Ab initio thermochemistry of large molecules

Dr. Dirk Bakowies

Laboratorium für Phyikalische Chemie
ETH Zürich, Switzerland

Motivation

Thermochemistry is a branch of thermodynamics concerned with the energy balance of chemical reactions. The elements in their standard states define the universal reference, establishing heats of formation as the primary quantity relating the heat content of one compound to that of another. Experimental access is usually provided through combustion calorimetry, supplemented by measurements of heats of vaporization or sublimation. On the theoretical side, heats of formation may be obtained from atomization energies, but *ab initio* predictions to any useful accuracy (say 1-2 kcal/mol) have not been possible even for the smallest chemical systems until about 20 years ago. This is mainly attributable to the large error incurred if electron correlation is not dealt with accurately, as the total atomization of a molecule involves a significant change in electron correlation. Accurate treatment, however, requires expensive calculations using extended basis sets as the electron correlation energy is known to converge very slowly with the size of orbital-based expansions.

Our interest in accurate *ab initio* thermochemical approaches reflects the need for supplying high-quality reference data for semiempirical method development. Traditionally such reference data have been obtained almost exclusively from experiment, but a reliable *ab initio* protocol would offer several advantages: (a) Heats of formation, previously used but theoretically not well justified, can be replaced by more appropriate atomization energies which are easily available only from calculation. (b) The accuracy of experimental data is often hard to quantify, and while many data are very well established, occasionally large errors do occur. (c) Experimental data are entirely unavailable for several important classes of molecules, and, in particular for biologically relevant model systems such as peptides. The latter point is of particular importance as one of the most promising fields of application for improved semiempirical methodology is in biochemistry.

Related work in the literature

A number of research groups have demonstrated that accurate atomization energies can be obtained from composite 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 coworkers,[1] the Wn (Weizmann) protocols of Martin[2] and the recently introduced HEAT protocol[3] are based on such combinations of various levels of theory. These methods have proven to be very accurate, but they are limited to fairly small molecules with just a few heavy atoms as they still require expensive CCSD(T) calculations with sizeable basis

sets. On the other hand, a number of simpler approaches have been introduced by the groups of Pople[4] (G1, G2, G3,...), Petersson[5] (CBS-Q, CBS-q, ...), and Truhlar[6] (MCCM/3, SAC/3). Although these approaches follow similar ideas, they are less demanding computationally and applicable to medium-sized organic molecules with 10 and more heavy atoms. The downside 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. Such calibration is not entirely satisfactory if the method is supposed to provide high-accuracy reference data, as larger errors may occur occasionally, and, more importantly, as quantities for which the method was not calibrated (such as the atomization energy) may be reproduced less accurately.

Research outline

We are developing fully *ab initio* protocols applicable to large molecules. Our research activities concentrate on two aspects: (a) The convergence behavior of electron correlation energies is analized carefully, employing both results known from analytical theory and a large body of high-quality *ab initio* data. The results of this analysis are used to improve on existing formulas to extrapolate to the complete basis set limit. (b) The bond-separation reaction scheme is implemented in an *ab initio* framework, using only high-level *ab initio* data to correct for systematic errors of more affordable composite approaches. Both aspects of our research have depended heavily on the supercomputing facilities of CSCS to generate the necessary high-quality reference data for analysis and validation.

Extrapolation of electron correlation energies

From the analytical work of Schwartz on helium[7] we know that the correlation energy of two-electron systems converges asymptotically as $f(L)=(L+1)^{-3}$ with the maximum angular momentum L included in the basis set, at least in the limit that perfect radial saturation has been reached. This work has motivated the formulation of various extrapolation techniques, but it can only serve as a guideline. Even if one accepts without proof that the same asymptotic convergence also applies to molecules, a number of further issues affect practical implementations in extrapolation models. One needs to realize that the inverse cubic term is only the leading term of the asymptotic expansion, that different singlet pair energies in N-electron atoms are generally associated with different values of higher expansion coefficients, and that triplet pair energies contribute only with $(L+1)^{-5}$ and higher terms. Further «perturbations» of the asymptotic convergence behavior are caused by the replacement of partial-wave expansions with correlation-consistent basis sets and by the somewhat arbitrary identification of the quantum number L with a particular value of the basis set cardinal number X. All these effects influence the effective convergence behavior which differs from its asymptotic limit for all practical values of X. Thus it appears legitimate to introduce some degree of empiricism and modify extrapolation formulas for optimal agreement with high-level reference data. One way of doing this is to allow for a calibrated exponent β in $f(X)=X^{-\beta}$ that is found to be optimal for a given base and target of extrapolation.

The analysis of a large body of reference data for small to medium-sized first-row molecules (HCNOF) has led to a number of conclusions relevant to the design of thermochemical models:[8] (a) Coupled-cluster (CCSD) treatments always converge faster to the complete basis set limit than perturbational (MP2) treatments. (b) Convergence behavior varies substantially with molecular composition, thus requiring either specialized treatments or large-basis-set calculations for accurate extrapolations. (c) Atomization energies converge less regularly than total energies and require, in general, a different set of optimal extrapolation exponents.

Motivated by observation (b), an improved formula, $f(X)=(X+\xi_i)^{-\beta}$, has been developed for small basis set extrapolations which includes a variable angular momentum offset $\xi=\xi_i$ expressed in terms of a calibrated function relating the number of correlated electrons to the number and type of atoms contributing them (Fig.1).[9] In essence, the offset reflects differences between correlation-consistent basis sets for hydrogen and for first-row atoms. Application of this formula allows for remarkably accurate extrapolations, which are matched by calibrated one-parameter formulas and simple asymptotic limit formulas only if used with significantly larger basis sets. Extensive validation shows that the new formula is robust enough for general application to neutral systems containing hydrogen and all first-row elements from B to Ne.

A new ab *initio* thermochemical protocol

The concept of bond separation reactions (BSRs)[10] may be used in conjunction with appropriate extrapolation techniques to further improve upon low-level treatments of atomization energies. [11] High-level calculations are performed to supply accurate data for a limited set of small parent compounds, essentially reducing the challenging task of calculating accurate atomization energies to the much simpler task of calculating reaction energies. BSRs can be mapped onto a system of corrective bond increments, thus providing for a high-level description of the process of bond dissociation embedded in a lower-level treatment of the perturbation through the chemical environment.

BSRs have often been used before, but their application was limited by the availability of accurate experimental reference data. The use of high level *ab initio* calculations eliminates this restriction, extending the range of possible applications to all molecules for which valence structures can be drawn. Such an approach not only profits from the full potential of BSRs, it also remains truly *ab initio*, as
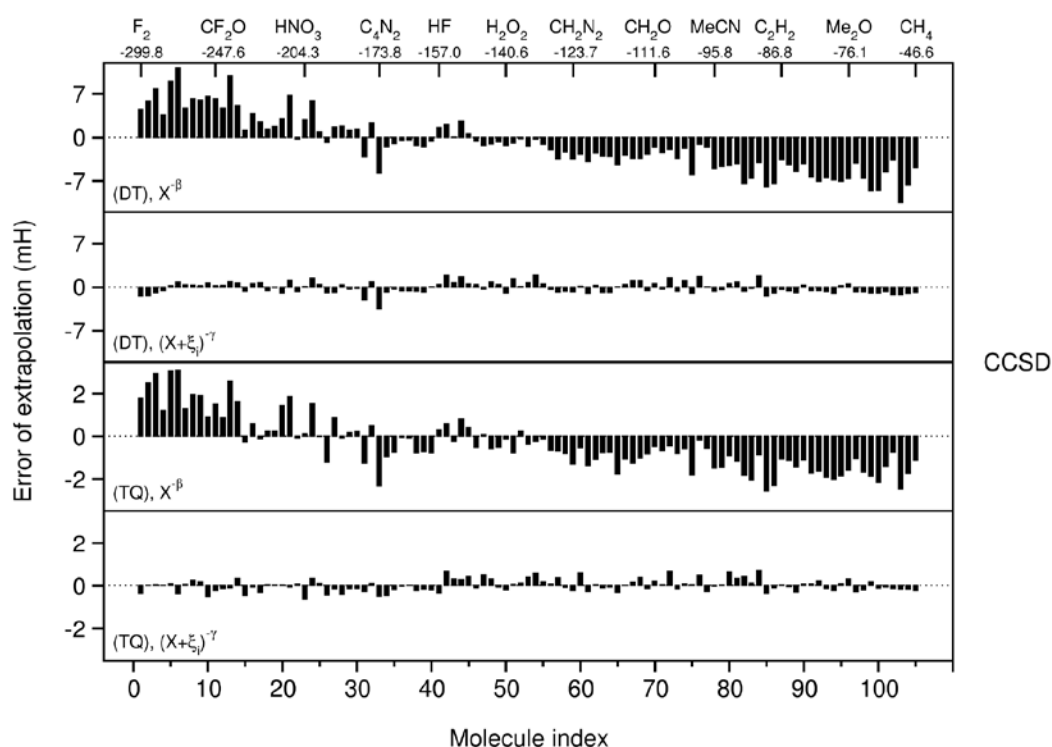

Fig. 1: Extrapolation errors for total correlation energies at the CCSD level as evaluated for a large set of 105 molecules (HCNOF). Molecules have been sorted according to their correlation energy per atom (see top bar for examples), an index that places hydrogen-rich molecules on the right and molecules with electronegative atoms on the left. The upper two panels show extrapolations from double- and triple-zeta basis sets, the lower two panels those from triple- and quadruple-zeta basis sets. The first and third panels refer to the standard one-parameter extrapolation formula, and the second and fourth panels show results for the new extrapolation formula. Energies are given in mH (1 mH ≈ 0.6 kcal/mol).

neither experimental data nor empirical calibration are required. Extensive analysis has shown that BSR corrected low-level atomization energies reproduce high-level reference data quite accurately, not only in total but also component by component, thus largely eliminating the risk of fortuitous error cancellation. Furthermore, good results are obtained not only for „simple organic" molecules, but also for more complex systems involving, for example, charge-separated valence structures. Corrections beyond the CCSD(T) model are treated under the assumption that the corresponding BSRs are thermoneutral, reducing calculations to simple summations of bond increments. This works extremely well for scalar relativistic corrections and typically provides for reasonable qualitative or semi-quantitative estimates of diagonal Born-Oppenheimer corrections and higher-order electron correlation contributions (Fig. 2). Validation with a large set of experimental heats of formation (neutral, closed shell, HCNOF) shows that the new thermochemical approach is competitive with or even slightly superior to the popular and accurate, but empirically calibrated G3 protocol, at comparable computational cost.[11] Further possible improvements are currently being investigated, and extensions to second-row elements, to ionic systems, and to radicals are planned.
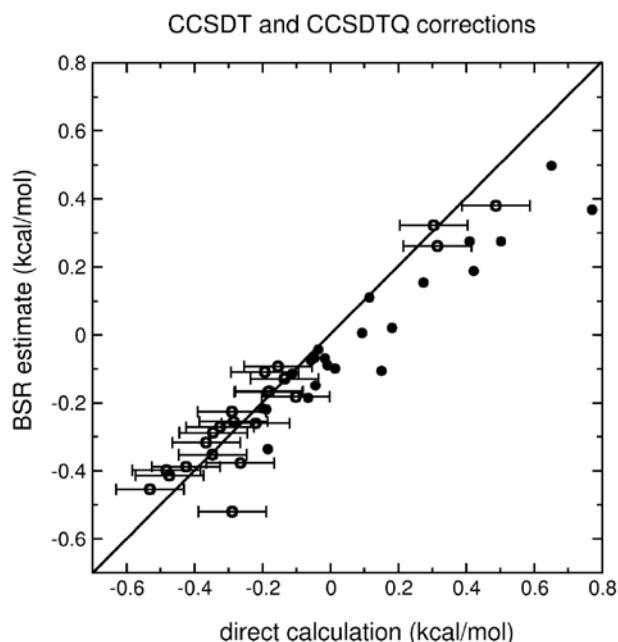

CCSDT and CCSDTQ corrections

Fig. 2: Comparison between estimated (BSR scheme, assuming thermoneutral reactions) and calculated sums of CCSDTQ-CCSDT and CCSDT-CCSD(T) corrections to atomization energies (44 molecules, HCNOF). Open circles refer to molecules for which full CCSDTQ calculations proved to be unfeasible in which case they were approximated from CCSDT(Q) and CCSDT[Q] results. The error bars of 0.1 kcal/mol indicate the uncertainty expected from a thorough analysis for smaller systems. Note that the BSR estimate involves no fitting or empirical calibration. As the BSRs are assumed to be thermoneutral, however, the estimate reduces to a simple summation over bond increments uniquely determined from precalculated data for a limited set of small parent compounds.

References

1. East, A. L. L.; Allen, W. D. J. Chem. Phys. 1993, 99, 4638.

2. Karton, A.; Rabinovich, E.; Martin, J. M. L.; Ruscic, B. J. Chem. Phys. 2006, 125, 144108.

3. Tajti, A et al J. Chem. Phys. 2004, 121, 11599.

4. Curtiss, L. A.; Redfern, P. C.; Frurip, D. J. Rev. Comp. Chem. 2000, 15, 147.

5. Ochterski, J. W.; Petersson, G. A.; Montgomery Jr., J. A. J. Chem. Phys. 1996, 104, 2598.

6. Lynch, B. J.; Truhlar, D. G. J. Phys. Chem. A 2003, 107, 3898.

7. Schwartz, C. Phys. Rev. 1962, 126, 1015.

8. Bakowies, D. J. Chem. Phys. 2007, 127, 084105.

9. Bakowies, D. J. Chem. Phys. 2007, 127, 164109.

10. Hehre, W. J.; Ditchfield, R.; Radom, L.; Pople, J. A. J. Am. Chem. Soc., 1970, 92, 4796.

11. Bakowies, D. manuscript in preparation.