

Improved multi-level electronic structure methods (MLSE $n+d$) for atomization energies and reaction energy barriers

Tsung-Hui Li, Hui-Ru Chen, Wei-Ping Hu *

Department of Chemistry and Biochemistry, National Chung Cheng University, San Hsing 160, Min-Hsiung, Chia-Yi 621, Taiwan

Received 14 June 2005; in final form 5 July 2005

Available online 1 August 2005

Abstract

We have improved our multi-level electronic structure methods MLSE n for calculating the atomization energies and reaction energy barriers for neutral systems by using improved correlation-consistent basis sets for second-row elements. The re-parameterization of the improved methods MLSE $n+d$ was based on updated databases of 109 atomization energies, 38 hydrogen-transfer barrier heights, and 22 neutral reaction barrier heights from a recently developed database of non-hydrogen-transfer reactions. The improved methods perform very well on all three types of energies with mean unsigned errors of 0.70, 0.87, and 0.69 kcal/mol by the MLSE4 + d method.

© 2005 Elsevier B.V. All rights reserved.

1. Introduction

From research on electronic structure calculation in the past two decades, it has become increasingly apparent that the most efficient way to achieve a particular accuracy in relative energies is by multi-level approaches [1–12]. For example, the GAUSSIAN- n [6–8] and CBS methods [9,10] and their various variants have provided high-quality energies for thermochemical and kinetics at affordable costs. In these methods, additive corrections are applied to a base energy to account for the incomplete treatment of the correlation energies and the incompleteness of the basis-set sizes. While the accuracy and efficiency of these multi-level methods has improved over the years, the applicability of the most accurate methods is still limited to relatively small molecules. From a series study by Truhlar et al. [3,4,13,14], it was recently recognized that by using scaled energy components in the multi-level methods higher accuracy and sometimes higher efficiencies can be achieved. In these

so-called multi-coefficient methods, such as G3S/3 [4], G3SX [15], and those in MCCM/3 suite [4], all the scaling factors for various energy components were optimized against databases of experimentally derived or high-level theoretical energies. In the mean time, various efficient hybrid DFT methods, such as MPWB1K [16] and BB1K [17], have been parameterized to give accurate energies of reactions and barrier heights (BHs). Very recently, Truhlar et al. [18] successfully developed a series of MCCM-DFT methods which combined hybrid DFT calculation with the multi-coefficient methods and pushed the performance/cost ratios of the high-accuracy methods to a new level.

In a previous study [5], we have developed a set of efficient multi-level methods (MLSE n) for neutral systems. The MLSE n methods were parameterized to give very accurate atomization energies (AEs), and the most accurate MLSE4 method gives a mean unsigned error (MUE) of ~ 0.8 kcal for a set of 109 AEs. Even though the methods were not parameterized for BHs, the performance on a set of 42 BHs was still impressive with ~ 1 kcal MUEs. The performance of the MLSE4 method is comparable to MCG3/3 [4], the most accurate

* Corresponding author. Fax: +886 5 2721040.

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

method in the MCCM/3 suite, at slightly less cost. Analysis of the errors in the MLSE n methods showed that a major source of errors comes from the molecules containing second-row elements. To improve the accuracy for the second-row molecules, we used a newer version of the Dunning's correlation consistent basis sets, cc-pV($n + d$)Z [19] which include an additional set of d functions for the second-row elements. These basis sets were developed to improve the convergence behaviors for extrapolation to the complete basis set limit. In the current study, we report a new parameterization for the MLSE n methods, named MLSE $n + d$, using the improved basis sets. (Some minor modifications were also made as discussed in the following section.) The parameterization was carried out against an updated versions of an atomization energy database (MGAE109/04) [18], a hydrogen-transfer barrier height database (HTBH38/04) [18], and a set of 22 neutral reaction barriers from a recently developed database for non-hydrogen-transfer barrier heights (NHTBH38/04) [20].

2. Method

The MLSE $n + d$ ($n = 1-4$) electronic energies are calculated by

$$\begin{aligned}
 E(\text{MLSE}_{n+d}) = & C_{\text{HF}} \times E(\text{HF}/\text{cc-pV}(\text{D} + d)\text{Z}) \\
 & + C_{\Delta\text{HF}} \times [E(\text{HF}/\text{cc-pV}(\text{T} + d)\text{Z}) \\
 & - E(\text{HF}/\text{cc-pV}(\text{D} + d)\text{Z})] \\
 & + C_{E_2} \times [E_2/\text{cc-pV}(\text{D} + d)\text{Z}] \\
 & + C_{E_{34}} \times [E(\text{MP4SDQ}/\text{cc-pV}(\text{D} + d)\text{Z}) \\
 & - E(\text{MP2}/\text{cc-pV}(\text{D} + d)\text{Z})] \\
 & + C_{\text{QCI}} \times [E(\text{QCISD}(\text{T})/\text{cc-pV}(\text{D} + d)\text{Z}) \\
 & - E(\text{MP4SDQ}/\text{cc-pV}(\text{D} + d)\text{Z})] \\
 & + C_{\text{B}} \times \gamma_{E_2} \times [E_2/\text{cc-pV}(\text{T} + d)\text{Z} \\
 & - E_2/\text{cc-pV}(\text{D} + d)\text{Z}] \\
 & + C_{+} \times [E_2/\text{aug-cc-pV}(\text{D} + d)\text{Z} \\
 & - E_2/\text{cc-pV}(\text{D} + d)\text{Z}] + E_{\text{SO}},
 \end{aligned}
 \tag{1}$$

where $E(\text{theory}/\text{basis set})$ denotes the single-point Born–Oppenheimer energy calculated with the combination of the particular theory and basis set, and $E_2/\text{basis set}$ denotes the second-order energy correction calculated at the MP2 theory using the particular basis set. All the HF, MP, and QCI calculations are unrestricted for open-shell systems. For $n = 1$ or 2, the factor γ_{E_2} is set to unity, while for $n = 3$ or 4, γ_{E_2} is set to $[E_2/\text{cc-pV}(\text{T} + d)\text{Z}]/[E_2/\text{cc-pV}(\text{D} + d)\text{Z}]$, as discussed in the previous study [5]. In the MLSE2 + d and MLSE4 + d methods, separate C_{E_2} were used for closed-shell and open-shell species. Because of this separate treatment of open and closed-shell systems, these two methods do

not predict continuous potential energy surfaces along a bond dissociation coordinate that yields doublet or triplet radicals, whereas the MCCM and G3S methods predict continuous potential energy surfaces in all cases. The C_{+} terms in Eq. (1) was changed from $[E(\text{MP2}/\text{aug-cc-pVDZ}) - E(\text{MP2}/\text{cc-pVDZ})]$ in the previous version to be consistent with the Hartree–Fock energy extrapolation by the first two terms in Eq. (1). The treatment of the spin-orbital coupling is now explicit (E_{SO}) for MLSE1 + d and MLSE2 + d for selected open-shell species [21] and remains implicit ($E_{\text{SO}} = 0$) for MLSE3 + d and MLSE4 + d methods. The molecular geometries used were those optimized at QCISD/MG3 level by Truhlar et al. [4,20]. Two ways of parameterization of the MLSE $n + d$ methods were carried out. In the first (preliminary) parameterization, only the MGAE109/04 database was used as the training set. In the second parameterization, the MGAE109/04, HTBH38/04, and the 22 neutral reactions in NHTBH38/04 database were all used in the training set. Whether the barrier heights database should be used as the training set is debatable. Since barrier heights are not experimentally observable, the values in the databases have much larger uncertainties than in the AE database. On the other hand, to achieve a more balanced treatment for all the important regions on the potential energy surface and for the application to kinetics study; it is desirable to include barrier heights in the training set. Both parameterization approaches were carried out to minimize the unweighted MUEs relative to the database values.

3. Results

Table 1 shows the results from the first parameterization. The original MLSE n results were also included for comparison. For AEs, the improvements over the original methods are 0.33, 0.30, 0.14, and 0.12 kcal/mol for $n = 1 - 4$, respectively. Even the 0.12 kcal/mol improvement for MLSE4 is significant because it means a reduction of the sum of the magnitudes of all the errors by 13 kcal/mol. For hydrogen transfer reactions, the performance of the new methods are about the same as the original methods, ~ 1 kcal/mol of MUEs on BHs and ~ 0.5 kcal/mol on energies of reaction (ΔE_{rxn}). For non-hydrogen-transfer reactions, the performance of the new methods is slightly better, with MUE of ~ 0.8 kcal/mol for BHs and ~ 0.4 kcal/mol for ΔE_{rxn} by MLSE4 + d method. To the best of our knowledge, this is the first reported test using the N^7 multi-level methods on the NHTBH/04 database. The MLSE1 + d and MLSE2 + d methods benefit significantly (~ 0.2 kcal in MUEs) from the explicit consideration of the spin-orbital energies. However, the E_{SO} term does not bring further improvement for MLSE3 + d and MLSE4 + d and is thus not used in these methods.

Table 1
Mean unsigned errors (kcal/mol) using MGAE109/04 as the training set

Method	AE (109)	HT		NHT	
		$\Delta V^\ddagger(38)$	$\Delta E_{\text{rxn}}(19)$	$\Delta V^\ddagger(22)$	$\Delta E_{\text{rxn}}(11)$
MLSE1	1.21	0.70	0.83	1.54	0.81
MLSE2	1.07	0.92	0.71	1.08	0.82
MLSE3	0.92	0.90	0.53	1.15	0.62
MLSE4	0.81	0.96	0.45	0.90	0.32
MLSE1 + <i>d</i>	0.88	1.04	0.56	1.08	0.65
MLSE2 + <i>d</i>	0.77	0.91	0.44	0.93	0.57
MLSE3 + <i>d</i>	0.78	1.18	0.60	1.02	0.59
MLSE4 + <i>d</i>	0.69	0.95	0.52	0.83	0.37

Table 2 shows the results from the second parameterization in which all AEs and BHs data were used in the training set. The results by two other multi-level methods and two DFT methods were also included for comparison. The MUEs by the most accurate MLSE4 + *d* are 0.70, 0.87, and 0.69 kcal/mol for AEs, HTBHs, and NHTBHs, respectively. Since there is no unique way to balance the MUEs of the three sets of energies, different results would be obtained by using different weights on the MUEs of the energies. In the current study, equal weighting was used for all 169 data. The 0.70-kcal/mol MUEs of AEs are among the best results reported in the literature, slightly smaller than the values obtained by the new MCG3-MPWB methods [18]. The 0.87-kcal/mol MUE of the HTBHs is similar to the results obtained by the MCG3/3 and is ~ 0.3 -kcal/mol less accurate than the MCG3-MPWB methods. The 0.69-kcal/mol MUEs of the NHTBHs is quite satisfactory considering the extended range of the BHs (2–106 kcal/mol) in the database. We also used the MCG3/3 method to test the same set of reactions, and the MUE for the 22 BHs was found to be 1.05 kcal/mol. In Table 2, we see that the two hybrid DFT methods parameterized for kinetics, BB1K and MPWB1K, give satisfactory results on BHs (MUEs of 1.2–1.7 kcal/mol). However, the performance of the N^7 methods is twice as good for BHs and almost an order of magnitude better for AEs and ΔE_{rxn} .

Table 2
Mean unsigned errors (kcal/mol) using all 169 data as the training set

Method	AE (109)	HT		NHT	
		$\Delta V^\ddagger(38)$	$\Delta E_{\text{rxn}}(19)$	$\Delta V^\ddagger(22)$	$\Delta E_{\text{rxn}}(11)$
MLSE1 + <i>d</i>	0.94	0.82	0.45	0.89	0.52
MLSE2 + <i>d</i>	0.79	0.78	0.41	0.77	0.58
MLSE3 + <i>d</i>	0.82	0.95	0.51	0.82	0.50
MLSE4 + <i>d</i>	0.70	0.87	0.50	0.69	0.39
MCG3-MPWB ^a	0.75	0.54			
MCG3/3 ^b	1.04	0.90	0.82	1.05	1.16
BB1K ^c	6.31	1.16	1.28	1.55	2.67
MPWB1K ^c	4.57	1.29	1.31	1.69	2.97

^a From [18].

^b From [4] and this work.

^c From [16,20].

The coefficients obtained from the second parameterization for the MLSE n + *d* methods are listed in Table 3. The cost of MLSE n + *d* calculation on small to medium-sized molecules is modest. For example, it takes less than 20 min to calculate C₅H₅N on an entry-level workstation. We also found that the calculation can be speeded up significantly by 2-way SMP with $\sim 170\%$ efficiency. We estimate that systems up to 10–15 heavy atoms can be handled by today's high-end workstations or supercomputers.

The accuracy of the multi-level methods has been increasing over the years. It may be worthwhile to discuss the performance limits, if any, that the multi-level methods can achieve. Currently, most of the accurate multi-level methods are empirical in nature, and they require experimental or high-level theoretical data for parameterization. Thus, the performance of current multi-level methods is limited by the accuracy of the data in the training set. Most of the thermochemical data used to derive the AEs in the database have uncertainties on the order of ~ 1 or ~ 0.2 kcal/mol. If we assume that the evaluation of the thermo- and zero-point energies have uncertainties of ~ 0.1 kcal/mol, the uncertainty of the zero-point exclusive atomization energies in the training set would be ~ 0.3 kcal/mol. The uncertainty on the BHs is more difficult to estimate. The values in the HTBH38/04 were derived by a combination of experimental/theoretical approach [22]. If we assume there is on average 20% uncertainty on the experimental rate constants at 600 K, at which most of the experimental and calculated rate constants were compared in building the database, this would cause ~ 0.2 kcal/mol uncertainty on the energy barriers. If we further assume that both the errors on the potential energy surfaces and on the approximations used in the dynamics methods contribute 20% uncertainty in the calculated rate constants, then the uncertainty on the BHs would be ~ 0.6 kcal/mol. Most of the values in NHTBH38/04 database were obtained by the W1 or W2 method [11]. The calculated values are believed to have uncertainty within 1 kcal/mol [20]. Since many of the reactants and products of the reactions in NHTBH database also appear in the MGAE109/04 database, the difference between

Table 3
Optimized coefficients^a

Coefficients	MLSE1 + <i>d</i>	MLSE2 + <i>d</i>	MLSE3 + <i>d</i>	MLSE4 + <i>d</i>
C _{HF}	1.00021	0.99912	0.99915	0.99824
C _{AHF}	1.18052	1.24905	1.30781	1.34648
C _{E2} (closed shell)	0.93604	0.94275	1.06976	1.07863
C _{E2} (open shell)	0.93604	0.94823	1.06976	1.08412
C _{E34}	0.81331	0.82680	0.91724	0.92781
C _{QC1}	1.35886	1.25715	1.28114	1.20327
C _B	1.87486	1.96266	1.01357	1.05754
C ₊	-0.42589	-0.53605	-0.19522	-0.22709

^a The coefficients were obtained by optimizing to all 169 data.

the calculated forward and reverse BHs can be compared with the ΔE_{rxn} obtained from the AE database. In most cases, the ΔE_{rxn} obtained from the two databases differ by less than 0.3 kcal/mol. For the $\text{H} + \text{F}_2 \rightarrow \text{HF} + \text{F}$ reaction, however, the difference is 0.9 kcal/mol. Since the error on the ΔE_{rxn} is the difference of the signed errors of the forward and reverse barriers, it seems reasonable to place the uncertainty of the BHs in the database at ~ 0.5 kcal/mol. While the above estimations are semi-quantitative at best, they can nonetheless give us rough ideas on the performance limits one can expect from the multi-level methods using these databases as training sets. For example, there seems still room for improvement for calculating AEs by the multi-level approaches. Also, the comparison of performance is only meaningful when the difference in MUEs by two methods is a significant fraction of the estimated uncertainty. For example, 0.1–0.2-kcal/mol difference in MUEs for BHs might not be very meaningful.

Acknowledgment

This work is supported by the National Science Council of Taiwan, Grant No. NSC 93-2113-M-194-002.

References

- [1] F.B. Brown, D.G. Truhlar, Chem. Phys. Lett. 117 (1985) 307.
- [2] M.S. Gordon, D.G. Truhlar, J. Am. Chem. Soc. 108 (1986) 5412.
- [3] P.L. Fast, J.C. Corchado, M.L. Sánchez, D.G. Truhlar, J. Phys. Chem. A 103 (1999) 5129.
- [4] B.J. Lynch, D.G. Truhlar, J. Phys. Chem. A 107 (2003) 3898.
- [5] T.-H. Li, C.-H. Mou, W.-P. Hu, Chem. Phys. Lett. 397 (2004) 364.
- [6] J.A. Pople, M. Head-Gordon, D.J. Fox, K. Raghavachari, L.A. Curtiss, J. Chem. Phys. 90 (1989) 5622.
- [7] L.A. Curtiss, K. Raghavachari, G.W. Trucks, J.A. Pople, J. Chem. Phys. 94 (1991) 7221.
- [8] L.A. Curtiss, K. Raghavachari, P.C. Redfern, V. Rasslov, J.A. Pople, J. Chem. Phys. 109 (1998) 7764.
- [9] J.A. Montgomery Jr., M.J. Frisch, J.W. Ochterski, G.A. Petersson, J. Chem. Phys. 112 (2000) 6532.
- [10] J. Cioslowski (Ed.), Quantum-Mechanical Prediction of Thermochemical Data: Understanding Chemical Reactivity Series, vol. 22, Kluwer, Dordrecht, 2001.
- [11] S. Parthiban, J.M.L. Martin, J. Chem. Phys. 114 (2001) 6014.
- [12] A.D. Boese, M. Oren, O. Atasoylu, J.M.L. Martin, J. Chem. Phys. 120 (2004) 4129.
- [13] C.M. Tratz, P.L. Fast, D.G. Truhlar, PhysChemComm 2 (1999) 14.
- [14] P.L. Fast, D.G. Truhlar, J. Phys. Chem. A 104 (2000) 6111.
- [15] L.A. Curtiss, P.C. Redfern, K. Raghavachari, J.A. Pople, J. Chem. Phys. 114 (2001) 108.
- [16] Y. Zhao, D.G. Truhlar, J. Phys. Chem. A 108 (2004) 6908.
- [17] Y. Zhao, B.J. Lynch, D.G. Truhlar, J. Phys. Chem. A 108 (2004) 2715.
- [18] Y. Zhao, B.J. Lynch, D.G. Truhlar, Phys. Chem. Chem. Phys. 7 (2005) 43.
- [19] T.H. Dunning Jr., K.A. Peterson, A.K. Wilson, J. Chem. Phys. 114 (2001) 9244.
- [20] Y. Zhao, N. González-García, D.G. Truhlar, J. Phys. Chem. A 109 (2005) 2012.
- [21] P.L. Fast, J. Corchado, M.L. Sanchez, D.G. Truhlar, J. Phys. Chem. A 103 (1999) 3139.
- [22] B.J. Lynch, P.L. Fast, M. Harris, D.G. Truhlar, J. Phys. Chem. A 104 (2000) 4811.