Accurate multi-level electronic structure methods (MLSE-DFT) for atomization energies and reaction energy barriers

Yi-Lun Sun, Tsung-Hui Li, Jien-Lian Chen, Kuo-Jui Wu, Wei-Ping Hu *

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

Abstract

We have developed a set of new multi-level electronic structure methods by including energies calculated from several density functional theory methods. The parameterization of the improved methods MLSE-DFT was based on updated databases of 109 atomization energies, 38 hydrogen-transfer barrier heights, and 22 neutral non-hydrogen-transfer reaction barrier heights. The best method, MLSE-TPSS1KCIS, performed impressively on the above three types of energies with mean unsigned errors of 0.62, 0.55, and 0.69 kcal/mol, respectively. We found that the hybrid versions of DFT are not absolutely necessary, and the performance can be improved significantly using two different basis sets in DFT calculation.

© 2007 Elsevier B.V. All rights reserved.

1. Introduction

One of the important traditional goals for developing accurate electronic structure methods is to obtain relative molecular energies to within chemical accuracy or ~1 kcal/mol at affordable cost. With the rapid advances in computer technologies and the innovation in quantum chemical methods in the past decade, this goal seems to be finally realized for small to medium-sized (~10 heavy atoms) molecules by the so-called multi-level methods [1–15]. 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. For example, the GAUSSIAN-n [6–8] and CBS methods [9,10] and their various variants have provided high-quality energies for thermochemistry and thermochemical kinetics at reasonable costs. 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 meth-

ods, 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. Very recently, Truhlar et al. found that incorporating hybrid DFT energies in the multi-coefficient methods can substantially increase the accuracy with only modest increase in the computational cost [16]. In fact, the MCG3-MPWB, MCG3-TS, and MCG3-MPW methods [16] which include the energies from the hybrid versions of the MPWB95, TPSSKCIS, and MPWPW91 functionals are among the most accurate all-purpose multi-level electronic structure methods currently available.

In previous studies [5,17] we have developed a set of efficient multi-level methods (MLSEn, MLSE$n + d$) for neutral systems. The MLSEn methods were parameterized to give very accurate atomization energies (AEs) and barrier heights (BHs). The most accurate MLSE4 + d method gives a mean unsigned error (MUE) of 0.70 kcal/mol for a set of 109 AEs, 0.87 kcal/mol for a set of 38 hydrogen-transfer BHs, and 0.69 for a set of 22 non-hydrogen-transfer BHs. The performance of the MLSE4 + d method for neutral systems is comparable to MCG3-MPWB at slightly less cost [5,17]. In light of the success of the MCG3-DFT

* Corresponding author. Fax: +886 5 272 1040.
  E-mail address: chewph@ccu.edu.tw (W.-P. Hu).

method, we would like to reformulate our MLSE$n + d$ methods by including energy components from hybrid DFT calculation. The reformulation, the increase in accuracy, and the comparison to other multi-level methods will be discussed.

2. Method

The MLSE-DFT electronic energies are calculated by

$$
\begin{aligned}
E(\text{MLSE-DFT}) = {}& C_{\text{WF}}\{E(\text{HF/cc-pV(D}+d)\text{Z}) \\
& + C_{\Delta\text{HF}}[E(\text{HF/cc-pV(T}+d)\text{Z}) \\
& - E(\text{HF/cc-pV(D}+d)\text{Z})] \\
& + C_{\text{E2}}[E2/\text{cc-pV(D}+d)\text{Z}] \\
& + C_{\text{E34}}(E(\text{MP4SDQ/cc-pV(D}+d)\text{Z}) \\
& - E(\text{MP2/cc-pV(D}+d)\text{Z})] \\
& + C_{\text{QCI}}[E(\text{QCISD(T)/cc-pV(D}+d)\text{Z}) \\
& - E(\text{MP4SDQ/cc-pV(D}+d)\text{Z})] \\
& + C_{\text{B}}[E2/\text{cc-pV(T}+d)\text{Z} \\
& - E2/\text{cc-pV(D}+d)\text{Z}] \\
& + C_{\text{HF+}}[E(\text{HF/aug-cc-pV(D}+d)\text{Z}) \\
& - E(\text{HF/cc-pV(D}+d)\text{Z})] \\
& + C_{\text{E2+}}[E2/\text{aug-cc-pV(D}+d)\text{Z} \\
& - E2/\text{cc-pV(D}+d)\text{Z}]\} \\
& + (1 - C_{\text{WF}})\{E(\text{DFT}X/\text{cc-pV(D}+d)\text{Z}) \\
& + C_{\text{B1}}[E(\text{DFT}X/\text{cc-pV(T}+d)\text{Z} \\
& - \text{DFT}X/\text{cc-pV(D}+d)\text{Z}]\} + E_{\text{SO}} \quad (1)
\end{aligned}
$$

where E (theory/basis set) denotes the single-point Born–Oppenheimer energy calculated with the combination of the particular theory and basis set, and $E2$/basis set denotes the second-order energy correction calculated at the MP2 theory using the particular basis set. The basis sets used in Eq. (1) are Dunning's correlation consistent basis sets with one additional set of d functions for the second-row elements [18]. The DFTX denotes a particular hybrid DFT method with X% of Hartree–Fock exchange energy. Four hybrid DFT methods were used in this work, and they are MPW1B95 [19,20], MPW1PW91 [19], TPSS1KCIS [21,22], and B1B95 [20,23]. The '1's in the above functionals signify that they are one-parameter (X) hybrid DFT with the names of exchange and correlation functionals before and after '1', respectively. These hybrid DFT methods were found to perform very well for thermochemistry and thermochemical kinetics [16,24–26]. The first three methods have also been used in the MCG3-DFT methods by Truhlar and coworkers [16]. The treatment of the spin–orbital coupling is explicit (E_{SO}) for selected open-shell species [27]. The molecular geometries used were those optimized at QCISD/MG3 level by Truhlar et al. [4,28]. The MLSE energy in Eq. (1) is partitioned into a wavefunction (WF) part and a DFT part. In comparison to our previous MLSE$n + d$ methods, a term represents the contribution by the diffuse functions to the Hartree–

Fock energy (the $C_{\text{HF+}}$ term) in the wavefunction part was added. The complication of using different C_{E2} for open- and closed-shell species, and the nonlinear scaling factor γ_{E2} was eliminated. As a result, the new MLSE-DFT methods will predict continuous potential energy surface in all cases [5]. The DFT part consists of a base energy using the cc-pV(D + d)Z basis set and a basis-set correction term using the cc-pV(T + d)Z basis set. We found that using basis sets of two different sizes in the DFT part may improve the overall accuracy significantly. The parameterization of the coefficients (including X) in Eq. (1) was based on the MGAE109/05 [29], HTBH38/04 [16], and the 22 neutral reactions in NHTBH38/04 databases [28], and was carried out to minimize the unweighted mean unsigned errors (MUEs) of all data relative to the database values. The electronic structure calculation was performed using the GAUSSIAN 03 program [30].

3. Results

Table 1 shows the MUEs obtained in the current work and by G3SX, MCG3/3, MCG3-MPWB [16,31], and MLSE1 + d [17] methods. The MLSE method in the table is based on Eq. (1) without using DFT energies ($C_{\text{WF}} = 1.0$). The MLSE method outperformed the MLSE1 + d method by 0.07 kcal/mol on the average MUE because of the additional $C_{\text{HF+}}$ term in Eq. (1), and of using the updated calculated energies and database values. Including DFT energies in Eq. (1) lowers the average MUE by ∼0.2 kcal/mol, as shown in Table 1 for the four MLSE-DFT methods. The MLSE-TPSS1KCIS method gives the best overall performance of 0.61 kcal/mol on the average MUE. In comparison, including the MPW1B95 energies in the MCG3 method (MCG3-MPWB) lowers the average MUE by ∼0.3 kcal/mol. The average MUE obtained by the MLSE-TPSS1KCIS method is 0.12 and 0.17 kcal/mol lower than those by the MCG3-MPWB and G3SX methods, respectively. Our new methods predict lower MUEs for atomization energies than the two very accurate meth-

Table 1
Mean unsigned errors (kcal/mol) obtained by various multi-level methods

Method	AE (109)	HTBH (38)	NHTBH (22)	Average[e]
G3SX[a]	0.85	0.67	0.66	0.78
MCG3/3[a]	1.04	0.84	1.05	1.00
MCG3-MPWB[a]	0.75	0.54	0.97	0.73
MLSE1 + d[b]	0.94	0.82	0.89	0.91
MLSE[c]	0.75	1.07	0.83	0.83
MLSE-TPSS1KCIS[d]	0.62	0.55	0.69	0.61
MLSE-MPW1B95[d]	0.66	0.51	0.76	0.64
MLSE-B1B95[d]	0.68	0.54	0.73	0.65
MLSE-MPW1PW91[d]	0.66	0.56	0.75	0.65

[a] From Ref.[16,31].
[b] From Ref. [17].
[c] Current result without using energies from DFT.
[d] Current work.
[e] Unweighted average of all 169 data.

ods. The MLSE-DFT methods also predict significantly lower non-hydrogen-transfer barrier heights than the MCG3-MPWB method.

The optimized coefficients for the four MLSE-DFT methods are listed in Table 2. The C_{WF} in these methods are ~0.8, that is, the wavefunction part contributes approximately 80% of its total energy to the MLSE-DFT energy while the DFT part contributes approximately 20% of its total energy. These percentages are similar to those obtained in the MCG3-DFT methods [16]. The optimized C_{B1} values are close to 2.0 in the four new methods, and this signifies the importance of scaling the basis-set effects even in the DFT part. In fact, if only the cc-pV(T + d)Z basis set was used in the DFT part, the average MUEs would be ~0.04 kcal/mol higher for these MLSE-DFT methods. Thus, the use of the dual basis sets in the DFT part of the MLSE-DFT methods contributes ~20% of the improvement in accuracy over the MLSE method. The optimized X values for the MLSE-TPSS1KCIS and MLSE-MPW1B95 methods are 15 and 24, respectively. These are compared to the somewhat higher X values of 19 and 34 in the MCG3-TS and MCG3-MPWB methods [16], respectively. Fig. 1 shows the obtained average MUEs as a function of the X values for the MLSE-TPSS1KCIS method. Interestingly, the accuracy of the method is *not* sensitive to the X values in the range tested. Using $X = 0$ or pure DFT, the average MUE only increases ~0.01 kcal/mol. Similar behaviors were also found for other MLSE-DFT methods. Thus it seems that in the MLSE-DFT methods, the use of hybrid versions of DFT might not be absolutely necessary. This is in contrast to the cases when DFT was used alone to calculate the molecular energies, where the hybrid DFT methods clearly outperform the pure DFT [26].

The computational cost of the MLSE-TPSS1KCIS method is compared to that of the G3SX, MCG3, MCG3-MPWB, and MLSE methods in Table 3. In fact, all the MLSE-DFT methods have very similar cost. For medium-sized molecules, the cost of the MLSE-DFT methods is almost identical to that of the MCG3-MPWB method, and it is ~40% higher than that of the MLSE method. The computational cost of the G3SX method is
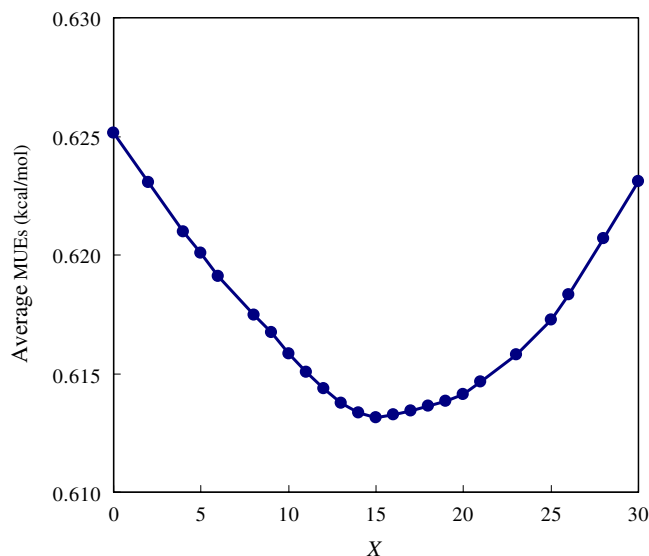

Fig. 1. Average mean unsigned errors (kcal/mol) obtained by the MLSE-TPSS1KCIS method as a function of X.

Table 3
Computational cost[a]

	Cost[b]
G3SX	9875
MCG3/3	1907
MCG3-MPWB	2334
MLSE	1673
MLSE-TPSS1KCIS	2374

[a] Tested using computers with Intel E6600 CPU and a version of GAUSSIAN 03 program optimized for the hardware architecture.
[b] Total CPU time in seconds to calculate C_5H_5N, C_2Cl_4, C_4H_4O, C_4H_4S, C_4H_5N, CF_3CN, and $SiCl_4$.

significantly higher. Of course, the relative cost depends strongly on the computer hardware and software used. On machines with two physical CPUs, we found that the two-way SMP efficiencies of our new methods were up to ~170%, and on machines with a single dual-core CPU, the efficiencies were up to ~160%. Similar efficiencies were found for the G3SX and MCG3-DFT methods. Thus, these multi-level methods should benefit significantly from parallel computing.

4. Concluding remarks

With the recent development of the MCG3-DFT and MLSE-DFT methods, and to a lesser extent the doubly hybrid DFT methods [32], it seems apparent that the most cost effective way to obtain highly accurate molecular energies is to combine energies calculated from multi-level wavefunction theory and density functional theory. (Hybrid versions of DFT methods are not absolutely necessary in the MLSE-DFT methods.) The average errors of the MCG3-DFT and MLSE-DFT methods are well below the so-called chemical accuracy, and they are

Table 2
Optimized coefficients of the MLSE-DFT methods using different hybrid functionals

	TPSS1KCIS	MPW1B95	B1B95	MPW1PW91
C_{WF}	0.80153	0.78033	0.79501	0.83861
$C_{\Delta HF}$	0.95769	0.98709	1.00975	1.02709
C_{E2}	0.96311	0.94093	0.93187	0.95799
C_{E34}	0.79367	0.85816	0.84394	0.86172
C_{QCI}	1.25570	1.27802	1.24388	1.26839
C_B	1.75278	1.90565	2.02000	1.82790
C_{HF+}	0.31310	0.25695	0.26427	0.27190
C_{E2+}	−0.70904	−0.81470	−0.82929	−0.68866
C_{B1}	2.33721	2.10151	1.91304	2.22611
X	15	24	20	18

approximately equal to $k_B T$ at room temperature. Due to the uncertainties in the experimentally derived database values, it is unlikely to further improve the accuracies of these empirical multi-level methods dramatically with similar approaches. However, with even better density functionals that might be developed in the near future, it is hopeful that some of the expensive wavefunction calculations can be eliminated without significant sacrifice in the accuracy, and the accurate multi-level methods can then be applied to much larger chemical systems.

Acknowledgements

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, Phys. Chem. Comm. 2 (1999) (article 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, B.J. Lynch, D.G. Truhlar, Phys. Chem. Phys. 7 (2005) 43.
[17] T.-H. Li, H.-R. Chen, W.-P. Hu, Chem. Phys. Lett. 412 (2005) 430.
[18] T.H. Dunning Jr., K.A. Peterson, A.K. Wilson, J. Chem. Phys. 114 (2001) 9244.
[19] C. Adamo, V. Barone, J. Chem. Phys. 108 (1998) 664.
[20] A.D. Becke, J. Chem. Phys. 104 (1996) 1040.
[21] J. Tao, J.P. Perdew, V.N. Staroverov, G.E. Scuseria, Phys. Rev. Lett. 91 (2003) 146401.
[22] T. Toulouse, A. Savin, C. Adamo, J. Chem. Phys. 117 (2002) 10465.
[23] A.D. Becke, Phys. Rev. A 38 (1988) 3098.
[24] Y. Zhao, D.G. Truhlar, J. Phys. Chem. A 108 (2004) 6908 (MPWB1K).
[25] Y. Zhao, B.J. Lynch, D.G. Truhlar, J. Phys. Chem. A 108 (2004) 2715 (BB1K).
[26] Y. Zhao, D.G. Truhlar, J. Phys. Chem. A 109 (2005) 5656.
[27] P.L. Fast, J. Corchado, M.L. Sanchez, D.G. Truhlar, J. Phys. Chem. A 103 (1999) 3139.
[28] Y. Zhao, N. González-García, D.G. Truhlar, J. Phys. Chem. A 109 (2005) 2012.
[29] Y. Zhao, N.E. Schultz, .D.G. Truhlar, J. Chem. Theory Comput. 2 (2006) 364.
[30] M.J. Frisch et al., GAUSSIAN 03, Revision D.02, Gaussian, Inc., Wallingford CT, 2004.
[31] J. Zheng, Y. Zhao, D.G. Truhlar, J. Chem. Theory Comput. 3 (2007) 569.
[32] B.J. Lynch, D.G. Truhlar, J. Phys. Chem. A 108 (2004) 4786.