Energies of Organic Molecules and Atoms in Density Functional Theory

Gábor I. Csonka,*,[1] Adrienn Ruzsinszky,[1] Jianmin Tao,[2] and John P. Perdew,*,[2]

[1]*Department of Inorganic Chemistry, Budapest University of Technology and Economics, H-1521 Budapest, Hungary and*
[2]*Department of Physics and Quantum Theory Group, Tulane University, New Orleans, Louisiana 70118*

Abstract: In the assessment of density functional approximations for the exchange-correlation energy, great weight is usually given to the accuracy of molecular atomization energies, or the enthalpies of formation constructed from calculated atomization energies. Here we show that a recent non-empirical functional, the meta-generalized gradient approximation of Tao, Perdew, Staroverov, and Scuseria, achieves remarkably accurate atomization energies even for the larger organic molecules of the G3-3 test set. But we also present strong evidence that most of the error of previous non-empirical functionals resides in the energy of the free atom, and so cancels out of typical reaction energies. Finally, we suggest that enthalpies of formation calculated without any reference to the free atoms would provide a fairer assessment of the performance of approximate density functionals.

keywords: density functional
           exchange-correlation energy
           molecular atomization energies
           enthalpies of formation
           meta-generalized gradient approximations

Introduction

Density functional theory (DFT) is now widely used for electronic structure calculations in both quantum chemistry and condensed matter physics. In the assessment of approximate functionals for the exchange-correlation energy, great weight is usually given to the accuracy of molecular atomization energies, or of enthalpies of formation[1-3] based on calculated atomization energies. In this communication, we have two points to make: (1) Nonempirically constructed functionals are finally producing chemically useful atomization energies for small and large molecules.[4] (A comprehensive self-consistent study with large basis sets for all G3 molecules will be reported elsewhere[5]). A recent such functional,[4-7] whose moderate errors cancel those of the modest 6-311G(d,p) Gaussian basis set, achieves extraordinary accuracy even for the larger organic molecules of the G3-3 set,[1] providing consistent energies from atoms to small and large molecules. (2) For the molecules studied, most of the error of earlier fully or nearly non-empirical functionals[8,9] resides in the energies of the free atoms,[10] and cancels out in the computation of energies of reactions[11] that do not involve free atoms. To a considerable extent, smaller sources of error such as basis-set incompleteness or relativistic

effects can also cancel out. DFT-calculated enthalpies of formation can often be greatly improved by introducing one empirical parameter to represent the energy of each type of free atom. Since DFT calculations can be performed for any standard state (e.g., graphite), we suggest that accurate non-empirical DFT molecular enthalpies of formation might be found from direct calculations which require neither atomic nor experimental input. This suggestion is consistent with earlier conclusions by Delley,[12] and has been implemented to compute the heats of formation of compound solids, e.g., Ref. 13, although not always with the expected accuracy.

Our work is related to that of Ref. 5 in the following ways: Ref. 5 was conceived as a comparative assessment of the new Tao-Perdew-Staroverov-Scuseria (TPSS) functional,[4-7] its hybrid,[5] and 14 other functionals against the G2 set of molecular enthalpies of formation and other experimental data. After we discovered the remarkable accuracy of TPSS for most of the larger organic molecules of the G3-3 set via nonselfconsistent calculations, the full G3-3 set was incorporated into the self-consistent assessment of Ref. 5. Ref. 5 does not analyze or identify the atomic source of the errors in standardly-calculated enthalpies of formation and does not analyze the effect of the use of a simpler 6-311G(d,p) basis set instead of the more expensive 6-311++G(3df,3pd) basis set.

We begin by discussing some problems with approximate functionals, focusing on the most popular empirically-constructed ones such as B3LYP. Earlier studies[1] have shown that generalized gradient approximation (GGA) and GGA-hybrid methods, that were parametrized for the enthalpies of formation of the relatively small molecules selected from the G2/97 test set, fail seriously for larger molecules of the G3-3 test set. The G3-3 test set[1,2] contains 75 enthalpies of formation with a quoted uncertainty of ±1 kcal or less for 13 molecules without hydrogen, 16 hydrocarbons, 44 substituted hydrocarbons, and 2 radicals. The G3-3 molecules contain three to ten non-hydrogen atoms, e.g. naphthalene.[1]

The following significant increase in errors of the DFT methods for the species in the G3-3 test set has been noted in the literature: The B3LYP, BLYP and SVWN/6-311+G(3df,2p) mean absolute errors (m.a.e.) for the G3-3 subset are over two times larger than those of the G2/97 test set (8.21 vs 3.29 kcal/mol, 13.32 vs 6.17 kcal/mol, and 216.49 vs 91.93 kcal/mol, respectively).[1] The increase in the errors is largest for the hydrocarbons and substituted hydrocarbons. If the test set is changed from G2/97 to G3-3, the m.a.e. of the B3LYP/6-311+G(3df,2p) method increases from 2.92 to 9.64 kcal/mol for the hydrocarbons, and from 2.22 to 7.15 kcal/mol for substituted hydrocarbons, respectively.[1] The increase in B3LYP

error with molecular size is evident from comparison of the errors for the n-alkanes methane, ethane,… up to octane (see Figure 1).[1]

Results

We performed a series of calculations on the 50 most problematic hydrocarbons and substituted hydrocarbons (containing only H, C, N, O, and F) from the G3-3 test set,[2] for which the mean atomization energy was 1500 kcal/mol. We used SVWN5, PBE GGA,[8] PKZB[9] meta-GGA, and the new Tao-Perdew-Staroverov-Scuseria (TPSS) meta-GGA[4] functionals. Except for PKZB (which has one empirical parameter), these are non-empirical density functionals constructed to satisfy exact constraints on the exchange-correlation energy, with no fitting to chemical data. These functionals fall respectively on the first (SVWN5), second (PBE GGA), and third (PKZB and TPSS meta-GGA) rungs of Jacob's Ladder of approximations,[4,14] where higher rungs have more (and more sophisticated) local ingredients. The performance of the PKZB meta-GGA for atomization energies was found to be satisfactory for the G2/97 test set.[15] However, it was observed that the geometries and frequencies calculated with the PKZB meta-GGA are worse than with the PBE or its hybrid,[15] and the PKZB meta-GGA provides a poor description of hydrogen bonds.[5,16] The errors are[7] mainly from PKZB exchange. By introducing a requirement that the meta-GGA exchange potential be finite at the nucleus for one- or two-electron densities (an exact constraint lost in GGA), an improved meta-GGA can be obtained. This is the key new element in the TPSS meta-GGA, which could explain its improved description of bond formation.[5]

All our reported results were obtained with our modified version of the CADPAC code,[17] using the standard 6-311G(d,p) basis set and self consistent PBE GGA orbitals. We have selected this moderate basis set, because for large molecules economic basis sets must be tested and our separate basis-set dependence study has shown that this triple zeta polarized basis set works well with the TPSS functional, achieving a remarkable cancellation of moderate errors not only for the G3-3 molecules but also for smaller ones. In a study of 21 typical G2/97 molecules, we compared our calculations to those much closer to the basis-set limit[4] and found that the TPSS mean overbinding (4 kcal/mol) of a molecule was cancelled almost perfectly by the mean underbinding due to our use of the moderate 6-311G(d,p) basis set. A similar comparison shows that the intrinsic TPSS mean overbinding for our 50 G3-3 molecules is 5 kcal/mol. Since B3LYP and PKZB underbind G3-3 molecules, their errors are not cancelled but reinforced by basis-set limitations. To evaluate the atomic energy, we follow the standard approach, using the broken-symmetry real orbitals of lowest energy.

For the larger G3-3 molecules, we have selected the B3LYP/6-31G(d) equilibrium geometries from our database[18] because the conventional MP2 geometry optimization is impractical beyond a certain size of the molecules, and the GGA-hybrid equilibrium geometries have proven to be quite useful.[1] Comparison of the energies obtained for the B3LYP and MP2 equilibrium geometries shows that the energy differences for those two equilibrium geometries are negligible for classical molecules, in agreement with ref. 1. In order to calculate the enthalpy of formation from the atomic and molecular energies, we used the G3X[19] zero-point energy and thermal corrections calculated from scaled B3LYP/6-31G(2df,p) vibrational frequencies, using a scale factor of 0.9854. The calculations were performed on IBM RS6000/AIX and SUN/SunOS 5.9 systems.

For the set of 50 molecules selected from the G-3 subset and for 21 molecules containing the same five elements, but selected from the G2/97, we observed the following: In agreement with the earlier observations, the m.a.e. of SVWN5/6-311G(*d,p*) is about three times larger for the 50 larger molecules than for the 21 smaller ones. For PBE and PKZB the m.a.e. increases from 10.9 to 32.2 kcal/mol, and from 5.5 to 17.1 kcal/mol, respectively. This is the usual behavior of the DFT error, which scales with increasing molecular size or atomization energy. The PBE shows systematic overbinding and the PKZB shows systematic underbinding. However, with the same basis set the TPSS meta-GGA shows a quite different, and much more desirable behavior. For the TPSS meta-GGA, the m.a.e. **decreases** from 3.2 (21 molecules of the G2/97 test set) to 2.2 kcal/mol (50 molecules of the G3-3 subset). The TPSS meta-GGA does not show under- or overbinding, the average error is very close to zero, and the r.m.s. deviation is small (2.7 kcal/mol) as shown in Figure 2.

For larger molecules, the errors in the reaction enthalpies are considerably more important than the errors in the enthalpies of formation. Inconsistencies between the calculated atomic and molecular energies will lead to large errors in calculated enthalpies of formation. However, for reaction energies of chemical interest, the inconsistencies between molecular and atomic energies do not play any role. Consistent behavior of the calculated molecular total energies of a method makes it possible to obtain good reaction enthalpies, after the usual zero-point and thermal corrections (cf. homodesmic and isodesmic reactions).

To check the internal consistency of the molecular energies, we suggest a procedure that helps to evaluate the performance of a method for relative molecular energies. We use corrected atomic energies in order to obtain the best possible agreement in a least square sense with experimental enthalpies of formation as shown in equation 1 for a given molecule, M:

$$\Delta H_f^0(M) = E_T(M) + E_{ZP}(M) + \Delta E_{therm}(M)$$

$$+ \sum_{A \in M} [\Delta H_f^0(A, \text{Expt.}) - E_{Tcorr}(A) - \Delta E_{therm}(A, \text{Expt.})], \qquad (1)$$

where $E_T(M)$ is the calculated total electronic energy, $E_{ZP}(M)$ is the calculated zero point vibration energy (ZPE), and the $\Delta E_{therm}(M)$ is the calculated difference between the enthalpies of the molecule at $T = 298.15$ and 0 K (calculated from the molecular heat capacity). The summation is over all atoms (A) of the molecule. The $\Delta H_f^0(A, \text{Expt.})$ are the experimental standard enthalpies of formation of the constituent atoms of molecule M, the $E_{Tcorr}(A)$ are the corrected total energies of these free atoms, and the $\Delta E_{therm}(A, \text{Expt.})$ are the experimental differences between the enthalpies at $T = 298.15$ and 0 K (calculated from the elemental heat capacities). The calculated and the corrected atomic energies are shown in Table 1. The corrected atomic energies were obtained from the least-square fit of the calculated enthalpies of formations to the experimental ones for the 50 molecules shown in the Appendix. Note that the fit corrections to the SVWN5 atomic energies are all negative, as expected since the atom is more inhomogeneous than the molecule.

After these corrections, the errors of the calculated enthalpies of formation arising due to inconsistencies between the atomic and molecular energies are eliminated. The remaining errors show the contribution of a given molecule to a reaction enthalpy error (as atomic energies fall out in a chemical reaction enthalpy calculation). We observe that SVWN5, PBE, and PKZB require considerable correction (cf. Figure 1), and thus there are serious inconsistencies between atomic and molecular energies in these methods. We also note that, after elimination of these inconsistencies via simple atomic corrections, PBE and PKZB showed an excellent performance for the 50 selected molecules of the G3-3 test set. Atomic corrections reduced the m.a.e. of SVWN5 from 217.7 to 4.4, of PBE from 32.2 to 2.2, of PKZB from 17.1 to 2.3, and of TPSS from 2.2 to 2.0 kcal/mol. The small improvement for TPSS shows that TPSS with the 6-311G(d,p) basis set provides consistent energies from atoms to large organic molecules (Figure 1).

Our empirical atomic energy corrections are reminiscent of the "atom equivalents" used by Mole et al.[20] for the hydrocarbons, following an idea by Dewar and Storch.[21] They differ, however, in important ways: Our corrections represent principally the free-atom energies, and do not include any other enthalpy term. They do to some extent absorb errors due to basis-set limitations (as also found in ref. 20) or neglect of relativistic effects. Still, they clearly

demonstrate that even methods that predict poor atomization energies can predict accurate energy differences among molecules.

Appendix

List of the 50 G3-3 molecules used in the test:

Methyl allene (C_4H_6)

Isoprene (C_5H_8)

Cyclopentane (C_5H_{10})

n-Pentane (C_5H_{12})

Neopentane (C_5H_{12})

1,3 Cyclohexadiene (C_6H_8)

1,4 Cyclohexadiene (C_6H_8)

Cyclohexane (C_6H_{12})

n-Hexane (C_6H_{14})

3-Methyl pentane (C_6H_{14})

Toluene ($C_6H_5CH_3$)

n-Heptane (C_7H_{16})

Cyclooctatetraene (C_8H_8)

n-Octane (C_8H_{18})

Naphthalene ($C_{10}H_8$)

Acetic acid methyl ester (CH_3COOCH_3)

t-Butanol ($CH_3)_3COH$

Aniline ($C_6H_5NH_2$)

Phenol (C_6H_5OH)

Divinyl ether (C_4H_6O)

Tetrahydrofuran (C_4H_8O)

Cyclopentanone(C_5H_8O)

Benzoquinone($C_6H_4O_2$)

Pyrimidine($C_4H_4N_2$)

Butanedinitrile(NC-CH_2-CH_2-CN)

Pyrazine($C_4H_4N_2$)

Acetyl acetylene (CH_3-C(=O)-C°CH)

Crotonaldehyde (CH_3-CH=CH-CHO)

Acetic anhydride (CH_3-CO-O-CO-CH_3)

Isobutane nitrile((CH_3)$_2$CH-CN)

Methyl ethyl ketone(CH_3-CO-CH_2-CH_3)

Isobutanal((CH_3)$_2$CH-CHO)

1,4-Dioxane($C_4H_8O_2$)

Tetrahydropyrrole(C_4H_8NH)

Nitro-s-butane (CH_3-CH_2-CH(CH_3)-NO_2)

Diethyl ether(CH_3-CH_2-O-CH_2-CH_3)

Dimethyl acetal(CH_3-CH(OCH_3)$_2$)

t-Butylamine ((CH_3)$_3$C-NH_2)

N-methyl pyrrole (cyc-C_4H_4N-CH_3)

Tetrahydropyran($C_5H_{10}O$)

Diethyl ketone (CH_3-CH_2-CO-CH_2-CH_3)

Isopropyl acetate (CH_3-C(=O)-O-CH(CH_3)$_2$)

Piperidine(cyc-$C_5H_{10}NH$)

t-Butyl methyl ether((CH_3)$_3$C-O-CH_3)

1,3-Difluorobenzene($C_6H_4F_2$)

1,4-Difluorobenzene($C_6H_4F_2$)

Fluorobenzene (C_6H_5F)

Di-isopropyl ether ((CH_3)$_2$CH-O-CH(CH_3)$_2$)

Ethane,-hexafluoro- (C_2F_6)

Azulene ($C_{10}H_8$)

Table 1.

Calculated and corrected H, C, N, O, and F atomic energies (in hartrees) for four DFT methods combined with 6-311G(d,p) basis set. The corrected atomic energies, E_{Tcorr}(A) of eq. 1, were obtained from a least-squares fit to 50 experimental enthalpies of formation. The molecules are shown in the Appendix.

| | SVWN5 | | PBE | | PKZB | | TPSS | |
	calculated	corrected	calculated	corrected	calculated	corrected	calculated	corrected
H	-0.47790	-0.48493	-0.49962	-0.49533	-0.49593	-0.49579	-0.49974	-0.50028
C	-37.46306	-37.51002	-37.79287	-37.80737	-37.78627	-37.78178	-37.86115	-37.86020
N	-54.12659	-54.16772	-54.52706	-54.54207	-54.50246	-54.50107	-54.60761	-54.60809
O	-74.51388	-74.55905	-74.99902	-75.01102	-74.95013	-74.94790	-75.09370	-75.09236
F	-99.09057	-99.12723	-99.65300	-99.66067	-99.57678	-99.57309	-99.75670	-99.75725

Figures
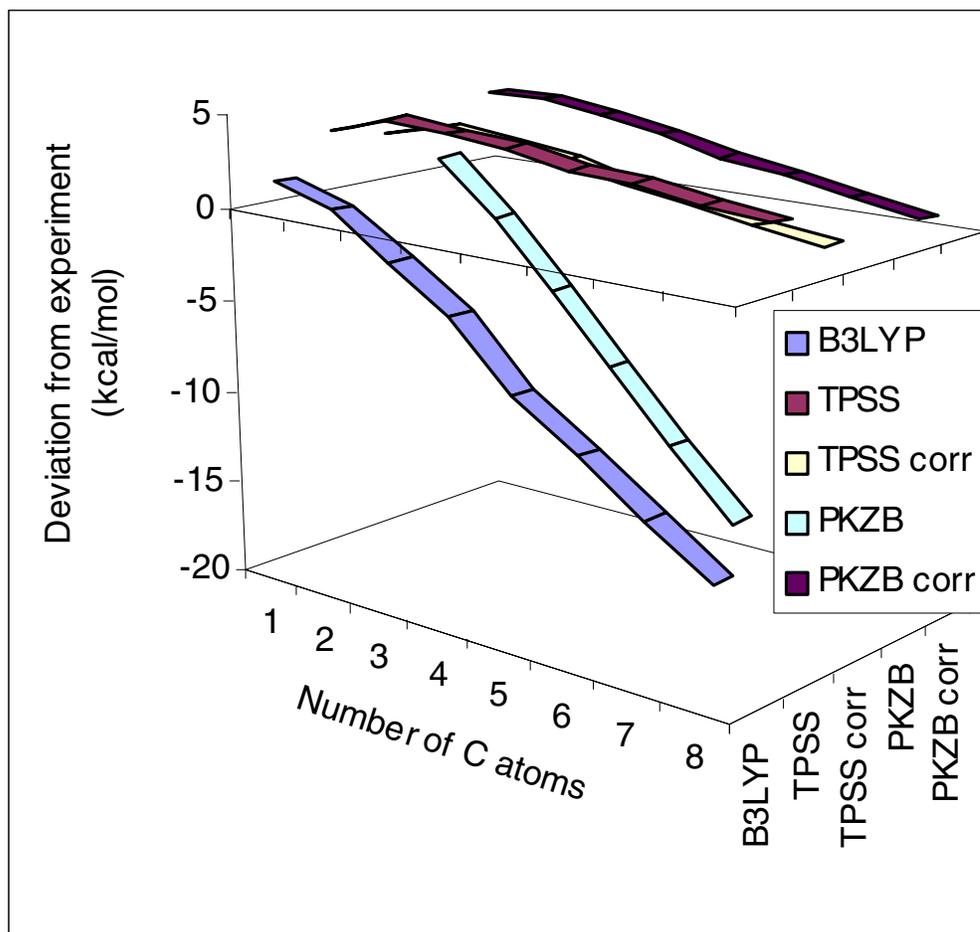
Figure 1. Deviation (experiment – theory) for enthalpies of formation of n-alkanes containing 1 to 8 C atoms calculated with various DFT methods and the modest 6-311G(d,p) basis set. The B3LYP deviations are from ref. 1. Corr denotes the use of corrected atomic energies.
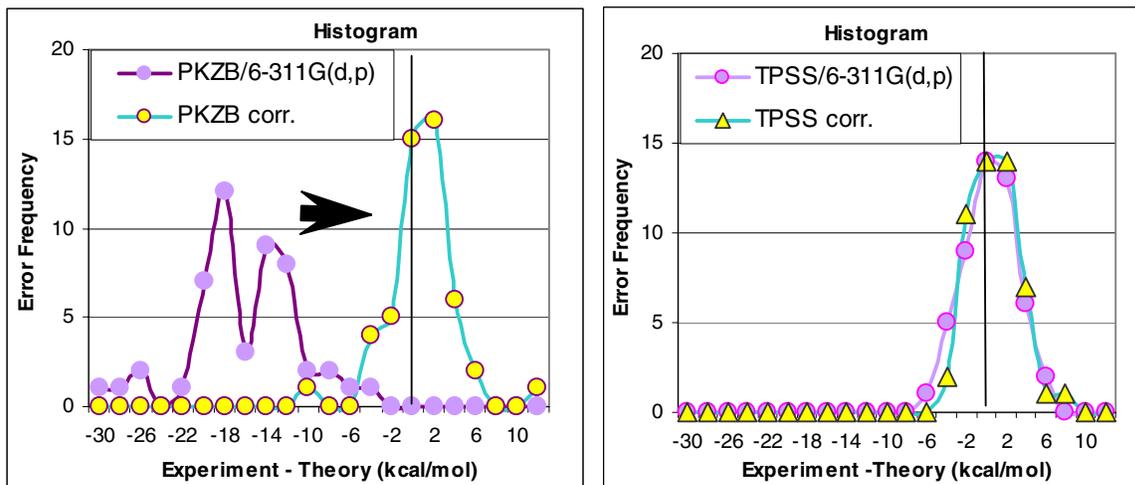
Figure 2. The histograms of the errors of the uncorrected and corrected (corr.) PKZB and the new non-empirical TPSS meta-GGA for 50 large hydrocarbons and substituted hydrocarbons (containing H, C, N, O, and F) selected from the G3-3 test set, using the modest 6-311G(d,p) basis set.

References

1   L. A. Curtiss, K. Raghavachari, P. C. Redfern, J. A. Pople, J. Chem. Phys. 112, 7374 (2000).

2   L. A. Curtiss, http://chemistry.anl.gov/compmat/g3-99.htm.

3   M. Ernzerhof, G. E. Scuseria, J. Chem. Phys. 110, 5029 (1999).

4   J. Tao, J. P. Perdew, V. N. Staroverov, G. E. Scuseria, Phys. Rev. Lett. 91, 146401 (2003).

5   V. N. Staroverov, G. E. Scuseria, J. Tao, J. P. Perdew, J. Chem. Phys. 119, 12129 (2003).

6   V.N. Staroverov, G.E. Scuseria, J. Tao, and J.P. Perdew, Phys. Rev. B (to appear).

7   J.P. Perdew, J. Tao, V.N. Staroverov, and G.E. Scuseria, submitted.

8   J. P. Perdew, K. Burke, M. Ernzerhof, Phys. Rev. Lett. 77, 3865 (1996).

9   J. P. Perdew, S. Kurth, A. Zupan, P. Blaha, Phys. Rev. Lett. 82, 2544 (1999).

10 E.J. Baerends, V. Branchadell, M. Sodupe, Chem. Phys. Lett. 265, 481 (1997).

11 A.C. Scheiner, J. Baker, J.W. Andzelm, J. Comp. Chem. 18, 775 (1997).

12 B. Delley, J. Chem. Phys. 113, 7756 (2000).

13 M. Fuchs, J.L.F. Da Silva, C. Stampl, J. Neugebauer, and M. Scheffler, Phys. Rev. B 65, 245212 (2002).

14 J.P. Perdew and K. Schmidt, in "Density Functional Theory and its Application to Materials", edited by V.E. Van Doren, K. Van Alsenoy, and P. Geerlings (American Institute of Physics, 2001).

15 C. Adamo, M. Ernzerhof, G. E. Scuseria, J. Chem. Phys. 112, 2643 (2000).

16 A. D. Rabuck, G. E. Scuseria, Theor. Chem. Acc. 104, 439 (2000).

17 N. C. Handy, R. D. Amos, I. L. Alberts, J. S. Andrews, et al., CADPAC6, Cambridge 1995.

18 G. I. Csonka, A. Ruzsinszky, J. Oláh, C. Van Alsenoy, J. Mol. Struct. THEOCHEM 589-590, 1 (2002).

19 L. A. Curtiss, P. C. Redfern, K. Raghavachari, J. A. Pople, J. Chem. Phys. 114, 108 (2001).

20 S.J Mole, X. Zhou, R. Liu, J. Phys. Chem. 100, 14665 (1996).

21 M.J.S. Dewar, D. M. Storch, J. Am. Chem. Soc. 107, 3898 (1985).