

Atomization energies from *ab initio* calculations without empirical corrections



Dr. Dirk Bakowies

Laboratorium für Physikalische Chemie
ETH Zürich, Switzerland

Scientific Abstract

Accurate thermochemistry has turned out to be a major challenge for standard *ab initio* quantum chemistry. High levels of electron correlation combined with very large basis sets are required to adequately treat the formation of a molecule from its constituent atoms. The application of current approaches is thus limited to very small molecules unless empirical corrections are permitted to account for the average effects of basis set incompleteness and higher order electron correlation. Here we test economical compound methods based entirely on *ab initio* electronic structure theory and void of empirical corrections, taking advantage of the observations that (a) higher-order electron correlation corrections are much less basis-set dependent than low-order (MP2) correlation energies, and (b) basis set incompleteness errors can largely be eliminated through extrapolation techniques. These compound methods should be accurate enough even for larger molecules so that they provide useful references for the parametrization of more approximate methods, particularly in semiempirical quantum chemistry.

Motivation

The calibration of semiempirical quantum-chemical methods requires highly accurate reference data, including energies, geometries, and electronic properties. For parametrizations geared towards biomolecular applications it appears mandatory to provide these data for reasonably sized models of typical biomolecules. Atomization and relative (conformational, complexation, activation) energies are usually not available from experiment and need to be obtained from high-level *ab initio* theory.

Model compounds of interest to biomolecular simulation include hydrogen-bonded complexes, nucleic acid base pairs, as well as conformations of small peptides, sugars, lipids, and transition states of typical chemical reactions. Most of these molecules are too large to be studied at reliable levels of standard *ab initio* theory. Here we explore the accuracy of less expensive, yet fully *ab initio* compound methods, taking advantage of the observation that higher-order electron correlation corrections are much less basis set dependent than low-order (MP2) correlation energies.

Related Work in the Literature

Atomization energies – also referred to as dissociation energies – are available from experiment only for very few small molecules since they require the precise knowledge of anharmonic zero-point energies which are typically not known experimentally. They can be calculated by *ab initio* electronic structure theory, but high levels of electron correlation and basis set saturation are usually required to achieve chemical accuracy.

A number of research groups have demonstrated that accurate atomization energies can be obtained from compound methods, combining the results from less expensive electron correlation

methods extrapolated to the full basis set limit with higher order correlation corrections obtained with medium-sized basis sets.

The success of these methods relies on the observation that the complete correlation energy is much more sensitive to basis set saturation than correlation corrections beyond the MP2 or CCSD levels of theory. The focal point analysis introduced by Allen and coworkers¹ and the W-1 (Weizmann 1) theory of Martin² are based on such combinations of various levels of theory.

However, these methods are still limited to fairly small molecules with just a few heavy atoms as they still involve CCSD(T) calculations with sizeable basis sets. On the other hand, a number of approaches introduced by the groups of Pople (G1, G2, G3,...),³ Petersson (CBS-Q, CBS-q, ...),⁴ and Truhlar (MCCM/3, SAC/3)⁵ follow similar ideas, but are less computationally demanding and applicable to medium-sized organic molecules with 10 and more heavy atoms. The caveat is that they require additional empirical corrections for acceptable accuracy and that these corrections have been determined by calibration with experimental heats of formation for large sets of mostly organic molecules. Zero-point energies and thermal corrections are obtained from scaled frequencies computed typically at the Hartree-Fock level to convert calculated atomization energies to heats of formation.

These approaches predict heats of formation of many molecules – also outside the initial training sets – very well, but are not necessarily well suited to calculate atomization energies for which accurate reference data are very limited. It is quite likely – and a recent study by Feller supports this point of view⁶ – that the empirical corrections also absorb non-negligible deficiencies in the calculated zero-point energies, rendering predicted atomization energies significantly less reliable than calculated heats of formation.

Project

We test economical compound methods to calculate atomization energies which are based on *ab initio* electronic structure theory but do not contain any empirical corrections. Their accuracy is assessed with respect to available experimental atomization energies for various small molecules, with respect to very high-level calculations for an extended set of reference molecules, and with respect to experimentally available heats of formation for an even larger set of reference molecules.

The latter comparison is certainly only of approximate nature as it requires the theoretical evaluation of zero-point energies and thermal corrections and thus introduces additional uncertainties. However, we are using a higher level of theory than typically applied in, e.g., the Gaussian (Gx) series of theories,³ to calculate equilibrium geometries and harmonic frequencies. This level of theory, MP2/cc-pVTZ, requires much smaller scaling factors (around 0.98 instead of 0.90) for frequencies to reproduce experimental zero-point energies of small molecules and thus suggests a higher level of accuracy for larger molecules.

A hierarchy of various levels of compound methods will be analyzed and suggested, balancing computational expense with desired accuracy. These approaches should be able to provide reference data of known accuracy for the calibration of semiempirical methods. Relative energies, i.e. differences in atomization energies between various conformations, between a hydrogen-bonded complex and its constituent molecules, or between transition states, reactants and products of a chemical reaction should be significantly more accurate than the atomization energies themselves, as remaining deficiencies in the electron correlation treatment will largely cancel. This certainly plausible hypothesis will be tested in detail and quantified by comparing various compound methods with each other and by studying convergence of basis-set saturation and electron correlation

treatment. In the end, we shall be able to choose a particular compound method based on the size of a molecule and available computational resources and estimate the expected level of accuracy for atomization, activation, and relative conformational energies.

Preliminary Results

It may appear too ambitious to design accurate compound methods that are significantly more economical than e.g. the focal point analysis of Allen¹ and yet do not require empirical corrections like other popular approaches as e.g. G2 theory.³ But a preliminary study for 14 small molecules⁷ with accurately known atomization energies (see Figure 1) shows that we can achieve results of very useful accuracy (< 1 - 2 kcal/mol error) defining compound methods of type

$$\begin{aligned} \Delta E(X) &= E(\text{MP2/cc-pV(TQ)Z}) \\ &+ \Delta E_X(\text{CCSD(T) - MP2}) + \Delta E(\text{core}) \end{aligned}$$

where

$$\begin{aligned} \Delta E_X(\text{CCSD(T) - MP2}) &= E(\text{CCSD(T)/cc-pVXZ}) \\ &- E(\text{MP2/cc-pVXZ}) \end{aligned}$$

$$\begin{aligned} \Delta E(\text{core}) &= E(\text{MP2(full)/cc-pCVTZ}) \\ &- E(\text{MP2/cc-pVTZ}) \end{aligned}$$

X defines the type of Dunning's correlation consistent valence polarized basis sets used (D, T, Q) for higher order correlation corrections, and MP2/cc-pV(TQ)Z refers to the MP2 energy at the basis set limit obtained from calculations with triple- and quadruple-zeta basis sets following the extrapolation protocol suggested by Halkier *et al*⁶. Note that all compound methods shown are significantly more accurate (X=D:1.5, T:0.6, Q:0.4 kcal/mol) than HF (102.8 kcal/mol), MP2 (5.9), CCSD(T) (4.6) and even CCSD(T)(full) (3.5) using basis sets of quad-

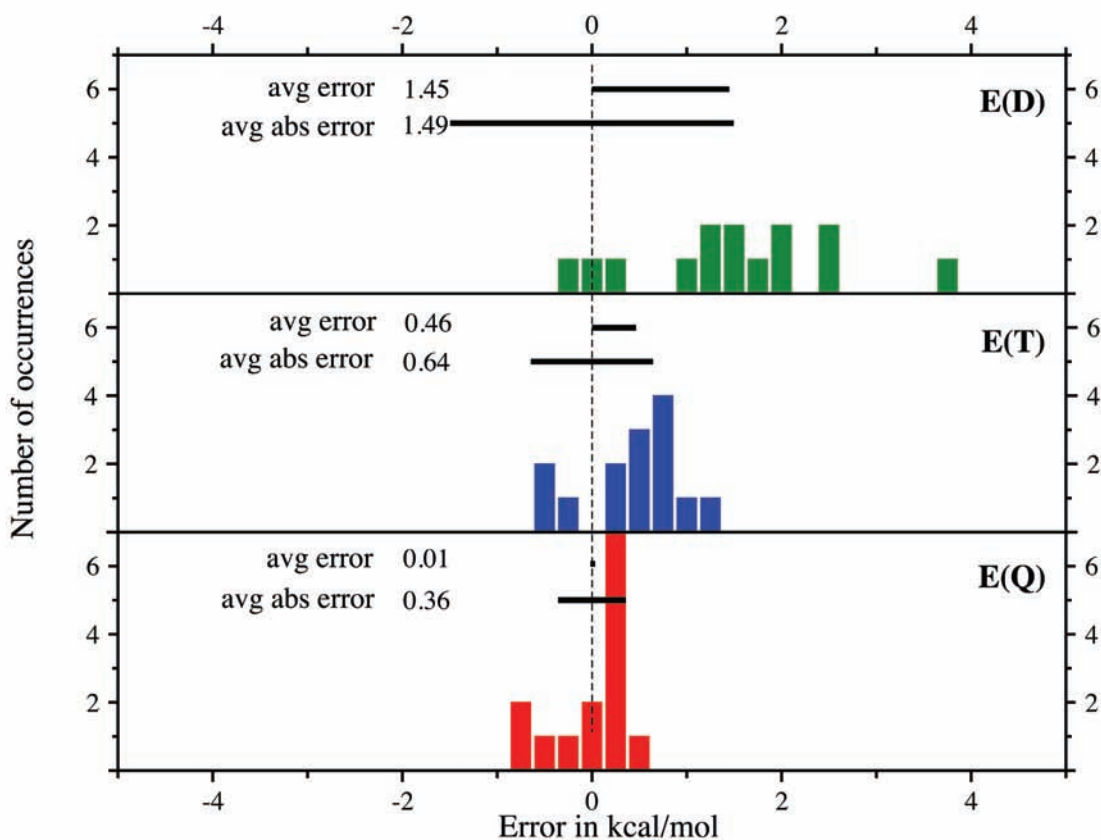


Figure 1: Distribution of errors in calculated atomization energies for 14 molecules with accurately known experimental data. E(D), E(T), E(Q) refer to the level of theory as defined above. All values are in kcal/mol.

triple-zeta quality and that extrapolation to the basis set limit is just as essential as consideration of higher-order electron correlation and core-electron correlation. Using scaled MP2/cc-pVTZ frequencies, we have also calculated heats of formation for a subset of 53 molecules out of the G2/97⁹ and G3¹⁰ reference sets which cover a large number of types of organic and inorganic molecules containing H, C, N, and O atoms. The average errors are somewhat larger (X=D:2.3, T:1.3, Q:0.8 kcal/mol) than before, reflecting both the larger size of the molecules and the additional uncertainty in the calculated vibrational frequencies, but again significantly smaller than those of HF (200.6 kcal/mol), MP2 (6.7) and CCSD(T) (9.0) at the cc-pVQZ basis set level.

In summary, we are quite confident to be able to calculate reliable atomization energies with average errors not exceeding 1 – 2 kcal/mol, using computationally feasible compound methods. With state-of-the-art personal computers we can treat

molecules with up to about 13 heavy atoms (C,N,O) using the E(D) level of theory, and up to about 8 heavy atoms using the E(T) level of theory.

Current Activities

We are extending our preliminary set of reference calculations to include highly accurate CCSD(T)/cc-pV(T,Q)Z and CCSD(T)(full)/cc-pCV(T,Q)Z calculations for a larger subset of molecules taken from the G2/97 and G3 reference sets. This will enable us to gauge less expensive compound methods like E(D) and E(T) against a larger set of reliable data for atomization energies. Furthermore we quantify how accurately atomization energies and relative energies (conformational, complexation, activation, see above) are predicted by any of the proposed compound methods. Production calculations at the E(D) and E(T) levels of theory are planned for larger molecules of biochemical interest to provide accurate reference data also for molecules that are too large for treatments of benchmark quality.

References

- [1] East, A. L. L.; Allen, W. D. J. Chem. Phys. 1993, 99, 4638-4650.
- [2] Martin, J. M. L.; de Olivera, G. J. Chem. Phys. 1999, 111, 1843-1856.
- [3] Curtiss, L. A.; Redfern, P. C.; Frurip, D. J. Rev. Comp. Chem. 2000, 15, 147-210.
- [4] Ochterski, J. W.; Petersson, G. A.; Montgomery Jr., J. A. J. Chem. Phys. 1996, 104, 2598-2619.
- [5] Lynch, B. J.; Truhlar, D. G. J. Phys. Chem. A 2003, 107, 3898-3906.
- [6] Feller, D.; Peterson, K. A. J. Chem. Phys. 1998, 108, 154-176.
- [7] Bak, K. L.; Jørgensen, P.; Olsen, J.; Helgaker, T.; Klopper, W. J. Chem. Phys. 2000, 112, 9229-9242.
- [8] Halkier, A.; Klopper, W.; Helgaker, T.; Jørgensen, P.; Taylor, P. R. J. Chem. Phys. 1999, 111, 9157-9167.
- [9] Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Pople, J. A. J. Chem. Phys. 1997, 106, 1063-1079.
- [10] Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Pople, J. A. J. Chem. Phys. 2000, 112, 7374-7383.