

Pragmatic ab initio prediction of enthalpies of formation for large molecules: accuracy of MP2 geometries and frequencies using CCSD(T) correlation energies

Robert W. Molt Jr. · Alexandre Bazanté ·
Thomas Watson Jr. · Rodney J. Bartlett

Received: 21 September 2012 / Accepted: 22 October 2012 / Published online: 20 November 2012
© Springer-Verlag Berlin Heidelberg 2012

Abstract We have addressed the accuracy of calculating the enthalpy of formation of an arbitrary single reference molecule using practical ab initio methodologies. It is known that MP2 geometries with a triple zeta basis set are almost as reliable as CCSD(T) geometries. It is also known that CCSD(T) correlation energies, with basis extrapolation, feature chemical accuracy for single-reference molecules. We investigate what accuracy one might expect in enthalpies of formation from a MP2 geometry, MP2 harmonic vibrational frequencies, a CCSD(T) correlation energy using triple zeta basis sets. It is far from obvious, a priori, as to which error source contributes most significantly. We observe that the accuracy in calculating enthalpies of formation of single-reference molecules with this protocol is 4 kcal mol⁻¹; our error analysis shows this comes almost exclusively from the correlation energy basis extrapolation, rather than errors intrinsic to MP2.

Keywords Coupled cluster · Enthalpies of formation

Introduction

We have recently reported detailed coupled cluster (CC) studies of the nitramine explosives RDX, HMX, and CL-20 [1–3]. These are large molecules for CC level applications, though ACES III makes it possible to obtain accurate energies and consequently enthalpies of formation. However, it remains impractical to search potential energy surfaces at the CCSD(T) level, instead suggesting that geometry optimization and harmonic vibrational frequencies be obtained by using second

order perturbation theory, MBPT2 (MP2 if a Hartree-Fock reference). This raises the issue of what errors are introduced into the calculation and how to keep them under control.

Use of a many-body calculation without guidance to the limits of its accuracy is pointless; for each intended study of some property (ionization energies, multiplet gaps, etc.), a benchmark is necessary. Given that all methodologies have some uncertainty, if one's answers are within that uncertainty, this needs to be documented. Choices in the many-body method used should be determined by the accuracy required. One should not choose CCSDT to get all the conformers of met-enkephalin, as this is overkill. One should not choose MD simulations to estimate energy differences known to be on the order of 1 kcal mol⁻¹. This is in contrast to the logical fallacy of “This is the best we can afford to do.”

We are interested in the accuracies of the enthalpies of formation (ΔH_f°) for arbitrary compounds using ab initio methodologies (here defined to mean a parameter-less methodology, systematically improvable as needed). Geometries of molecules calculated using second-order many-body perturbation theory [4] (MBPT(2); otherwise known as Möller-Plesset perturbation theory if the reference chosen happens to be Hartree-Fock) have been benchmarked to be accurate to within $\pm 0.01 \text{ \AA}$ for single reference molecules [5] in a triple zeta basis set, with a reasonably tight distribution of errors. As such, MBPT(2) geometries are trustworthy for an arbitrary single-reference molecule. Because the methodology is size-extensive formally, one need not worry about the size of the system studied [6]. It is also desirable because, unlike most (but not all) styles of Kohn-Sham density functional theory (KS-DFT [7–10]), it is not biased to work only on some test set of molecules used in its parameterization, succeeding markedly in some cases, and failing in others. This is another virtue of single-reference ab initio methods: they generally can be trusted to have the same accuracy

R. W. Molt Jr. (✉) · A. Bazanté · T. Watson Jr. · R. J. Bartlett
Quantum Theory Project,
Gainesville, FL 32611, USA
e-mail: r.molt.chemical.physics@gmail.com

independent of the system, as long as the system is genuinely single-reference and the basis set is comparable for the intended purpose. Many popular functionals are ambiguous in their quality outside of the molecules against which they were parameterized (references to M06 family [11], B3LYP [12], TPSS [13], B97D [14], CAM-B3LYP [15]), compared to the first-principles functionals (LDA [16], *PBE* β [17], VWN [18], PBE [19], and others). Vibrational harmonic frequencies using MBPT(2) are generally qualitatively correct for single-reference molecules, but perform rather poorly for the absolute numbers predicted for vibrations [20], again relative to single-reference systems. The accuracy of coupled cluster with perturbative triples CCSD(T) [21] or CCSD[T] [22, 23] for geometries and frequencies is even better [5, 20], but prohibitively expensive for systems on the order of 20–200 atoms. The use of CCSD(T) with a triple-zeta basis set for single-point energies, however, is quite manageable with highly parallelized software for coupled cluster [24].

Accuracies for enthalpies of formation have been studied in the use of composite methods and KS-DFT [25, 26]. Accuracies of enthalpies of formation also exist for situations in which one provides the correct geometries, zero-point energies, anharmonic contributions, etc. to MBPT(2), CCSD, and CCSD(T) estimates of the internal energy [27]. Additionally, the studies associated with the HEAT algorithm [28] describe calculating the enthalpy of formation without any experimental input for systems with below ~20 atoms, using very sophisticated calculations. Our work in this paper will closely follow the work of reference [27], in general; Helgaker et al. studied ab initio enthalpies in terms of only correlation energies for given geometries and frequencies. We study the case of not having the luxury of geometries and frequencies and how the errors may compound.

Table 1 Error in enthalpies of formation CCSD(T)/cc-pVTZ/MBPT(2)/6-311++G(d,p); units are kcal mol⁻¹. Ground state always chosen

Fluorine	1.3
Hydrogen fluoride	2.2
Methylene	10.3
Azanone	8.1
Nitrogen	7.8
Water	4.5
Carbon monoxide	2.4
Hydrogen peroxide	6.2
Ammonia	5.7
Hydrogen cyanide	2.3
Formaldehyde	3.7
Carbon dioxide	4.2
Ethyne	1.5
Methane	2.0
Ethene	2.0

We report to the first decimal place in this table for the sake of computational comparison; we do *not* make the claim that these calculations are accurate to within tenths of a kcal mol⁻¹

Table 2 Error in zero-point energies MBPT(2)/6-311++G(d,p); kcal mol⁻¹

Fluorine	0.02
Hydrogen	0.13
Hydrogen fluoride	0.09
Methylene	0.36
Azanone	0.03
Nitrogen	1.47
Water	0.14
Carbon monoxide	0.06
Hydrogen peroxide	0.24
Ammonia	0.24
Hydrogen cyanide	0.06
Formaldehyde	0.11
Carbon dioxide	0.00
Ethyne	0.06
Methane	0.28
Ethene	0.11

Methods

Our goal is to consider the enthalpies of formation obtained from the following protocol:

- 1) Optimization of geometry using MBPT(2)/ triple-zeta basis set of choice
- 2) Calculation of zero-point energies and energetic minimum confirmation using MBPT(2)/ triple-zeta basis set of choice
- 3) Use of CCSD(T)/cc-pVTZ for the internal energy as a single-point calculation

This protocol is of interest given the accuracy of MBPT(2) for geometries and CCSD(T) for correlation energies. The MBPT(2) geometries and frequencies may be improved by using CCSD(T) in the same basis set [5]; the internal

Table 3 Error in enthalpies of formation using MP2 geometries, frequencies, zero-point energies, and CCSD(T) complete-basis set extrapolation

Fluorine	0.6
Hydrogen	0.9
Hydrogen fluoride	1.9
Methylene	2.0
Azanone	2.6
Nitrogen	0.4
Water	0.9
Carbon monoxide	0.3
Hydrogen peroxide	1.0
Ammonia	0.2
Hydrogen cyanide	0.5
Formaldehyde	0.3
Carbon dioxide	2.1
Ethyne	0.6
Methane	0.4
Ethene	0.6

energy estimate may be improved by using a quadruple-zeta basis set and performing an extrapolation [5]. Thus, if our studies on new molecules require higher accuracy, we know what we need to do. This is of particular importance in the study of large organic explosives, in which the enthalpy of formation is very desirable to predict computationally (and thus safely) for newly proposed fuels and explosives [26].

Results

However, it is not clear which of the three above necessary calculations for the enthalpy of formation is weakest, nor what the net accuracy is for these three approximations built upon one another. We thus calculate the enthalpies of formation for a 15 molecule test set with results in Table 1. We emphasize that since all of our methods are size-extensive, the number of particles in question are irrelevant; specifically, in the sense that the method scales properly with particle number. The same error will exist regardless of the number of atoms, ignoring basis set sensitivity issues. We thus may choose small systems without loss of generality. A general introduction to the issues of energy reference problems in electromagnetics vs. thermodynamics is provided, given the difference conventions [29]. All calculations follow the protocol for calculation enthalpies of formation as described in reference [1]. All optimizations are performed using MBPT(2)/6-311++G(d,p), as well as the harmonic vibrational frequencies; single-point energies are calculated using CCSD(T)/cc-pVTZ. All calculations use ACES II [30].

The average un-signed error is 4 kcal mol⁻¹, with a standard deviation of 3. This should be compared to previous work which established that the error in the enthalpies of formation of CCSD(T)/cc-pVTZ internal energies is 8 kcal mol⁻¹ when given the *correct* zero-point energies and geometries [5]. Thus the results include some cancellation of error.

We next consider the error in the zero-point energies calculated using MBPT(2)/6-311++G(d,p); compared to the experimental values [31–43] in Table 2.

The mean un-signed error is generally negligible, less than 0.2 kcal mol⁻¹ (although the error in nitrogen is oddly high). Previous work showed that if one performs a basis set extrapolation for the CCSD(T) internal energy, the mean un-signed error reduces to less than a kcal mol⁻¹ [27]. It would thus seem that improving accuracy is solely a matter of improving the correlation energy estimate. This is somewhat surprising; given the inaccuracy of harmonic vibrational frequencies using MBPT(2), it was suspected that the zero-point energies might also be poor.

As a consistency check, we performed a complete basis set extrapolation on the CCSD(T) energies from the MP2 structures, and used the MP2 harmonic frequencies and ZPE

to calculate the enthalpies of formation. If our claims are correct, then the only error should be in the geometries from MP2 rather than CCSD(T) (which we claim is a small source of error above). The results are given in Table 3. We used the complete basis set extrapolation scheme of Helgaker et al. [27] in order to be consistent. All of these calculations were done dropping the core orbitals (logically necessary given that we do not have core functions in the aforementioned basis).

The average un-signed error is 1.0 kcal mol⁻¹, in perfect agreement with previous work [27]. We have thus demonstrated control over all error variables for typical molecules at equilibrium geometries. We do, however, make no claim about non-standard cases of multi-reference systems.

Conclusions

As ever, the correlation energy remains the critical issue in single-reference quantum chemistry. One may confidently rely on MP2 geometries and MP2 harmonic vibrational frequencies in terms of thermochemistry contributions if using a triple-zeta basis set. The average un-signed error for our ab initio protocol of using MP2 geometries/frequencies with CCSD(T) correlation energies in triple-zeta basis sets is 4 kcal mol⁻¹, with a standard deviation of 3.

Acknowledgments This research is sponsored through the Army Research Office (ARO).

References

1. Molt RW, Watson T, Lotrich VF, Bartlett RJ (2011) RDX geometries, excited states, and revised energy ordering of conformers via MP2 and CCSD(T) methodologies: insights into decomposition mechanism. *J Phys Chem A* 115:884–890. doi:10.1021/jp109695v
2. Molt RW, Watson T, Lotrich VF, Bartlett RJ (2012) Conformers of CL-20 explosive and ab initio refinement using perturbation theory; implications to detonation mechanisms. *J. Phys. Chem A*, accepted, in press
3. Molt RW, Watson T, Lotrich VF, Bartlett RJ (2012) The great diversity of HMX conformers: probing the PES using CCSD(T) *J Chem Phys* In press
4. Bartlett RJ, Musial M (2007) Coupled-cluster theory in quantum chemistry. *Rev Mod Phys* 79:291–352. doi:10.1103/RevModPhys.79.291
5. Helgaker T, Gauss J, Jørgensen P, Olsen J (1997) The prediction of molecular equilibrium structures by the standard electronic wave functions. *J Chem Phys* 106:6430–6440. doi:10.1063/1.473634
6. Bartlett RJ, Shavitt I (2009) Many-body methods in chemistry and physics. Cambridge University Press, Cambridge, Ch 6
7. Kohn W, Sham LJ (1965) Self-consistent equations including exchange and correlation effects. *Phys Rev* 140:A1133–A1138. doi:10.1103/PhysRev.140.A1133
8. Hohenberg P, Kohn W (1964) Inhomogeneous electron gas. *Phys Rev* 136:B864–B871. doi:10.1103/PhysRev.136.B864

9. Cramer C (2008) Essentials of computational chemistry, 2nd edn. Wiley, West Sussex, Ch 8
10. Martin RM (2004) Electronic structure. Cambridge University Press, Cambridge
11. Zhao Y, Truhlar DG (2008) The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, non-covalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theor Chem Acc* 120:215–241. doi:10.1007/s00214-007-0310-x
12. Becke AD (1993) Density-functional thermochemistry. III. The role of exact exchange. *J Chem Phys* 98:5648–5652. doi:10.1063/1.464913
13. Tao JM, Perdew JP, Staroverov VN, Scuseria GE (2003) Climbing the density functional ladder: nonempirical meta-generalized gradient approximation designed for molecules and solids. *Phys Rev Lett* 91:146401-1–146401-4. doi:10.1103/PhysRevLett.91.146401
14. Grimme S (2006) Semiempirical GGA-type density functional constructed with a long-range dispersion correction. *J Comp Chem* 27:1787–1799. doi:10.1002/jcc.20495
15. Yanai T, Tew D, Handy N (2004) A new hybrid exchange-correlation functional using the Coulomb-attenuating method (CAM-B3LYP). *Chem Phys Lett* 393:51–57. doi:10.1016/j.cplett.2004.06.011
16. Parr RG, Weitao Y (1994) Density-functional theory of atoms and molecules. Oxford University Press, Ch 6.
17. Vela A, Trickey SB, Gázquez JL, del Campo JM (2012) Non-empirical improvement of PBE and its hybrid PBE for general description of molecular properties. *J Chem Phys* 136:104108-1–104108-8. doi:10.1063/1.3691197
18. Vosko SH, Wilk L, Nusair M (1980) Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis. *Can J Phys* 58:1200–1211. doi:10.1139/p80-159
19. Perdew JP, Burke K, Ernzerhof M (1996) Generalized gradient approximation made simple. *Phys Rev Lett* 77:3865–3868. doi:10.1103/PhysRevLett.77.3865, Erratum: Perdew JP, Burke K, Ernzerhof M (1997) Comment on generalized gradient approximation made simple. *Phys Rev Lett* 78:1396. doi:10.1103/PhysRevLett.78.1396
20. Adamo C, Barone VJ (1998) Exchange functionals with improved long-range behavior and adiabatic connection methods without adjustable parameters: the mPW and MPW1PW models. *J Chem Phys* 108:664–675. doi:10.1063/1.475428
21. Pople JA, Head-Gordon M, Raghavachari K (1987) Quadratic configuration interaction. A general technique for determining electron correlation energies. *J Chem Phys* 87:5968–5975. doi:10.1063/1.453520
22. Bartlett RJ, Stanton J (1994) Applications of post-Hartree-Fock methods: a tutorial. *Rev Comput Chem* 5:65–169
23. Urban M, Noga J, Cole S, Bartlett RJ (1985) Towards a full CCSDT model for electron correlation. *J Chem Phys* 83:4041–4046. doi:10.1063/1.449067
24. Lotrich V, Flocke N, Ponton M, Yau A, Perera A, Deumens E, Bartlett RJ (2008) *J Chem Phys* 128:194104-1–194104-15. doi:10.1063/1.2920482
25. Curtiss LA, Raghavachari K, Redfern PC, Pople JA (1997) Assessment of Gaussian-2 and density functional theories for the computation of enthalpies of formation. *J Chem Phys* 106:1063–1079. doi:10.1063/1.473182
26. Wang NX, Wilson AK (2004) The behavior of density functionals with respect to basis set. I. The correlation consistent basis sets. *J Chem Phys* 121:7632–7646. doi:10.1063/1.1792071
27. Bak KL, Jørgensen P, Olsen J, Helgaker T, Klopper W (2000) Accuracy of atomization energies and reaction enthalpies in standard and extrapolated electronic wave function/basis set calculations. *J Chem Phys* 112:9229–9242. doi:10.1063/1.481544
28. Tajti A, Szalay PG, Császár AG, Kállay M, Gauss J, Valeev EF, Flowers BA, Vázquez J, Stanton JF (2004) HEAT: High accuracy extrapolated ab initio thermochemistry. *J Chem Phys* 121:11599–11613. doi:10.1063/1.1811608
29. Cramer C (2008) Essentials of computational chemistry, 2nd edn. Wiley, West Sussex, p 369
30. ACES II is a program product of the Quantum Theory Project, University of Florida. Authors: J.F. Stanton, J. Gauss, S.A. Perera, A.D. Yau, J.D. Watts, M. Noojien, N. Oliphant, P.G. Szalay, W.J. Lauderdale, S.R. Gwaltney, S. Beck, A. Balková, D.E. Bernholdt, K.-K. Baeck, P. Rozyczko, H. Sekino, C. Huber, J. Pittner, and R.J. Bartlett. Integral packages included are VMOL (J. Almlöf and P.R. Taylor), VPROPS (P.R. Taylor), ABACUS (T. Helgaker, H.J. Aa. Jensen, P. Jørgensen, J. Olsen, and P.R. Taylor).
31. Huber KP, Herzberg G (1979) Constants of diatomic molecules. Reinhold, New York
32. Jensen P, Bunker PR (1988) The potential surface and stretching frequencies of \tilde{x}^3B_1 methylene (CH_2) determined from experiment using the Morse oscillator-rigid bender internal dynamics Hamiltonian. *J Chem Phys* 89:1327–1332. doi:10.1063/1.455184
33. Dateo CE, Lee TJ, Schwenke DW (1994) An accurate quartic force field and vibrational frequencies for HNO and DNO. *J Chem Phys* 101:5853–5859. doi:10.1063/1.467301
34. Hoy AR, Mills IM, Strey G (1972) Anharmonic force constant calculations. *Mol Phys* 24:1265–1290. doi:10.1080/00268977200102361
35. Kuhn B, Rizzo TR, Luckhaus D, Quack M, Suhm MA (1999) A new six-dimensional analytic potential up to chemically significant energies for the electronic ground state of hydrogen peroxide. *J Chem Phys* 111:2565–2587. doi:10.1063/1.479534
36. Martin JML, Lee TL, Taylor PR (1992) An accurate ab initio quartic force field for ammonia. *J Chem Phys* 97:8361–8371. doi:10.1063/1.463406
37. Lee TJ, Dateo CE, Gazdy B, Bowman JM (1993) Accurate quartic force fields and vibrational frequencies for hydrogen cyanide and hydrogen isocyanide. *J Phys Chem* 97:8937–8943. doi:10.1021/j100137a018
38. Carter S, Mills IM, Handy NC (1993) Vibration-rotation variational calculations: precise results on HCN up to 25000cm^{-1} *J Chem Phys* 99:4379–4390 doi: 10.1063/1.466091
39. Reisner DE, Field RW, Kinsey JL, Dai HL (1984) Stimulated emission spectroscopy: a complete set of vibrational constants for \tilde{x}^1A_1 formaldehyde. *J Chem Phys* 80:5968–5978. doi:10.1063/1.446677
40. Teffo JL, Sulakshina ON, Perevalov VJ (1992) Effective Hamiltonian for rovibrational energies and line intensities of carbon dioxide. *J Mol Spectrosc* 156:48–64. doi:dx.doi.org/10.1016/0022-2852(92)90092-3
41. Smith BC, Winn JS (1988) The C-H overtone spectra of acetylene: Bend/stretch interactions below 10000cm^{-1} . *J Chem Phys* 89:4638–4645. doi:10.1063/1.455683
42. Gray DL, Robiette AG (1979) The anharmonic force field and equilibrium structure of methane. *Mol Phys* 37:1901–1920. doi:10.1080/00268977900101401
43. Martin JML, Taylor PR (1996) The geometry, vibrational frequencies, and total atomization energy of ethylene. A calibration study. *Chem Phys Lett* 248:336–344. doi:0009-2614/96/