a

	Energy	Gradient
MLSE n ^b	36 ^c (37) ^d	273 ^c (288) ^d
MCG3/3	51 (53)	335 (347)
G3S/3	171 (184)	2572 (2763)

^a CPU time relative to MP2/cc-pVDZ energy or gradient calculation for seven molecules: C_5H_5N , C_2Cl_4 , C_4H_4O , C_4H_4S , C_4H_5N , CF_3CN and $SiCl_4$.

^b The costs for energy calculation are the same for MLSE n ($n = 1-4$), and the cost for gradient calculation is for the MLSE1 method.

^c Total cost/7.

^d Average of the seven individual costs.

of reactions, other MLSE n methods can also be applied to obtain the relative energies on the PES.

Acknowledgement

This work is supported in part by the National Science Council of Taiwan, Grant No. NSC 92-2113-M194-019.

References

- [1] L.A. Curtiss, K. Raghavachari, P.C. Redfern, V. Rasslov, J.A. Pople, *J. Chem. Phys.* 109 (1998) 7764.
- [2] T.H. Dunning Jr., *J. Chem. Phys.* 90 (1989) 1007.
- [3] R.A. Kendall, T.H. Dunning Jr., R.J. Harrison, *J. Chem. Phys.* 96 (1992) 6796.
- [4] D.E. Woon, T.H. Dunning Jr., *J. Chem. Phys.* 98 (1993) 1358.
- [5] J.A. Pople, M. Head-Gordon, D.J. Fox, K. Raghavachari, L.A. Curtiss, *J. Chem. Phys.* 90 (1989) 5622.
- [6] L.A. Curtiss, K. Raghavachari, G.W. Trucks, J.A. Pople, *J. Chem. Phys.* 94 (1991) 7221.
- [7] J. Cioslowski (Ed.), *Quantum-Mechanical Prediction of Thermochemical Data: Understanding Chemical Reactivity Series*, vol. 22, Kluwer, Dordrecht, 2001.
- [8] P.L. Fast, J.C. Corchado., M.L. Sánchez, D.G. Truhlar, *J. Phys. Chem. A* 103 (1999) 5129.
- [9] L.A. Curtiss, K. Raghavachari, P.C. Redfern, J.A. Pople, *J. Chem. Phys.* 112 (2000) 1125.
- [10] L.A. Curtiss, P.C. Redfern, K. Raghavachari, J.A. Pople, *J. Chem. Phys.* 114 (2001) 108.
- [11] B.J. Lynch, D.G. Truhlar, *J. Phys. Chem. A* 107 (2003) 3898.
- [12] C.M. Tratz, P.L. Fast, D.G. Truhlar, *PhysChemComm* 2 (1999), Article 14.
- [13] P.L. Fast, D.G. Truhlar, *J. Phys. Chem. A* 104 (2000) 6111.
- [14] Y. Zhao, B.J. Lynch, D.G. Truhlar, *J. Phys. Chem. A* 108 (2004) 4786.