

Cite as: A. Karton, *WIREs Comput. Mol. Sci.* 6, 292–310 (2016). <http://dx.doi.org/10.1002/wcms.1249>

A computational chemist's guide to accurate thermochemistry for organic molecules

Amir Karton

School of Chemistry and Biochemistry, The University of Western Australia, Perth, WA
6009, Australia; amir.karton@uwa.edu.au

Abstract

Composite ab initio methods are multistep theoretical procedures specifically designed to obtain highly accurate thermochemical and kinetic data with confident sub-kcal mol⁻¹ or sub-kJ mol⁻¹ accuracy. These procedures include all energetic terms that contribute to the molecular binding energies at these levels of accuracy (e.g., CCSD(T), post-CCSD(T), core-valence, relativistic, spin-orbit, Born–Oppenheimer, and zero-point vibrational energy corrections). Basis-set extrapolations (and other basis-set acceleration techniques) are used for obtaining these terms at sufficiently high levels of accuracy. Major advances in computer hardware and theoretical methodologies over the past two decades have enabled the application of these procedures to medium-sized organic systems (e.g., ranging from benzene and hexane to amino acids and DNA bases). With these advances, there has been a proliferation in the number of developed composite ab initio methods. We give an overview of the accuracy and applicability of the various types of composite ab initio methods that were developed in recent years. General recommendations to guide selection of the most suitable method for a given problem are presented, with a special emphasis on organic molecules.

ACCURATE THEORETICAL THERMOCHEMICAL AND KINETIC DATA

Information about the energetic properties of atoms and molecules is at the heart of many chemical investigations. Accurate thermochemical and kinetic properties such as reaction energies and barrier heights are of great value when we are interested in the feasibility and spontaneity of chemical transformations. This is especially true in complex chemical environments involving multiple competing reaction pathways.^{1,2} The emergence of powerful supercomputers, alongside important theoretical developments over the past two decades, has made computational chemistry one of the most powerful approaches for generating highly accurate thermochemical data for organic molecules. Composite ab initio approaches are particularly useful in cases where experimental data is lacking or the experimental errors are large. For example, when investigating transient species such as free radicals, reactive intermediates, and transition structures. In addition, highly accurate energetic data can be used to generate accurate spectroscopic properties (e.g., bond lengths, harmonic frequencies, anharmonicities, and rotational constants) as well as electrical properties (e.g., dipole moments, polarizabilities, and hyperpolarizabilities).^{3,4}

A further motivation for obtaining highly accurate energetic data is for the parameterization and testing of approximate electronic structure methods, in particular, density functional theory (DFT) approximations.^{5,6} Traditionally, experimental thermochemical and kinetic data has been used for this purpose, however, the available experimental data is not always sufficiently diverse and/or sufficiently accurate. Composite ab initio methods offer an alternative since they can achieve accuracies of at least one order of magnitude better than the best contemporary DFT functionals. The use of benchmark experimental data is additionally complicated by the fact that it has to be converted into nonrelativistic electronic energies for a direct comparison with the DFT data. This involves removing energetic contributions (such as zero-point vibrational energies and relativistic contributions) from the experimental data, which is not always a trivial task.

The influential computational chemist, Paul von Ragué Schleyer, defined computational chemistry as the branch of chemistry that “attempts to model all aspects of real chemistry as closely as possible by using calculations rather than experiment”.⁷ Indeed, one of the primary challenges of computational chemistry is to develop quantum chemical methods that can narrow the gap between theory and benchmark experimental data. Composite ab initio methods are multistep theoretical procedures that are specifically designed to obtain highly accurate thermochemical and kinetic data. These procedures are perhaps the most accurate ab initio methods that are still applicable to medium-sized organic systems. With major advances in computer hardware and theoretical methodologies over the past two decades, there has been a proliferation in the number of developed composite ab initio methods. This review gives an overview of the various types of composite methods and offers some strategies for choosing the most appropriate method for a given problem. Within the scope of this review it is impossible to cover in detail all the methods that have been developed in the past two decades. We will therefore focus on a restricted subset of fixed-recipe methods (such as the *Wn* and HEAT protocols), rather than give a superficial coverage of all the methods that were developed. Additional discussions can be found in a number of excellent reviews.^{4,8,9,10,11,12,13,14,15,16}

OVERVIEW AND CLASSIFICATION OF COMPOSITE AB INITIO METHODS

Composite thermochemical methods, also known as compound thermochemistry methods, are theoretical procedures that aim at obtaining accurate thermochemical and kinetic data. These procedures can be viewed as three-dimensional. The first two dimensions describe the nonrelativistic electronic energy, and the third dimension comprises of any residual energetic contributions (e.g., scalar relativistic, spin-orbit, zero-point vibrational energy, and Born–Oppenheimer corrections). In principle, if we treat all three dimensions in an exact manner, we would be able to obtain any thermochemical quantity exactly.

Let us start with a brief overview of the two dimensions describing the nonrelativistic electronic energy. Coupled cluster (CC) theory is perhaps the most reliable, yet computationally affordable, ab initio method for solving the electronic Schrödinger equation. As a consequence, truncated CC theory constitutes the basis for most contemporary composite ab initio methods. Application of CC theory involves two approximations. The first is the use of a finite basis set to express the orbitals in the Hartree–Fock wavefunction (a.k.a. the one-particle truncation error). The second approximation is the truncation of the cluster operator to include only n -tuple excitations (a.k.a. the n -particle truncation error).¹⁷ The levels of these truncations dictate the accuracy of the calculation as illustrated schematically in a modified Pople diagram (Figure 1).¹⁸ The horizontal axis describes the size of the one-particle basis set, whilst the vertical axis describes the level of electron correlation included in the wavefunction. As one traverses along the vertical and horizontal axes the accuracy of the electronic energy increases, however, the computational cost of the CC calculation also increases factorially with the number of electrons and basis-set functions. Two key points on the Pople diagram are:

- The exact electronic energy lies in the top right corner. This energy is obtained by carrying out a full configuration interaction (FCI) calculation in conjunction with an infinitely large basis set
- For systems dominated by a single reference determinant, the CCSD(T) method provides a very good approximation to the FCI energy at a significantly reduced computational cost

Composite ab initio methods can be primarily classified according to the points on the Pople diagram they are trying to approximate:

- post-CCSD(T) composite methods attempt to approximate the FCI energy
- CCSD(T) composite methods approximate the CC energy with singles, doubles, and quasiperturbative triple excitations

Figure 1 shows these points for a number of popular composite methods, along with examples of molecules to which these methods have been applied using current mainstream computer hardware.

Let us now turn to the third dimension of composite methods. Energetic properties obtained on the nonrelativistic electronic potential energy surface are widely used for the validation and

parameterization of computationally cost-effective methods (such as DFT and semiempirical methods).^{6,19} However, in order to reproduce accurate experimental thermochemical properties secondary energetic contributions have to be considered. The third dimension of composite ab initio methods consists of energetic contribution required for converting the nonrelativistic electronic energy to thermochemical quantities. These contributions may include: spin-orbit, scalar relativistic, zero-point vibrational energy, Born–Oppenheimer, thermal, and entropic corrections. In certain cases additional corrections may be needed, for example: (i) conformational corrections to the enthalpy for floppy molecules;^{20,21} or (ii) tunneling contributions for certain reaction barrier heights.²² In principle, any secondary component that can reasonably affect the molecular binding energies at the target level of accuracy should be explicitly (or implicitly) included in the third dimension of the composite method.

NONRELATIVISTIC ELECTRONIC ENERGY

Post-CCSD(T) composite methods

There are two main types of post-CCSD(T) composite methods:

- FCI/CBS composite methods attempt to approximate the CC energy with up to quintuple (or higher) excitations at the infinite basis-set limit
- CCSDT(Q)/CBS composite methods attempt to approximate the CC energy with single, double, triple, and quasiperturbative quadruple excitations at the infinite basis-set limit

FCI/CBS composite methods

FCI/CBS composite methods are capable of confident sub-kJ mol⁻¹ predictions of total atomization energies and are generally applicable to any molecule including highly multireference systems (such as O₃, X_mO_n (X = H, N, F, Cl),^{8,23,24,25} and C_n clusters).²⁶ However, due to their steep computational cost their use becomes prohibitive for molecules containing more than a handful of non-hydrogen atoms.^{8,23,24,25,27,28} A number of FCI/CBS composite methods have been developed to date, most notably, the highly accurate extrapolated ab initio thermochemistry (HEAT) methods of Stanton and co-workers,^{29,30,31} and the Weizmann-4 (W4) family of methods of Martin and co-workers.^{28,32} The W4 methods include the W4lite,³² W4,³² W4.2,³² W4.3,³² and W4.4²⁸ variants. The HEAT methods include the HEAT-345QP^{29,30} and HEAT-456QP³¹ methods. Both the HEAT and *Wn* methods may be regarded as fixed-recipe composite methods since they prescribe a sequence of well-defined steps for obtaining highly accurate thermochemical properties. In this context, the pioneering work of Allen, East, and Császár on the focal-point analysis (FPA) approach,^{33,34,35,36,37,38,39,40} and the extensive work of Feller, Peterson, and Dixon on the FPD approach^{4,11,24,25,27,41} must also be mentioned. These latter two provide more flexible strategies, rather than a prescribed set of steps for obtaining accurate thermochemical data. An important feature of these more flexible approaches is that they can be easily modified to include (or exclude) contributions for achieving a given level of accuracy and for extending their applicability to species across the periodic table. For example, by inclusion of second-order spin-orbit corrections for treating heavy main-group, transition-metal, and f-block systems.^{4,41} In addition, various

contributions can be calculated at very high levels of theory as needed. For example, in the FPD approach the valence CCSD(T) correlation energy is extrapolated from basis sets as large as aug-cc-pV8Z and aug-cc-pV9Z in some cases.^{4,25,27} However, since application of these methods may require extensive hands-on experience in high-level ab initio electronic structure methods, we will focus on the fixed-recipe composite methods, which are usually more straightforward to apply. Excellent reviews of the FPA and FPD approaches can be found elsewhere.^{4,12,27,41}

Generally speaking, the fixed-recipe FCI/CBS composite methods share a number of design goals:

- To compute total atomization energies with confident sub-kJ mol⁻¹ accuracy (i.e., 95% confidence intervals < 1 kJ mol⁻¹, and maximal errors below ~1 kJ mol⁻¹ even for pathologically multireference systems such as ozone, halogen oxides, and carbon clusters)
- All components of the electronic energy up to quintuple excitations (or higher) are calculated at (or close to) the complete basis-set limit
- All secondary components that can reasonably affect the molecular binding energies at the sub-kJ mol⁻¹ range are considered explicitly
- Completely devoid of empirical parameters derived from experiment

Table 1 gives an overview of the characteristics of the HEAT and W4 composite methods. Before proceeding to a detailed discussion of the W4 method we note that it yields significant improvements over the earlier W3 theory for a relatively small increase in the computational cost.^{32,42} In W4 theory the SCF and valence CCSD energies are extrapolated with basis sets of up to aug'-cc-pV(6+d)Z (the notation aug' indicates the combination of a non-augmented basis set on hydrogen and an augmented basis set on all other elements).^{43,44,45} We note that the CCSD singlet- and triplet-coupled pair energies are extrapolated separately to the infinite basis-set limit,³² since they exhibit different basis set convergence rates.⁴⁶ The quasiperturbative triple excitations, (T), are extrapolated with basis sets of up to aug'-cc-pV(5+d)Z. The higher-order connected triples, T₃-(T), correlation contribution is extrapolated from the cc-pVnZ basis sets (*n* = D, T). The parenthetical connected quadruple, (Q), and higher-order connected quadruple, T₄-(Q), correlation contributions are calculated with the cc-pVTZ and cc-pVDZ basis sets, respectively. The connected quintuple, T₅, correlation contribution is calculated with the *sp* part of the cc-pVDZ basis set (denoted by DZ).

It has been found that contributions from successively higher cluster expansion terms (i.e., CCSD → (T) → T₃-(T) → (Q) → T₄-(Q) → T₅) converge increasingly faster to the complete basis-set limit, since they increasingly reflect non-dynamical rather than dynamical correlation.^{28,32} Indeed, the fact that W4 theory (and related FCI/CBS composite methods) can be carried out at a realistic computational cost for systems like CF₄ and acetic acid hinges on this behaviour.⁸ In practice, all the components of the valence CCSDTQ5 correlation energy in W4 theory are calculated very close to (or at) the infinite basis-set limit.^{28,32}

An important feature of W4 theory is that the correlation components of the CCSDTQ5 energy are obtained within the frozen-core approximation. That is, the inner-shell orbitals (1s for first-row atoms, and 1s, 2s, and 2p for second-row atoms) are constrained to be doubly occupied in all configurations. In practice, this partitioning between the inner-shell and valence-shell electrons makes W4 theory applicable to molecules containing second-row elements at a realistic

computational cost. To account for inner-shell correlation, W4 theory includes a core-valence (CV) correction term. The CV correction is extrapolated from the aug'-cc-pwCVnZ basis sets at the CCSD(T) level ($n = T, Q$).⁴⁷

What about the higher W4.x methods? In W4.2 and W4.3 the CV correction is obtained at the CCSDT level,³² and in W4.4 it is obtained at the CCSDT(Q) level.²⁸ Post-CCSD(T) contributions to the CV component are practically nil for systems dominated by dynamical correlation, but can be non-negligible at the kJ mol^{-1} level for systems with moderate-to-severe nondynamical correlation effects. For example, in W4.2 theory they increase the atomization energies by 0.01 (CH_4), 0.03 (C_2H_4), 0.08 (HO_3), 0.13 (O_3), and 0.19 ($\text{C}_2, {}^1\Sigma_g^+$) kcal mol^{-1} . W4.3 and W4.4 theories also use larger basis sets than W4 theory to calculate the valence post-CCSD(T) contributions, and in addition they include contributions of connected sextuple excitations from CCSDTQ5(6)/DZ or CCSDTQ56/DZ calculations.^{28,32}

Moving to the HEAT theories, we note that the notation HEAT-xyzA indicates that the SCF energy is extrapolated from the aug-cc-pCVnZ basis sets ($n = x, y, z$), the all-electron CCSD(T) correlation energy is extrapolated from the aug-cc-pCVnZ basis sets ($n = y, z$), and the valence T_3 -(T) correlation contribution is extrapolated from the cc-pVTZ and cc-pVQZ basis sets. The letter(s) after the xyz numbers indicate which valence post-CCSDT contributions are included in the method. For example, 'QP' denotes a CCSDTQ5-CCSDT correction, and '(Q)' denotes a CCSDT(Q)-CCSDT correction. These higher-level correlation contributions are calculated with the cc-pVDZ basis set. Overall, the HEAT and W4 approaches are attempting to approximate similar points on the modified Pople diagram (Figure 1). In particular, HEAT-456QP and W4 approximate the CCSDTQ5/CBS energy; HEAT-456(Q) and W4lite approximate the CCSDT(Q)/CBS energy; and HEAT-345(Q) and W3.2 approximate the CCSDT(Q) energy further away from the CBS limit (i.e., using smaller basis sets than those used in HEAT-456(Q) and W4lite in the CCSD(T) extrapolations). Nevertheless, there are two main differences between the HEAT and W4 approaches. (i) The HEAT methods employ the unrestricted Hartree-Fock (HF) formalism, whereas the W4 methods employ the restricted open-shell formalism for open-shell systems. (ii) In the HEAT methods all electrons are correlated in the CCSD and (T) calculations, and the post-CCSD(T) contributions are obtained within the frozen-core approximation. As mentioned above, the W4 methods partition between the inner-shell and valence-shell electrons in the correlation calculations, and include a CV correction term, which is obtained at the CCSD(T) (W4), CCSDT (W4.2 and W4.3), and CCSDT(Q) (W4.4) level.

Validation of FCI/CBS composite methods against experiment

The confident sub- kJ mol^{-1} accuracies that are achieved by FCI/CBS composite methods make their validation against experimental data a nontrivial task. The validation of such methods requires a set of highly accurate experimental data with well-defined error bars that are sufficiently smaller than the intrinsic error of the method being evaluated. The Active Thermochemical Tables (ATcT) thermochemical network developed by Ruscic and co-workers^{48,49} provides reliable and internally consistent thermochemical values with the accuracy needed to evaluate the performance of FCI/CBS methods. Table 2 lists the error statistics for some of the W4 and HEAT variants. As customary in experimental thermochemistry, we will primarily look at 95% confidence intervals (CIs), rather than root-mean-square deviations (RMSDs) or mean-absolute deviations (MADs) when

validating the performance of these highly accurate theoretical procedures.⁵⁰ The 95% CIs reflect the expected uncertainty associated with any thermochemical quantity calculated with these methods. As noted in ref. 50, the 95% CI is approximately equal to twice the RMSD and to 2.5–3.5 times the MAD (depending on the error distribution).

The performance of W4 theory was evaluated in ref. 8 against 35 atomization energies from ATcT (associated with error bars ≤ 0.05 kcal mol⁻¹) including both first- and second-row species. Specifically, this set includes the following molecules: CH, CH₂, CH₃, CH₄, C₂H₂, C₂H₄, C₂H₆, H₂CO, H₂C=C=O, CH₃OH, NH, NH₂, NH₃, OH, H₂O, HO₂, H₂O₂, HCN, HNO, trans-HONO, CO, CO₂, NO, NO₂, N₂O, O₂, O₃, SO, SO₂, H₂, HF, HCl, N₂, F₂, and Cl₂. Against these highly accurate experimental values, W4 theory attains an RMSD of 0.085 kcal mol⁻¹, implying a 95% (2σ) confidence interval of 0.170 kcal mol⁻¹, and a 99% (3σ) confidence interval of 0.255 kcal mol⁻¹. The largest deviations (theory–experiment) are obtained for ozone (–0.23) and nitrous acid (+0.20 kcal mol⁻¹). For 29 of these systems, TAEs were obtained with the higher W4.x theories ($x = 2-4$). The RMSD for the W4.x theories is reduced to 0.060 kcal mol⁻¹, implying a 95% CI of merely 0.120 kcal mol⁻¹. In addition, the largest deviations are significantly reduced; only four deviations exceed 0.1 kcal mol⁻¹, namely: ozone (–0.14), hydrogen peroxide (–0.12), dichlorine (–0.11), and ethane (+0.11 kcal mol⁻¹).

What about the HEAT methods? The performance of the HEAT methods was evaluated in reference 31 against a set of 18 first-row ATcT TAEs (associated with error bars ≤ 0.06 kcal mol⁻¹). Specifically, this set includes the following molecules: CH, CH₂, CH₃, C₂H, C₂H₂, OH, H₂O, HO₂, H₂O₂, HCO, CO, CO₂, NO, O₂, H₂, HF, N₂, and F₂. We note that this test set does not include pathologically multireference systems such as ozone. Against these ATcT reference values, HEAT-456QP theory attains an RMSD of 0.100 kcal mol⁻¹, implying a 95% CI of 0.200 kcal mol⁻¹, and a 99% CI of 0.300 kcal mol⁻¹. Interestingly, the HEAT-345QP method results in better performance (RMSD = 0.068 and 95% CI = 0.135 kcal mol⁻¹) indicating that it benefits from some fortuitous cancellation of errors (see also discussions in references 4 and 31).

Magnitude of post-CCSDT(Q) contributions to molecular binding energies

In the previous section we have seen that FCI/CBS composite methods can achieve 95% CIs in the sub-kJ mol⁻¹ accuracy range. This means that atomization energies or heats of formation calculated with these methods are associated with error bars narrower than 1 kJ mol⁻¹. However, in practice, these theories are limited to systems with up to five non-hydrogen atoms with some symmetry (e.g., W4 theory has been applied to C₅, CF₄, F₂O₂, Cl₂O₂, CH₃COOH, SO₃, and S₄).^{8,26} Fully iterative CCSDTQ and CCSDTQ5 calculations place strenuous demands on computational resources since they exhibit asymptotic CPU time scaling proportional to $n_o^4 n_v^6 N_{iter}$ and $n_o^5 n_v^7 N_{iter}$, respectively (where n_o is the number of occupied orbitals, n_v is the number of virtual orbitals, and N_{iter} is the number of iterations required to achieve convergence). For comparison, the computational cost of the CCSDT(Q) method scales as $n_o^3 n_v^5 N_{iter}$ for the CCSDT part, followed by a one-time $n_o^4 n_v^5$ step for the (Q) part. This translates to significant savings in computational resources compared to the CCSDTQ and CCSDTQ5 calculations (in terms of CPU time, memory, and disk usage).

So can CCSDT(Q)/CBS composite methods provide a good approximation for the FCI/CBS composite methods? In order to answer this question we have to look at the magnitude of the

higher-order connected quadruple, T_4 -(Q), and connected quintuples, T_5 , contributions. It has been shown that T_4 -(Q) contributions tend to universally decrease the atomization energies, whereas T_5 contributions tend to increase them.^{28,30,32} Therefore, there is usually a significant degree of cancellation between these two contributions. Overall, post-CCSDT(Q) contributions tend to decrease the atomization energies. For systems that exhibit mild-to-moderate nondynamical correlation effects these contributions typically reach up to ~ 1 kJ mol⁻¹. For example, for typical organic molecules the post-CCSDT(Q) contributions decrease the atomization energies by: 0.03 (methanol and methanimine), 0.05 (acetaldehyde), 0.09 (ketene), 0.12 (formic acid and glyoxal), 0.17 (carbon dioxide), and 0.27 (cyano radical) kcal mol⁻¹ (for more examples see Table S1 of the Supporting Information, which summarizes the post-CCSDT(Q) contributions for a set of 124 species). However, for highly multireference systems the post-CCSDT(Q) contributions can decrease the atomization energies by up to ~ 1 kcal mol⁻¹, for example, by: 0.40 (dioxygen difluoride), 0.44 (tetrasulfur), 0.51 (ozone), and 0.79 (dicarbon, $^1\Sigma_g^+$) kcal mol⁻¹. Therefore, CCSDT(Q)/CBS composite methods should be applied with caution when modelling such systems.

CCSDT(Q)/CBS composite methods

A number of CCSDT(Q)/CBS composite methods have been developed in recent years. In these methods the post-CCSDT(Q) contributions are completely neglected. Examples of such methods include the *Wn* methods (*W3.2*³² and *W4lite*)³² and the HEAT methods (HEAT-345(Q)³⁰ and HEAT-456(Q)).³¹ These methods generally yield 95% CIs close to the 1 kJ mol⁻¹ mark. However, in contrast to the FCI/CBS composite methods, maximal errors can certainly exceed 1 kJ mol⁻¹. This is particularly true for highly multireference systems, systems with highly polar bonds, and for second-row compounds.²⁸

In the abovementioned *Wn* and HEAT methods the parenthetical quadruples are calculated with the cc-pVDZ basis set. The difference between these methods lies in the levels of theory used for obtaining the CCSD(T) and T_3 -(T) components. In the more rigorous CCSDT(Q)/CBS composite methods (i.e., *W4lite* and HEAT-456(Q)) the SCF and CCSD correlation energies are extrapolated from basis sets of up to sextuple- ζ quality, and the (T) correlation energy is extrapolated from basis sets of up to quintuple- ζ quality. In *W4lite* the valence T_3 -(T) correlation energy is extrapolated from the cc-pVnZ basis sets ($n = D, T$), whereas in HEAT-456(Q) theory it is extrapolated from larger basis sets ($n = T, Q$). The computationally more economical variants of these methods (i.e., *W3.2* and HEAT-345(Q)) use smaller basis sets for the SCF, CCSD, and (T) extrapolations.

Against the set of 18 first-row ATcT TAEs considered in reference 31, the HEAT-456(Q) method shows practically the same performance as the HEAT-456QP method (both methods attain a 95% CI of 0.20 kcal mol⁻¹, Table 2). Whereas the HEAT-345(Q) method shows a slight deterioration in performance relative to the HEAT-456(Q) method (95% CI = 0.17 kcal mol⁻¹). Moving to the *Wn* methods, *W4lite* shows similar performance to the HEAT-456(Q) and HEAT-345(Q) methods, with a 95% CI of 0.18 kcal mol⁻¹ (Table 2). However, the *W3.2* method shows significantly worse performance with a 95% CI of 0.30 kcal mol⁻¹, and a largest deviation (underestimation) of 0.38 kcal mol⁻¹ for O₂. The *Wn* methods were also assessed against a larger set of 35 ATcT TAEs,⁸ the 95% CIs for *W4lite* and *W3.2* for this set are 0.25 and 0.38 kcal mol⁻¹, respectively. The larger 95% CIs

obtained for W4lite and W3.2 theories may be attributed to the presence of highly multireference systems (e.g., ozone and nitrous acid) and second-row compounds in the extended ATcT test set.

The abovementioned CCSDT(Q)/CBS theories are still computationally demanding, for example, the largest systems W4lite has been applied to are SF₆, SF₆⁻,⁵¹ and C₆H₁₄,⁵² whilst HEAT-345(Q) has been applied to benzene.⁵³ The principal bottleneck in applying these methods to larger systems is the evaluation of the T₃-(T) term with the cc-pVQZ (HEAT-345(Q)) or even the cc-pVTZ basis set (W3.2 and W4lite). A number of cost-effective alternatives for these methods have been developed in recent years. These include methods such as W3.2lite,⁵⁴ W3-F12,⁵⁵ W3X,⁵⁶ and W3X-L.⁵⁷ These procedures employ one (or more) of the following basis-set acceleration techniques in order to reduce the computational cost:

- Truncating the basis sets used in the post-CCSD(T) calculations and scaling the post-CCSD(T) terms by empirically motivated scaling factors
- Using explicitly correlated methods in the CCSD(T) calculations
- Using MP2-based basis-set additivity schemes

The W3.2lite method obtains the T₃-(T) term from the cc-pVDZ and a truncated version of the cc-pVTZ basis set (e.g., without the *f* functions on non-hydrogen atoms and without the *d* functions on hydrogens).⁵⁴ Since the highest angular momentum (*L*) present in these basis sets is now the same, an extrapolation in terms of *L* is no longer possible. Alternatively, W3.2lite uses an expression of the form: $E_{\infty}^{T-(T)} \approx E_{cc-pVDZ}^{T-(T)} + a[E_{cc-pVDZ}^{T-(T)} - E_{cc-pVTZ(no\ f)}^{T-(T)}]$, which involves a single adjustable parameter (*a*). This scaling factor was determined from least squares fitting to basis-set-limit T₃-(T) energies. In addition, the (Q)/cc-pVDZ term in W3.2lite is scaled by an empirical parameter determined from least squares fitting to basis-set-limit T₄ + T₅ energies. These approximations significantly reduce the computational cost relative to W3.2 theory with little loss in accuracy.⁵⁴ W3.2lite was applied to organic molecules such as benzene, fulvene, phenyl radical, pyridine, furan, benzyne isomers, [1.1.1]propellane, and bicyclo[1.1.1]-pentane. W3-F12 is another variant of W3.2 theory, which instead of reducing the cost of the post-CCSD(T) steps, reduces the computational cost of the CCSD(T) steps by employing explicitly correlated F12 techniques.^{55,58} The performance of W3-F12 theory was assessed against the set of 35 highly accurate atomization energies from ATcT (set I, Table 2). W3-F12 theory attains an RMSD of 0.18 kcal mol⁻¹, implying a 95% (2σ) confidence interval of 0.36 kcal mol⁻¹. The largest deviation of +0.368 kcal mol⁻¹ (theory-experiment) is obtained for nitrous acid.

The cost-effective W3X procedure of Chan and Radom combines both the basis-set acceleration techniques used in W3-F12 and W3.2lite theories.⁵⁶ That is, it combines explicitly correlated techniques in the CCSD(T) steps (with basis sets smaller than those used in W3-F12) with computationally-economical approximations to the post-CCSD(T) contributions (in a similar manner to W3.2lite). In addition, W3X provides additional computational savings by calculating the core-valence correction at the MP2/cc-pCVTZ level of theory, rather than at the CCSD(T) level like in W3.2, W3.2lite, and W3-F12. Against the 140 accurate TAEs in the W4-11 database (*vide infra*), W3X attains a MAD of 0.45 kcal mol⁻¹, which implies a 95% CI of about 1.4 kcal mol⁻¹ (taken as three times the MAD).⁵⁰ In an attempt to improve on these results the W3X-L procedure extrapolates the CCSD-F12 energy from the aug'-cc-pVnZ basis sets (n = T, Q).⁵⁷ Note that in order to keep the

computational cost at a minimum the aug'-cc-pVnZ basis sets are used, rather than the larger cc-pVnZ-F12 basis sets (which were specifically optimized for explicitly correlated calculations).⁵⁹ In addition, W3X-L adds a cost-effective $\Delta(T)$ correction term to the MP2-based core-valence correction in W3X. Overall, W3X-L represents a significant improvement over W3X, namely the MAD over the W4-11 database is cut by $\sim 50\%$, implying a 95% confidence interval of about $0.6 \text{ kcal mol}^{-1}$. We note that the W3X and W3X-L procedures have been applied to organic systems such as toluene and benzene.⁵⁷

Finally, we note that Klopper and co-workers recently added post-CCSD(T) contributions to their CCSD(T)+F12+INT composite method, which approximates the CCSD(T)/CBS limit.^{60,61,62} In this scheme the T_3 -(T) and (Q) valence correlation energies are calculated in conjunction with the cc-pVTZ and cc-pVDZ basis sets, respectively. This method was evaluated against a set of 72 ATcT TAEs (associated with error bars $\leq 0.65 \text{ kcal mol}^{-1}$).⁶² Against this large set of ATcT reference values, the CCSD(T)+F12+INT (+post-CCSD(T) contributions) composite scheme attains an RMSD of $0.51 \text{ kcal mol}^{-1}$, implying a 95% confidence interval of about 1 kcal mol^{-1} .

Magnitude of post-CCSD(T) contributions to molecular binding energies

Before proceeding to a discussion of composite ab initio methods that attempt to approximate the CCSD(T) energy, it is useful to look at the magnitude of post-CCSD(T) contributions to the molecular binding energies. It is well known that the CCSD(T) method benefits from a systematic error cancellation between T_3 -(T) and post-CCSD(T) excitations ($T_4 + T_5$).^{15,28,29,30,31,32,42,63,64,65} These contributions are of similar orders of magnitude, but the T_3 -(T) contributions tend to universally decrease the molecular binding energies, whereas the $T_4 + T_5$ contributions universally increase them. Overall, the sum of the post-CCSD(T) contributions can be either negative or positive, depending on the relative magnitudes of the T_3 -(T) and $T_4 + T_5$ contributions. This error compensation works remarkably well for systems dominated by a single reference configuration. Table S2 (Supporting Information) gathers the post-CCSD(T) contributions for a large set of 173 molecules with up to nine non-hydrogen atoms. For just over half of these species the sum of the post-CCSD(T) contributions amounts to less than 1 kJ mol^{-1} . These include a range of organic systems involving various functional groups, e.g., ketene, allene, oxirane, oxirene, acetic acid, and the methylamine radical (for additional examples see Table S2). We note however that post-CCSD(T) contributions can exceed the half kcal mol^{-1} mark for medium-sized systems even though they are dominated by a single reference configuration.^{11,52} For example, it has been recently shown⁵² that for the homologous series of *n*-alkanes the post-CCSD(T) contributions from W4lite theory are: -0.13 (C_2H_6), -0.28 (C_3H_8), -0.40 (C_4H_{10}), -0.54 (C_5H_{12}), and -0.65 (C_6H_{14}) kcal mol^{-1} (see also discussion in ref. 11 regarding the post-CCSD(T) contributions to the TAEs of alkanes). For 90% of the systems in Table S2 the sum of the post-CCSD(T) contributions amounts to less than 1 kcal mol^{-1} . For example, the post-CCSD(T) contributions for a number of organic systems are: -0.40 (benzene and pyridine), -0.28 (pentane), 0.37 (ethynyl radical and *m*-benzyne), 0.58 (cyanogen), and 0.88 (cyano radical) kcal mol^{-1} . For pathologically multireference systems the post-CCSD(T) contribution can reach well over 1 kcal mol^{-1} ; for example, 1.31 (singlet tetracarbon), 1.69 (heptacarbon), 2.32 (singlet *p*-benzyne), 2.88 (ozone), 3.35 (nonacarbon), and 4.44 (singlet octacarbon) kcal mol^{-1} .

CCSD(T)/CBS composite methods

A large variety of composite ab initio methods that approximate the CCSD(T) energy have been developed over the past two decades. Generally speaking, they can be divided into two classes:

- CCSD(T)/CBS composite methods that attempt to approximate the CCSD(T) energy at the infinite basis-set limit
- CCSD(T)/TZ composite methods that attempt to approximate the CCSD(T) energy in conjunction with a triple- ζ -quality basis set

Here we will only consider the more rigorous CCSD(T)/CBS composite methods. Methods that belong to the CCSD(T)/TZ class include the popular Gaussian- n ^{9,66,67,68} and CBS^{69,70} family of methods. These methods are computationally very economical and can be routinely applied to very large systems. For example, the G4(MP2) method⁶⁷ has been recently applied for the calculation of the heat of formation of C₆₀.⁷¹

The most rigorous composite CCSD(T)/CBS methods use only CC calculations. In these methods the components of the CCSD(T) energy (i.e., HF, CCSD, and (T)) are extrapolated separately to the infinite basis-set limit from large Gaussian basis sets. Examples of such methods include the Wn and Wn -F12 theories ($n = 1, 2$).^{55,72} For example, in W2.2 theory³² (which is a slightly modified version of the original W2 theory)⁷² the HF and valence CCSD energies are extrapolated from the aug'-cc-pV($n+d$)Z basis sets ($n = Q, 5$). Similarly to the higher Wn methods,³² the CCSD singlet- and triplet-coupled pair energies are extrapolated separately to the infinite basis set limit. The valence (T) energy is extrapolated from the aug'-cc-pV($n+d$)Z basis sets ($n = T, Q$). The CV correction is extrapolated from the aug'-cc-pwCV n Z basis sets at the CCSD(T) level ($n = T, Q$). The performance of W2.2 theory was assessed against the set of 35 highly accurate atomization energies from ATcT (set I, Table 2). W2.2 theory attains an RMSD of 0.74 kcal mol⁻¹, implying a 95% (2σ) confidence interval of 1.49 kcal mol⁻¹. The largest deviations (theory–experiment) are of course obtained for highly multireference systems, most notably for ozone (–3.45) and nitrogen dioxide (–1.46 kcal mol⁻¹). Exclusion of these two outliers from the training set reduces the 95% CI to 0.80 kcal mol⁻¹.

Since the 95% CI for W2.2 theory (1.49 kcal mol⁻¹) is nearly 1 order of magnitude larger than that for W4 theory (0.17 kcal mol⁻¹, Table 2), it is sensible to evaluate the performance of theories such as W2.2 against reference data obtained at the W4 level. To this end, reference 8 generated a large and diverse dataset of 140 TAEs of small first- and second-row molecules obtained by means of the W4 or W4.x theories (known as the W4-11 database). The species in the W4-11 database cover a broad spectrum of bonding situations and multireference character, and as such it is an ideal benchmark set for the validation of CCSD(T) composite methods. The error statistics for W2.2 theory against the 140 TAEs in the W4-11 database are practically identical to those obtained against the 35 ATcT TAEs above; namely, RMSD = 0.74 and 95% CI = 1.48 kcal mol⁻¹. It should be stressed, however, that since the CCSD(T) method should not be used for systems with significant nondynamical correlation effects, such systems should be removed from the evaluation set. Omitting the 14 systems with severe multireference effects from the W4-11 database reduces the W2.2 error

statistics to RMSD = 0.42 and 95% CI = 0.85 kcal mol⁻¹.⁷³ This set of highly accurate 126 TAEs will be used below for the evaluation of the other CCSD(T) composite methods (and will be referred to as the W4-11nonMR database). Finally, we note that W2-F12 is an explicitly correlated version of W2.2, which attempts to reduce the computational cost of W2.2 without sacrificing the accuracy.⁵⁵ W2-F12 theory attains essentially the same error statistics as W2.2 against the W4-11nonMR dataset (namely, RMSD = 0.42 and 95% CI = 0.84 kcal mol⁻¹).⁵⁵ W2-F12 was successfully applied to medium-sized hydrocarbons (e.g., cubane),⁷⁴ as well as to systems of biological relevance (e.g., adenine⁵⁵ and methionine).²¹

W1 theory⁷² and its explicitly correlated version (W1-F12)⁵⁵ calculate the CCSD(T) energy with smaller basis sets compared to W2.2 and W2-F12, as illustrated in Figure 1. For example, in W1 theory the HF and valence CCSD energies are extrapolated from the aug'-cc-pV(*n*+d)Z basis sets (*n* = T, Q), and the valence (T) energy is extrapolated from the aug'-cc-pV(*n*+d)Z basis sets (*n* = D, T). The CV correction is calculated with a large triple- ζ basis set. The W1-type methods exhibit a noticeable deterioration in performance relative to the W2-type methods. For example, for the W4-11nonMR dataset they obtain 95% CIs of 1.26 (W1) and 1.48 (W1-F12) kcal mol⁻¹. However, it should be pointed out that W1-F12 performs poorly for a number of systems containing second-row elements; ostensibly due to the somewhat anemic character of the VDZ-F12 basis set. Upon removal of the second-row systems from the training set the 95% CI for W1-F12 is reduced to 0.90 kcal mol⁻¹.⁵⁵ W1 and W1-F12 have been applied to large hydrocarbons (e.g., sumanene,⁷¹ corannulene,⁷⁵ and dodecahedrane)⁷⁴ as well as to systems of biological relevance (e.g., guanine⁵⁵ and arginine),²¹ and other organic species (e.g., substituted benzenes,⁷⁶ phenols,⁷⁷ and anisoles).^{78,79}

In an attempt to further reduce the computational cost of CCSD(T)/CBS composite methods, a number of methods that combine moderate level CCSD(T) calculations in conjunction with MP*n* calculations have been developed. Examples of such methods include the correlation consistent Composite Approach (ccCA) methodology of Wilson and co-workers^{10,80,81,82,83} the W1X-*n* theories of Chan and Radom (*n* = 1, 2),⁸⁴ and the CCSD(T)+F12+INT scheme of Klopper and co-workers.^{61,85,86} Reference 10 provides an excellent review of the various ccCA variants. Here we will mention that the ccCA-PS3 variant extrapolates the HF energy from the aug-cc-pV(*n*+d)Z basis sets (*n* = T, Q). The valence CCSD(T) energy is calculated in conjunction with the cc-pV(T+d)Z basis set and an MP2-based basis-set-correction term (Δ MP2) is added in order to approximate the CCSD(T)/CBS energy. The Δ MP2 basis-set-correction term is calculated as: Δ MP2 = MP2/CBS – MP2/cc-pV(T+d)Z, where MP2/CBS is extrapolated from basis sets of up to aug-cc-pV(Q+d)Z. The CV correction term is calculated at the MP2/aug-cc-pCVTZ level of theory. Against the W4-11nonMR database the ccCA-PS3 attain an RMSD of 1.03 kcal mol⁻¹, implying a 95% CI of about 2 kcal mol⁻¹.⁸ The W1X-*n* methods (*n* = 1, 2) are modified versions of the W1-F12 protocol.⁵⁵ The main difference between W1-F12 and W1X-1, is that in W1X-1 the CV correction is calculated at the MP2 level rather than at the CCSD(T) level. W1X-2 additionally reduces the computational cost by extrapolating the valence CCSD-F12 and (T) components from the orbital aug'-cc-pV(*n*+d)Z basis sets (*n* = D, T), rather than the larger cc-pV*n*Z-F12 basis sets (similar to the approach adopted in W3X-L, *vide supra*). The performance of these methods was evaluated against the TAEs in the W4-11nonMR database; W1X-1 and W1X-2 attain MADs of 0.50 and 0.53 kcal mol⁻¹, respectively,⁵⁷ implying 95% CIs of about 1.5 and 1.6 kcal mol⁻¹, respectively (taken as three times the MADs).⁵⁰ The W1X-2 method has been recently applied to the dihydronaphthyl radical.⁸⁷ Finally, we note that the CCSD(T)+F12+INT composite approach^{85,86} approximates the CCSD(T)/CBS by adding an interference-corrected Δ MP2-F12 basis set correction

term to the CCSD(T)/cc-pVDZ-F12 base energy. The performance of this method has been evaluated for atomization energies with the addition of post-CCSD(T) contributions (*vide supra*).⁶¹

HOW TO CHOOSE THE RIGHT COMPOSITE METHOD?

In order to choose an appropriate composite ab initio method for a given molecule (i.e., CCSD(T)/CBS, CCSDT(Q)/CBS, or FCI/CBS), we first need to estimate the magnitude of post-CCSD(T) contributions. A number of diagnostics, which are based on the CCSD T_1 amplitudes, have been proposed for gauging the degree of multireference character in a system, e.g., the \mathcal{T}_1 and \mathcal{D}_1 diagnostics^{88,89} (see also discussion in ref. 32 for additional diagnostics). However, it has been found that these diagnostics are not necessarily good predictors for the magnitude of post-CCSD(T) contributions to the TAEs.^{3,8,23,32,90,91,92} On the other hand, it has been found that an energy-based diagnostic of the form: $\%TAE[(T)] = 100 \times (TAE[CCSD(T)] - TAE[CCSD])/TAE[CCSD(T)]$ provides a reliable *a priori* diagnostic for the importance of post-CCSD(T) contributions to the TAEs (where $TAE[CCSD]$ and $TAE[CCSD(T)]$ are the TAEs calculated at the CCSD and CCSD(T) levels, respectively).^{8,32} In particular, there is a high statistical correlation between the percentage of the TAE accounted for by parenthetical connected triple excitations ($\%TAE[(T)]$) and the percentage of the TAE accounted for by connected quadruple and quintuple excitations ($\%TAE[T_4+T_5]$). For the 140 molecules in the W4-11 database, a squared-correlation coefficient of $R^2 = 0.94$ is obtained.^{8,32} In practice, the $\%TAE[(T)]$ diagnostic can be converted into a useful *a priori* indicator of the magnitude of the post-CCSD(T) or $T_4 + T_5$ contributions to the TAE. This is illustrated in Figure 2 for the 140 molecules in the W4-11 database.

Let us begin with the $T_4 + T_5$ contributions to the TAE (upper pane of Figure 2). For the 65 molecules in the W4-11 set for which $\%TAE[(T)] \leq 2.5\%$, the magnitude of the $T_4 + T_5$ contribution to the TAE ≤ 1.0 kcal mol⁻¹ (with the exception of acetic acid for which this contribution amounts to 1.08 kcal mol⁻¹). For the 100 molecules in the W4-11 set for which $\%TAE[(T)] \leq 5\%$, the magnitude of the $T_4 + T_5$ contribution to the TAE ≤ 1.8 kcal mol⁻¹, again with one exception (NCCN) for which this contribution amounts to 2.49 kcal mol⁻¹. For molecules with $5\% \leq \%TAE[(T)] \leq 10\%$ the $T_4 + T_5$ contribution can reach up to 2.7 kcal mol⁻¹ (P₄), whereas for molecules with $\%TAE[(T)] \geq 10\%$ the $T_4 + T_5$ contribution can exceed 3.0 (FO₂ and ClOO) and even 4.0 kcal mol⁻¹ (O₃).

These results illustrate that the $\%TAE[(T)]$ diagnostic can serve as an upper bound for the $T_4 + T_5$ contribution to the TAE. However, for the purpose of choosing between a CCSD(T)/CBS and a post-CCSD(T)/CBS composite method, it is useful to predict the magnitude of the post-CCSD(T) contribution to the TAE (Figure 2, lower pane). For the 100 molecules in the W4-11 set for which $\%TAE[(T)] \leq 5\%$, the magnitude of the post-CCSD(T) contribution ≤ 0.5 kcal mol⁻¹, with the exception of two minor outliers (SiF₄ and NCCN for which it is 0.58 and 0.68 kcal mol⁻¹, respectively). For the 25 molecules in the W4-11 set with $5\% \leq \%TAE[(T)] \leq 10\%$ the magnitude of the post-CCSD(T) contributions is generally smaller than 1.0 kcal mol⁻¹; with one minor exception (NO₂ for which it is 1.07 kcal mol⁻¹), and two major exceptions (cis- and trans-HO₃ for which it is 1.85 and 2.31 kcal mol⁻¹, respectively). Finally, for molecules with $\%TAE[(T)] \geq 10\%$ the magnitude of the post-CCSD(T) contribution can approach (or even exceed) the 3.0 kcal mol⁻¹ mark. These empirical observations suggest the following general guidelines:

- %TAE[(T)] ≤ 5% indicates that post-CCSD(T) contributions should not exceed 0.5 kcal mol⁻¹
- %TAE[(T)] ≤ 10% indicates that post-CCSD(T) contributions should generally not exceed 1.0 kcal mol⁻¹
- %TAE[(T)] ≥ 10% indicates that post-CCSD(T) can certainly exceed 1.0 kcal mol⁻¹ by significant amounts

Three practical considerations are noted when applying the %TAE[(T)] diagnostic for estimating the magnitude of post-CCSD(T) contributions. (i) The %TAE[(T)] diagnostic exhibits a relatively weak basis set dependence; values obtained with a triple- ζ basis set are sufficiently converged for practical purposes, and even values obtained with a double- ζ basis set are generally within 1% of the basis set limit values.⁸ (ii) This diagnostic provides useful upper bounds (but not necessarily lower bounds) for the magnitude of post-CCSD(T) contributions. (iii) For systems that are strongly unbound at the SCF level the %TAE[(T)] diagnostic can be artificially high. This situation is encountered for two systems in the W4-11 database, namely, F₂ and Be₂ (see also discussion in ref. 8).

SECONDARY ENERGETIC CONTRIBUTIONS

The nonrelativistic electronic energy typically accounts for about 95% of the TAE. The secondary energetic contributions usually account for ~5% of the TAE, of which the ZPVE accounts for the lion's share. So far we discussed the levels of theory that are required for obtaining accurate nonrelativistic electronic energies. In order to obtain thermochemical and kinetic properties with high accuracy, each of the secondary contributions has to be converged to the same target accuracy as the electronic energy. This section gives a brief overview of some of the secondary energetic components in common composite ab initio methods, additional discussions can be found elsewhere.^{4,11,12,16,41}

Relativistic and DBOC corrections

Scalar relativistic corrections refer to two different effects, the so-called Darwin and mass-velocity (DMV) corrections. The mass-velocity correction is due to the relativistic dependence of the electronic mass on the velocity. The Darwin correction corresponds to what may be described as high-frequency oscillations of the electron around its mean position. Perhaps the simplest method to account for scalar relativistic contributions in an all-electron framework is the first-order perturbation theory of the one-electron DMV operators.^{93,94} In the HEAT theories this term is calculated at the CCSDT/aug-cc-pCVTZ level of theory²⁹ and a two-electron Darwin term is also added.³⁰ This approach works quite well for molecules containing first-row elements. The contemporary variants of the *W_n* methods³² use the second-order Douglas-Kroll-Hess (DKH) relativistic treatment.^{95,96} The DKH approximation has been shown to yield results that are in close agreement with the full relativistic treatment.^{97,98} In the *W_n* methods the scalar relativistic contribution is obtained from the difference between nonrelativistic CCSD(T)/aug'-cc-pV(Q+d)Z and relativistic CCSD(T)/aug'-cc-pV(Q+d)Z-DK calculations.⁹⁹

Scalar relativistic corrections universally reduce the atomization energies, and their importance generally grows with the number of bonds in the system. For example, for the CH_n series they reduce the TAEs by: 0.04 (CH), 0.15 (CH_2 , $^3\text{B}_1$), 0.17 (CH_3), and 0.19 (CH_4) kcal mol^{-1} .⁸ A similar trend is observed for the C_2H_n series, i.e., the TAEs are reduced by: 0.28 (C_2H_2), 0.33 (C_2H_4), and 0.39 (C_2H_6) kcal mol^{-1} . As can be seen scalar relativistic corrections can exceed 1 kJ mol^{-1} even for very simple organic molecules. Table S3 (Supporting Information) gives an overview of these corrections for a wide range of 110 organic molecules with up to 20 non-hydrogen atoms. The set of amino acids provides a nice illustration of the magnitude that scalar relativistic corrections can reach for medium-sized organic molecules; namely they range from 1.06 (glycine) to 2.61 (arginine) kcal mol^{-1} .²¹ We also note that for larger systems, such as dodecahedrane, scalar relativistic corrections can exceed the 3 kcal mol^{-1} mark.⁷⁴

The interactions between the magnetic moment arising from the electron spin and the magnetic field associated with the orbital motion of the electron result in spin-orbit (SO) splitting of the energy levels, and lowering of the ground state energy. For molecules containing first- and second-row elements only first-order SO corrections have to be considered, these corrections are nonzero for radicals in a degenerate ground state.^{32,29} Atomic and molecular SO corrections can make nontrivial contributions to the molecular binding energies. Table S3 (Supporting Information) gives an overview of these corrections for a wide range of organic molecules. Again, we consider the amino acids in order to get a feeling of the magnitude of these contributions to the TAEs; SO corrections reduce the TAEs by amounts ranging from 0.62 (glycine) to 1.43 (methionine) kcal mol^{-1} .²¹

Finally, we note in passing that a number of composite procedures (mostly CCSD(T)/TZ and CCSD(T)/CBS-type methods) have been developed for treating heavy main-group, transition-metal, and/or f-block systems.^{4,9,25,41} Prominent examples include the $G4$,^{100,101} Wn ($n = 1, 2$),^{102,103} ccCA ,^{104,105} and FPD ^{11,106,107,108,109,110} approaches. Extension of the applicability of post-CCSD(T) theories capable of kJ mol^{-1} accuracies to elements of the third-row and beyond may involve tasks such as: (i) development of more general basis set extrapolation schemes (within various pseudopotential approximations) for the core electrons, (ii) more rigorous treatment of relativistic contributions (e.g., using the third-order DKH approximation) and including second-order spin-orbit coupling effects, (iii) appropriate partitioning of core-core and core-valence contributions, (iv) consideration of Lamb shift contributions, and (v) for highly multireference transition-metal systems, DFT and even CCSD(T) may no longer be adequate for geometry optimizations.

Deviations from the Born–Oppenheimer approximation can be estimated by computing the so-called diagonal Born–Oppenheimer correction (DBOC). For the 110 organic molecules in Table S3 the DBOC contribution to the TAE can reach up to 0.42 kcal mol^{-1} (dodecahedrane). In the contemporary variants of the Wn and $Wn\text{-F12}$ methods³² ($n = 1\text{--}3$) the DBOC contribution is calculated at the HF/aug-cc-pVTZ level of theory. However, it should be pointed out that for systems with many hydrogens, correlation contributions to the DBOC will reduce the HF DBOC contribution by up to 50%.^{21,28,31,74,111,112,113} For example, correlation contributions calculated at the CCSD/cc-pVDZ level reduce the DBOC by 0.09 (propene),¹¹² 0.11 (propane),¹¹² 0.10 (glycine),²¹ 0.18 (valine),²¹ 0.22 (octahedrane),⁷⁴ and 0.36 (dodecahedrane) kcal mol^{-1} .⁷⁴ In the HEAT³¹ and $W4.4$ ²⁸ methods a correlation correction to the DBOC is calculated at the CCSD/cc-pVDZ level of theory.

Zero-point vibrational energies

The ZPVE is required whenever one wants to convert a ‘bottom of the well’ total atomization energy (TAE_e) to a total atomization energy at 0 K (TAE_0). In terms of absolute energies, the ZPVE contribution can get quite hefty even for simple organic systems. In particular, it increases with the size of the system and the number of hydrogens in the system, and usually approaches the 100 kcal mol⁻¹ regime for systems containing ~10 hydrogen atoms. For example, the ZPVE contribution for the first members of the *n*-alkane series is: 27.74 (CH₄), 46.39 (C₂H₆), 64.20 (C₃H₈), 81.26 (C₄H₁₀), and 98.88 (C₅H₁₂) kcal mol⁻¹.¹¹² For larger organic molecules the ZPVE can reach well over 100 kcal mol⁻¹; for example, it is 138.18 (arginine)²¹ and 222.97 (dodecahedrane)⁷⁴ kcal mol⁻¹. It has been pointed out that while ab initio methods have been refined to the point where they can determine accurate electronic energies for medium-sized organic systems, it becomes increasingly hard to converge the ZPVE to the same level of accuracy (particularly for systems containing many hydrogens).^{21,111,112,114} For example, for systems in which the ZPVE contribution amounts to ~100 kcal mol⁻¹ (e.g., pentane¹¹² and histidine),²¹ even a 1% error due to neglect of explicit anharmonicity can translate into a ~1 kcal mol⁻¹ error in the final ZPVE.

Obtaining ZPVEs with sub-kJ mol⁻¹ or even sub-kcal mol⁻¹ accuracies, in particular for molecules containing many hydrogens, requires consideration of vibrational anharmonicity effects. In brief, there are a number of ways in which the anharmonic ZPVE can be obtained:

- The combination of accurate theoretical harmonic frequencies (e.g., CCSD(T)/cc-pVQZ) with cubic and quartic anharmonic corrections (e.g., at the MP2 or CCSD(T) levels) can yield anharmonic ZPVEs with sub-kJ mol⁻¹ accuracies. Such approaches are used in the HEAT and in W4-type methods,^{29,30,31,32,111,112} however they become prohibitive for molecules containing more than a handful of non-hydrogen atoms. Notably, in the HEAT methods both the harmonic and anharmonic force fields are calculated at the CCSD(T)/cc-pVQZ level of theory.²⁹
- In cases where experimental fundamentals are available, a fairly good estimate can be obtained by taking a simple or weighted average of the harmonic ZPVE and the ZPVE determined from the experimental fundamentals,¹¹⁵ since the former tends to overestimate the anharmonic ZPVE and the latter tends to underestimate it.^{16,116}
- A fairly straightforward way to obtain the anharmonic ZPVE is to scale CCSD(T) or even DFT harmonic frequencies by a single empirical scaling factor. This simplistic method turns out to yield useful estimates of anharmonic ZPVEs. The purpose of scaling is twofold: (i) to correct for vibrational anharmonicity; and (ii) to correct for imperfections of the theoretical model. Many studies have derived such scaling factors for various levels of theory by a least-squares fit to benchmark ZPVEs obtained from experiment or high level quartic force field calculations.^{116,117,118,119,120,121,122,123,124,125} It has been recently shown that scaling of CCSD(T) or double-hybrid DFT harmonic frequencies can yield anharmonic ZPVEs with accuracies approaching those obtained from quartic force fields.¹²⁵ On the other hand, scaling harmonic

ZPVEs obtained from DFT calculations is a cost-effective way for obtaining anharmonic ZPVEs, consequently this approach is used in most of the CCSD(T)/CBS (e.g., W1-F12, W2-F12, ccCA, and W1X-*n*) and the computationally economical CCSDT(Q)/CBS (e.g., W3-F12 and W3.2lite) composite methods. Finally, we note that the scaling factor between the harmonic and anharmonic ZPVEs is not the same as the scaling factor between the harmonic and fundamental frequencies, and that simple scaling seems to work much better for anharmonic ZPVEs rather than fundamentals.¹²⁵

Conclusion

This review focuses on fixed-recipe composite ab initio methods in the context of accurate thermochemistry for organic molecules. Generally speaking, these methods can be classified according to the points on the Pople diagram (Figure 1) they are attempting to approximate:

- FCI/CBS
- CCSDT(Q)/CBS
- CCSD(T)/CBS
- CCSD(T)/TZ

We highlighted the rationale behind the design of these methods, their applicability and the accuracies they can achieve. FCI/CBS composite methods can achieve 95% confident intervals below (or even well below) 1 kJ mol⁻¹. CCSDT(Q)/CBS composite methods attain 95% CIs on the order of ~1 kJ mol⁻¹, however some caution is needed when applying them to pathologically multireference systems (e.g., O₃ and C_{*n*}) and/or second-row systems (e.g., S₄ and SO₃). CCSD(T)/CBS methods generally attain 95% CIs on the order of ~1 kcal mol⁻¹ for systems that are characterized by mild-to-moderate nondynamical correlation effects. We have shown that the *a priori* %TAE[(T)] diagnostic is a reliable diagnostic for the importance of post-CCSD(T) contributions. We make the following general recommendations:

- If benchmark accuracy is desired (i.e., 95% CI ≤ 1 kJ mol⁻¹):
 - In case the %TAE[(T)] diagnostics ≥ 10% use FCI/CBS composite methods (e.g., W4, W4.x, HEAT-456QP, and HEAT-345QP).
 - In case the %TAE[(T)] diagnostics ≤ 10% one can alternatively use the CCSDT(Q)/CBS composite methods in the following order of preference: HEAT-456(Q) ~ W4lite > HEAT-345(Q) ~ W3.2 > W3.2lite ~ W3-F12 ~ W3X-L > W3X
- If chemical accuracy is desired (i.e., 95% CI ≤ 1 kcal mol⁻¹):
 - In case the %TAE[(T)] diagnostics ≥ 5% use the CCSDT(Q)/CBS composite methods in the abovementioned order of preference
 - In case %TAE[(T)] diagnostics ≤ 5% one can use one of the CCSD(T)/CBS composite methods

References

- ¹ Vereecken L, Francisco JS. Theoretical studies of atmospheric reaction mechanisms in the troposphere. *Chem Soc Rev* 2012, 41:6259.
- ² Bell AT, Head-Gordon M. Quantum mechanical modeling of catalytic processes. *Annu Rev Chem Biomol Eng* 2011, 2:453.
- ³ Karton A, Martin JML. Performance of W4 theory for spectroscopic constants and electrical properties of small molecules. *J Chem Phys* 2010, 133:144102.
- ⁴ Peterson KA, Feller D, Dixon DA. Chemical accuracy in ab initio thermochemistry and spectroscopy: current strategies and future challenges. *Theor Chem Acc* 2012, 131:1079.
- ⁵ Cohen AJ, Mori-Sánchez P, Yang W. Challenges for Density Functional Theory. *Chem Rev* 2012, 112:289.
- ⁶ Goerigk L, Grimme S. A thorough benchmark of density functional methods for general main group thermochemistry, kinetics, and noncovalent interactions. *Phys Chem Chem Phys* 2011, 13, 6670.
- ⁷ Preface to Reviews in Computational Chemistry, K.B. Lipkowitz, D.B. Boyd (Eds.), Volume 1, VCH Publishers, New York (1990), pp. vii–xii.
- ⁸ Karton A, Daon S, Martin JML. W4-11: A high-confidence benchmark dataset for computational thermochemistry derived from first-principles W4 data. *Chem Phys Lett* 2011, 510:165.
- ⁹ Curtiss LA, Redfern PC, Raghavachari K. Gn theory. *WIREs Comput Mol Sci*, 2011, 1:810.
- ¹⁰ DeYonker N, Cundari TR, Wilson AK, in: Piecuch P, Maruani J, Delgado-Barrio G, Wilson S (Eds.), *Advances in the Theory of Atomic and Molecular Systems (Progress in Theoretical Chemistry and Physics, Vol. 19)*, Springer Netherlands, Dordrecht, 2009 pp. 197–224.
- ¹¹ Feller D, Peterson KA, Dixon DA. A survey of factors contributing to accurate theoretical predictions of atomization energies and molecular structures. *J Chem Phys* 2008, 129:204105.
- ¹² Helgaker T, Klopper W, Tew DP. Quantitative quantum chemistry. *Mol Phys* 2008, 106:2107.
- ¹³ Fabian WMF. Accurate thermochemistry from quantum chemical calculations? *Monatsch Chem* 2008, 139:309.
- ¹⁴ Martin JML. Computational Thermochemistry: A Brief Overview of Quantum Mechanical Approaches. *Annu Rep Comput Chem* 2005, 1:31.
- ¹⁵ Helgaker T, Klopper W, Bak KL, Halkier A, Jørgensen P, Olsen J. Highly accurate ab initio computation of thermochemical data. *Understanding Chemical Reactivity, Vol. 22: Quantum-Mechanical Prediction of Thermochemical Data*, J. Cioslowski, ed. (Kluwer, Dordrecht, 2001), pp. 1–30.

-
- ¹⁶ Martin JML, Parthiban, S. W1 and W2 theory and their variants: thermochemistry in the kJ/mol accuracy range. *Understanding Chemical Reactivity*, Vol. 22: Quantum-Mechanical Prediction of Thermochemical Data, J. Cioslowski, ed. (Kluwer, Dordrecht, 2001), pp. 31–65.
- ¹⁷ Shavitt I, Bartlett RJ. *Many-Body Methods in Chemistry and Physics: Many-Body Perturbation Theory and Coupled-Cluster Theory*. Cambridge, UK: Cambridge University Press; 2009.
- ¹⁸ Pople JA. Two-Dimensional Chart of Quantum Chemistry. *J Chem Phys* 1965, 13:S229.
- ¹⁹ Korth M, Thiel W. Benchmarking Semiempirical Methods for Thermochemistry, Kinetics, and Noncovalent Interactions: OMx Methods Are Almost As Accurate and Robust As DFT-GGA Methods for Organic Molecules. *J Chem Theory Comput* 2011, 7:2929.
- ²⁰ Ramabhadran RO, Sengupta A, Raghavachari K. Application of the Generalized Connectivity-Based Hierarchy to Biomonomers: Enthalpies of Formation of Cysteine and Methionine. *J Phys Chem A* 2013, 117:4973.
- ²¹ Karton A, Yu L-J, Kesharwani MK, Martin JML. Heats of formation of the amino acids re-examined by means of W1-F12 and W2-F12 theories. *Theor Chem Acc* 2014, 133:1483.
- ²² Ley D, Gerbig D, Schreiner PR. Tunnelling control of chemical reactions – the organic chemist’s perspective. *Org Biomol Chem* 2012, 10:3781.
- ²³ Karton A, Parthiban S, Martin JML. Post-CCSD(T) ab Initio Thermochemistry of Halogen Oxides and Related Hydrides XOX, XOOX, HOX, XOn, and HXOn (X = F, Cl), and Evaluation of DFT Methods for These Systems. *J Phys Chem A* 2009, 113:4802.
- ²⁴ Feller D, Peterson KA, Dixon DA. Refined Theoretical Estimates of the Atomization Energies and Molecular Structures of Selected Small Oxygen Fluorides. *J Phys Chem A* 2010, 114:613.
- ²⁵ Feller D, Peterson KA, Dixon DA. Further benchmarks of a composite, convergent, statistically-calibrated coupled cluster-based approach for thermochemical and spectroscopic studies. *Mol Phys* 2012, 110:2381.
- ²⁶ Karton A, Martin JML. Atomization energies of the carbon clusters C_n ($n = 2–10$) revisited by means of W4 theory as well as density functional, Gn, and CBS methods. *Mol Phys* 2009, 107:977.
- ²⁷ Feller D, Peterson KA, Ruscic B. Improved accuracy benchmarks of small molecules using correlation consistent basis sets. *Theor Chem Acc* 2014, 133:1407.
- ²⁸ Karton A, Taylor PR, Martin JML. Basis set convergence of post-CCSD contributions to molecular atomization energies. *J Chem Phys* 2007, 127:064104.
- ²⁹ Tajti A, Szalay PG, Császár AG, Kállay M, Gauss J, Valeev EF, Flowers BA, Vázquez J, Stanton JF. HEAT: High accuracy extrapolated ab initio thermochemistry. *J Chem Phys* 2004, 121:11599.
- ³⁰ Bomble YJ, Vázquez J, Kállay M, Michauk C, Szalay PG, Császár AG, Gauss J, Stanton JF. High-accuracy extrapolated ab initio thermochemistry. II. Minor improvements to the protocol and a vital simplification. *J Chem Phys* 2006, 125:064108.

-
- ³¹ Harding ME, Vázquez J, Ruscic B, Wilson AK, Gauss J, Stanton JF. High-accuracy extrapolated ab initio thermochemistry. III. Additional improvements and overview. *J Chem Phys* 2008, 128:114111.
- ³² Karton A, Rabinovich E, Martin JML, Ruscic B. W4 theory for computational thermochemistry: In pursuit of confident sub-kJ/mol predictions. *J Chem Phys* 2006, 125:144108.
- ³³ Allen WD, East ALL, Császár AG. Ab Initio Anharmonic Vibrational Analyses of Non-Rigid Molecules. In *Structures and Conformations of Non-Rigid Molecules*, edited by J. Laane, M. Dakkouri, B. van der Veken, and H. Oberhammer (Kluwer, Dordrecht, 1993), p. 343.
- ³⁴ East ALL, Allen WD. The heat of formation of NCO. *J Chem Phys* 1993, 99:4638.
- ³⁵ Klippenstein SJ, East ALL, Allen WD. A high level ab initio map and direct statistical treatment of the fragmentation of singlet ketene. *J Chem Phys* 1996, 105:118.
- ³⁶ Császár AG, Allen WD, Schaefer HF. In pursuit of the ab initio limit for conformational energy prototypes. *J Chem Phys* 1998, 108:9751.
- ³⁷ Tarczay G, Császár AG, Klopper W, Szalay V, Allen WD, Schaefer HF. The barrier to linearity of water. *J Chem Phys* 1999, 110:11971.
- ³⁸ Császár AG, Leininger ML, Burcat A. Enthalpy of Formation of ${}^2\Pi_{3/2}$ SH. *J Phys Chem A* 2003, 107:2061.
- ³⁹ Császár AG, Leininger ML, Szalay V. The standard enthalpy of formation of CH₂. *J Chem Phys* 2003, 118:10631.
- ⁴⁰ Császár AG, Furtenbacher T. From a Network of Computed Reaction Enthalpies to Atom-Based Thermochemistry (NEAT). *Chem. Eur. J.* 2010, 16:4826.
- ⁴¹ Dixon DA, Feller D, Peterson KA. A practical guide to reliable first principles computational thermochemistry predictions across the periodic table. *Annu Rep Comput Chem* 2012, 8:1.
- ⁴² Boese AD, Oren M, Atasoylu O, Martin JML, Kállay M, Gauss J. W3 theory: Robust computational thermochemistry in the kJ/mol accuracy range. *J Chem Phys* 2004, 120:4129.
- ⁴³ Dunning, TH. Gaussian Basis Sets for Use in Correlated Molecular Calculations. I. The Atoms Boron through Neon and Hydrogen. *J Chem Phys* 1989, 90:1007.
- ⁴⁴ Kendall RA, Dunning, TH, Harrison RJ. Electron Affinities of the First-Row Atoms Revisited. Systematic Basis Sets and Wave Functions. *J Chem Phys* 1992, 96:6796.
- ⁴⁵ Dunning, TH, Peterson KA, Wilson AK. Gaussian basis sets for use in correlated molecular calculations. X. The atoms aluminum through argon revisited. *J Chem Phys* 2001, 114:9244.
- ⁴⁶ Klopper W. Highly accurate coupled-cluster singlet and triplet pair energies from explicitly correlated calculations in comparison with extrapolation techniques. *Mol Phys* 2001, 99:481.

-
- ⁴⁷ Peterson KA, Dunning, TH. Accurate correlation consistent basis sets for molecular core-valence correlation effects. The second row atoms Al–Ar, and the first row atoms B–Ne revisited. *J Chem Phys* 2002, 117:10548.
- ⁴⁸ Ruscic B, Pinzon RE, Morton ML, von Laszewski G, Bittner SJ, Nijssure SG, Amin KA, Minkoff M, Wagner AF. Introduction to Active Thermochemical Tables: Several “Key” Enthalpies of Formation Revisited. *J Phys Chem A* 2004, 108:9979.
- ⁴⁹ Ruscic B, Pinzon RE, von Laszewski G, Kodeboyina D, Burcat A, Leahy D, Montoya D, Wagner AF. Active Thermochemical Tables: Thermochemistry for the 21st Century. *J Phys Conf Ser* 2005, 16:561.
- ⁵⁰ Ruscic B. Uncertainty Quantification in Thermochemistry, Benchmarking Electronic Structure Computations, and Active Thermochemical Tables. *Int J Quant Chem* 2014, 114:1097.
- ⁵¹ Karton A, Martin JML. Comment on “Revised electron affinity of SF₆ from kinetic data” [J. Chem. Phys. 136, 121102 (2012)]. *J Chem Phys* 2012, 136:197101.
- ⁵² Karton A. How large are post-CCSD(T) contributions to the total atomization energies of medium-sized alkanes? *Chem Phys Lett* 2016, 645:118.
- ⁵³ Harding ME, Vázquez J, Gauss J, Stanton JF, Kállay M. Towards highly accurate ab initio thermochemistry of larger systems: Benzene. *J Chem Phys* 2011, 135:044513.
- ⁵⁴ Karton A, Kaminker I, Martin JML. Economical Post-CCSD(T) Computational Thermochemistry Protocol and Applications to Some Aromatic Compounds. *J Chem Phys* 2009, 113:7610.
- ⁵⁵ Karton A, Martin JML. Explicitly correlated *W_n* theory: W1-F12 and W2-F12. *J Chem Phys* 2012, 136:124114.
- ⁵⁶ Chan B, Radom L. W3X: A Cost-Effective Post-CCSD(T) Composite Procedure. *J Chem Theory Comput* 2013, 9:4769.
- ⁵⁷ Chan B, Radom L. W2X and W3X-L: Cost-Effective Approximations to W2 and W4 with kJ mol⁻¹ Accuracy. *J Chem Theory Comput* 2015, 11:2109.
- ⁵⁸ Klopper W, Manby FR, Ten-no S, Valeev EF. R12 methods in explicitly correlated molecular electronic structure theory. *Int Rev Phys Chem* 2006, 25:427.
- ⁵⁹ Peterson KA, Adler TB, and Werner H-J. Systematically convergent basis sets for explicitly correlated wavefunctions. The atoms H, He, B–Ne, and Al–Ar. *J Chem Phys* 2008, 128:084102.
- ⁶⁰ Klopper W, Ruscic B, Tew DP, Bischoff FA, Wolfsegger S. Atomization energies from coupled-cluster calculations augmented with explicitly-correlated perturbation theory. *Chem Phys* 2009, 356:14.
- ⁶¹ Vogiatzis KD, Klopper W. Accurate non-covalent interactions with basis-set corrections from interference-corrected perturbation theory: comparison with the S22B database. *Mol Phys* 2013, 111:2299.

-
- ⁶² Vogiatzis KD, Haunschild R, Klopper W. Accurate atomization energies from combining coupled-cluster computations with interference-corrected explicitly correlated second-order perturbation theory. *Theor Chem Acc* 2014, 133:1446.
- ⁶³ J. F. Stanton. Why CCSD(T) works: a different perspective. *Chem Phys Lett* 1997, 281:130.
- ⁶⁴ Bak KL, Jørgensen P, Olsen J, Helgaker T, Gauss J. Coupled-cluster singles, doubles and triples (CCSDT) calculations of atomization energies, *Chem Phys Lett* 2000, 317:116.
- ⁶⁵ Raghavachari K. Historical perspective on: A fifth-order perturbation comparison of electron correlation theories [Volume 157, Issue 6, 26 May 1989, Pages 479–483]. *Chem Phys Lett* 2013, 589:35.
- ⁶⁶ Curtiss LA, Redfern PC, Raghavachari K. Gaussian-4 theory. *J Chem Phys* 2007, 126:084108.
- ⁶⁷ Curtiss LA, Redfern PC, Raghavachari K. Gaussian-4 theory using reduced order perturbation theory. *J Chem Phys* 2007, 127:124105.
- ⁶⁸ Chan B, Deng J, Radom L. G4(MP2)-6X: A Cost-Effective Improvement to G4(MP2). *J Chem Theory Comput* 2011, 7:112.
- ⁶⁹ Ochterski JW, Petersson GA, Montgomery Jr JA. A complete basis set model chemistry. V. Extensions to six or more heavy atoms. *J Chem Phys* 1996, 104:2598.
- ⁷⁰ Montgomery Jr JA, Frisch MJ, Ochterski JW, Petersson GA. A complete basis set model chemistry. VI. Use of density functional geometries and frequencies. *J Chem Phys* 1999, 110:2822.
- ⁷¹ Wan W, Karton A. Heat of formation for C₆₀ by means of the G4(MP2) thermochemical protocol through reactions in which C₆₀ is broken down into corannulene and sumanene. *Chem Phys Lett* 2016, 643:34.
- ⁷² Martin JML, de Oliveira G. Towards standard methods for benchmark quality ab initio thermochemistry—W1 and W2 theory. *J Chem Phys* 1999, 111:1843.
- ⁷³ The 14 highly multireference systems, for which the %TAE[(T)] ≥ 10%, are: Be₂, B₂, C₂(¹Σ_g⁺), BN(¹Σ⁺), OF, F₂O, FO₂, F₂O₂, Cl₂O, ClOO, OClO, O₃, S₃, and S₄.
- ⁷⁴ Karton A, Schreiner PR, and Martin JML. Heats of Formation of Platonic Hydrocarbon Cages by Means of High-Level Thermochemical Procedures. *J Comput Chem* 2016, 37:49.
- ⁷⁵ Karton A, Chan B, Raghavachari K, Radom L. Evaluation of the heats of formation of corannulene and C₆₀ by means of high-level theoretical procedures. *J Phys Chem A* 2013, 117:1834.
- ⁷⁶ Agapito F, Santos RC, Martinho Simões JA. Energetics of Nonbonded Ortho Interactions in Alkylbenzenes. *J Phys Chem A* 2013, 117:2873.
- ⁷⁷ Gonçalves EM, Agapito F, Almeida TS, Martinho Simões JA. Enthalpies of formation of dihydroxybenzenes revisited: Combining experimental and high-level ab initio data. *J Chem Thermodynamics* 2014, 73:90.

-
- ⁷⁸ Emel'yanenko VN, Zaitseva KV, Agapito F, Martinho Simões JA, Verevkin SP. Benchmark thermodynamic properties of methylanisoles: Experimental and theoretical study. *J Chem Thermodynamics* 2015, 85:155.
- ⁷⁹ Simões RG, Agapito F, Diogo HP, Minas da Piedade ME. Enthalpy of Formation of Anisole: Implications for the Controversy on the O–H Bond Dissociation Enthalpy in Phenol. *J Phys Chem A* 2014, 118:11026.
- ⁸⁰ DeYonker NJ, Cundari TR, Wilson AK. The correlation consistent composite approach (ccCA): An alternative to the Gaussian-n methods. *J Chem Phys* 2006, 124:114104.
- ⁸¹ DeYonker NJ, Grimes T, Yockel S, Dinescu A, Mintz B, Cundari TR, Wilson AK. The correlation-consistent composite approach: Application to the G3/99 test set. *J Chem Phys* 2006, 125: 104111.
- ⁸² DeYonker NJ, Wilson BR, Pierpont AW, Cundari TR, Wilson AK. Towards the intrinsic error of the correlation consistent Composite Approach (ccCA). *Mol Phys* 2009, 107:1107.
- ⁸³ Mahler A, Wilson AK. Explicitly Correlated Methods within the ccCA Methodology. *J Chem Theor Comput* 2013, 9:1402.
- ⁸⁴ Chan B, Radom L. W1X-1 and W1X-2: W1-Quality Accuracy with an Order of Magnitude Reduction in Computational Cost. *J Chem Theory Comput* 2012, 8:4259.
- ⁸⁵ Klopper W, Bachorz RA, Hättig C, Tew DP. Accurate computational thermochemistry from explicitly correlated coupled-cluster theory. *Theor Chem Acc* 2010, 126:289.
- ⁸⁶ Vogiatzis KD, Barnes EC, Klopper W. Interference-corrected explicitly-correlated second-order perturbation theory. *Chem Phys Lett* 2011, 503:157.
- ⁸⁷ Krechkivska O, Wilcox CM, Chan B, Jacob R, Liu Y, Nauta K, Kable SH, Radom L, Schmidt TW. H and D Attachment to Naphthalene: Spectra and Thermochemistry of Cold Gas-Phase 1-C₁₀H₉ and 1-C₁₀H₈D Radicals and Cations. *J Phys Chem A* 2015, 119:3225.
- ⁸⁸ Lee TJ, Taylor PR. A diagnostic for determining the quality of single-reference electron correlation methods. *Int J Quantum Chem* 1989, 36:199.
- ⁸⁹ Janssen CL, Nielsen IMB. New diagnostics for coupled-cluster and Møller–Plesset perturbation theory. *Chem Phys Lett* 1998, 290:423.
- ⁹⁰ Watts JD, Bartlett RJ. Coupled-Cluster singles, doubles, and triples calculations with hartree–fock and brueckner orbital reference determinants: A comparative study. *Int J Quantum Chem* 1994, 52:195.
- ⁹¹ Watts JD, Urban M, Bartlett RJ. Accurate electrical and spectroscopic properties of X ¹Σ⁺ BeO from coupled-cluster methods. *Theor Chim Acta* 1995, 90:341.
- ⁹² Zheng J, Zhao Y, Truhlar DG. Thermochemical Kinetics of Hydrogen-Atom Transfers between Methyl, Methane, Ethynyl, Ethyne, and Hydrogen. *J Phys Chem A* 2007, 111:4632.

-
- ⁹³ Cowan RD, Griffin M. Approximate relativistic corrections to atomic radial wave functions. *J. Opt. Soc. Am.* 1976, 66:1010.
- ⁹⁴ Martin RL. All-Electron Relativistic Calculations on AgH. An Investigation of the Cowan–Griffin Operator in a Molecular Species. *J Phys Chem* 1983, 87:750.
- ⁹⁵ Douglas M, Kroll NM. Quantum electrodynamical corrections to the fine structure of helium. *Ann Phys* 1974, 82:89.
- ⁹⁶ Heß BA. Relativistic electronic-structure calculations employing a two-component no-pair formalism with external-field projection operators. *Phys Rev A* 1986, 33:3742.
- ⁹⁷ Collins CL, Dyall KG, Schaefer HF. Relativistic and correlation effects in CuH, AgH, and AuH: Comparison of various relativistic methods. *J Chem Phys* 1995, 102:2024.
- ⁹⁸ Barysz M, Sadlej, AJ. Two-component methods of relativistic quantum chemistry: from the Douglas–Kroll approximation to the exact two-component formalism. *J Molec. Struct. (Theochem)* 2001, 573:181.
- ⁹⁹ de Jong WA, Harrison RJ, Dixon DA. Parallel Douglas–Kroll energy and gradients in NWChem: Estimating scalar relativistic effects using Douglas–Kroll contracted basis sets. *J Chem Phys* 2001, 114:48.
- ¹⁰⁰ Mayhall NJ, Raghavachari K, Redfern PC, Curtiss LA, Rassolov V. Toward accurate thermochemical models for transition metals: G3Large basis sets for atoms Sc–Zn. *J Chem Phys* 2008, 128:144122.
- ¹⁰¹ Mayhall NJ, Raghavachari K, Redfern PC, Curtiss LA. Investigation of Gaussian-4 theory for transition metal thermochemistry. *J Phys Chem A* 2009, 113:5170.
- ¹⁰² Parthiban S, de Oliveira G, Martin JML. Benchmark ab Initio Energy Profiles for the Gas-Phase S_N2 Reactions $Y^- + CH_3X \rightarrow CH_3Y + X^-$ (X, Y = F, Cl, Br). Validation of Hybrid DFT Methods. *J Phys Chem A* 2001, 105:895.
- ¹⁰³ Oren M, Iron MA, Burcat A, Martin JML. Thermodynamic Properties of C_1 and C_2 Bromo Compounds and Radicals: a Relativistic Ab Initio Study. *J Phys Chem A* 2004, 108:7752.
- ¹⁰⁴ DeYonker NJ, Peterson KA, Steyl G, Wilson AK, Cundari TR. Quantitative Computational Thermochemistry of Transition Metal Species. *J Phys Chem A* 2007, 111:11269.
- ¹⁰⁵ Jiang W, DeYonker NJ, Determan JJ, Wilson AK. Toward Accurate Theoretical Thermochemistry of First Row Transition Metal Complexes. *J Phys Chem A* 2012, 116:870.
- ¹⁰⁶ Li S, Hennigan JM, Dixon DA, Peterson KA. Accurate Thermochemistry for Transition Metal Oxide Clusters. *J Phys Chem A* 2009, 113:7861.
- ¹⁰⁷ Bross DH, Hill JG, Werner H–J, Peterson KA. Explicitly correlated composite thermochemistry of transition metal species. *J Chem Phys* 2013, 139:094302.

-
- ¹⁰⁸ Bross DH, Peterson KA. Composite thermochemistry of gas phase U(VI)-containing molecules. *J Chem Phys* 2014, 141:244308.
- ¹⁰⁹ Bross DH, Parmar P, Peterson KA. Multireference configuration interaction calculations of the first six ionization potentials of the uranium atom. *J Chem Phys* 2015, 143:184308.
- ¹¹⁰ Thanthiriwatte KS, Vasiliu M, Battey SR, Lu Q, Peterson KA, Andrews L, Dixon DA. Gas phase properties of MX₂ and MX₄ (X=F, Cl) for M = Group 4, Group 14, Ce, and Th. *J Phys Chem A* 2015, 119:5790.
- ¹¹¹ Karton A, Ruscic B, Martin JML. Benchmark atomization energy of ethane: Importance of accurate zero-point vibrational energies and diagonal Born–Oppenheimer corrections for a ‘simple’ organic molecule. *J Mol Struct Theochem* 2007, 811:345.
- ¹¹² Karton A, Gruzman D, Martin JML. Benchmark Thermochemistry of the C_nH_{2n+2} Alkane Isomers (n = 2–8) and Performance of DFT and Composite Ab Initio Methods for Dispersion-Driven Isomeric Equilibria. *J Phys Chem A* 2009, 113:8434.
- ¹¹³ Tajti A, Szalay PG, Gauss J. Perturbative treatment of the electron-correlation contribution to the diagonal Born–Oppenheimer correction. *J Chem Phys* 2007, 127:014102.
- ¹¹⁴ Schuurman MS, Allen WD, Schaefer HF. The ab initio limit quartic force field of BH₃. *J Comput Chem* 2005, 26:1106.
- ¹¹⁵ Csonka GI, Ruzsinszky A, Perdew JP. Estimation, computation, and experimental correction of molecular zero-point vibrational energies. *J Phys Chem A* 2005, 109:6779.
- ¹¹⁶ Grev RS, Jenssen CL, Schaefer HF. Concerning zero-point vibrational energy corrections to electronic energies. *J Chem Phys* 1991, 95:5128.
- ¹¹⁷ Pople JA, Scott AP, Wong, MW, Radom L. Scaling Factors for Obtaining Fundamental Vibrational Frequencies and Zero-Point Energies from HF/6-31G* and MP2/6-31G* Harmonic Frequencies. *Isr J Chem* 1993, 33:345.
- ¹¹⁸ Scott AP, Radom L. Harmonic Vibrational Frequencies: An Evaluation of Hartree–Fock, Møller–Plesset, Quadratic Configuration Interaction, Density Functional Theory, and Semiempirical Scale Factors. *J Phys Chem* 1996, 100:16502.
- ¹¹⁹ Sinha P, Boesch SE, Gu C, Wheeler RA, Wilson AK. Harmonic Vibrational Frequencies: Scaling Factors for HF, B3LYP, and MP2 Methods in Combination with Correlation Consistent Basis Sets. *J Phys Chem A* 2004, 108:9213.
- ¹²⁰ Merrick JP, Moran D, Radom L. An Evaluation of Harmonic Vibrational Frequency Scale Factors. *J Phys Chem A* 2007, 111:11683.
- ¹²¹ Irikura, KK; Johnson III, RD, Kacker RN, Kessel RJ. Uncertainties in scaling factors for ab initio vibrational zero-point energies. *J Chem Phys* 2009, 130:114102.

¹²² Alecu I, Zheng J, Zhao Y, Truhlar DG. Computational Thermochemistry: Scale Factor Databases and Scale Factors for Vibrational Frequencies Obtained from Electronic Model Chemistries. *J Chem Theory Comput* 2010, 6:2872.

¹²³ Laury ML, Carlson MJ, Wilson AK. Vibrational Frequency Scale Factors for Density Functional Theory and the Polarization Consistent Basis Sets. *J Comput Chem* 2012, 33:2380.

¹²⁴ Martin JML, Kesharwani MK. Assessment of CCSD(T)-F12 approximations and basis sets for harmonic vibrational frequencies. *J Chem Theory Comput* 2014, 10:2085.

¹²⁵ Kesharwani MK, Brauer B, Martin JML. Frequency and zero-point vibrational energy scale factors for double hybrid density functionals (and selected other methods): can anharmonic force fields be avoided? *J Phys Chem A* 2015, 119:1701.

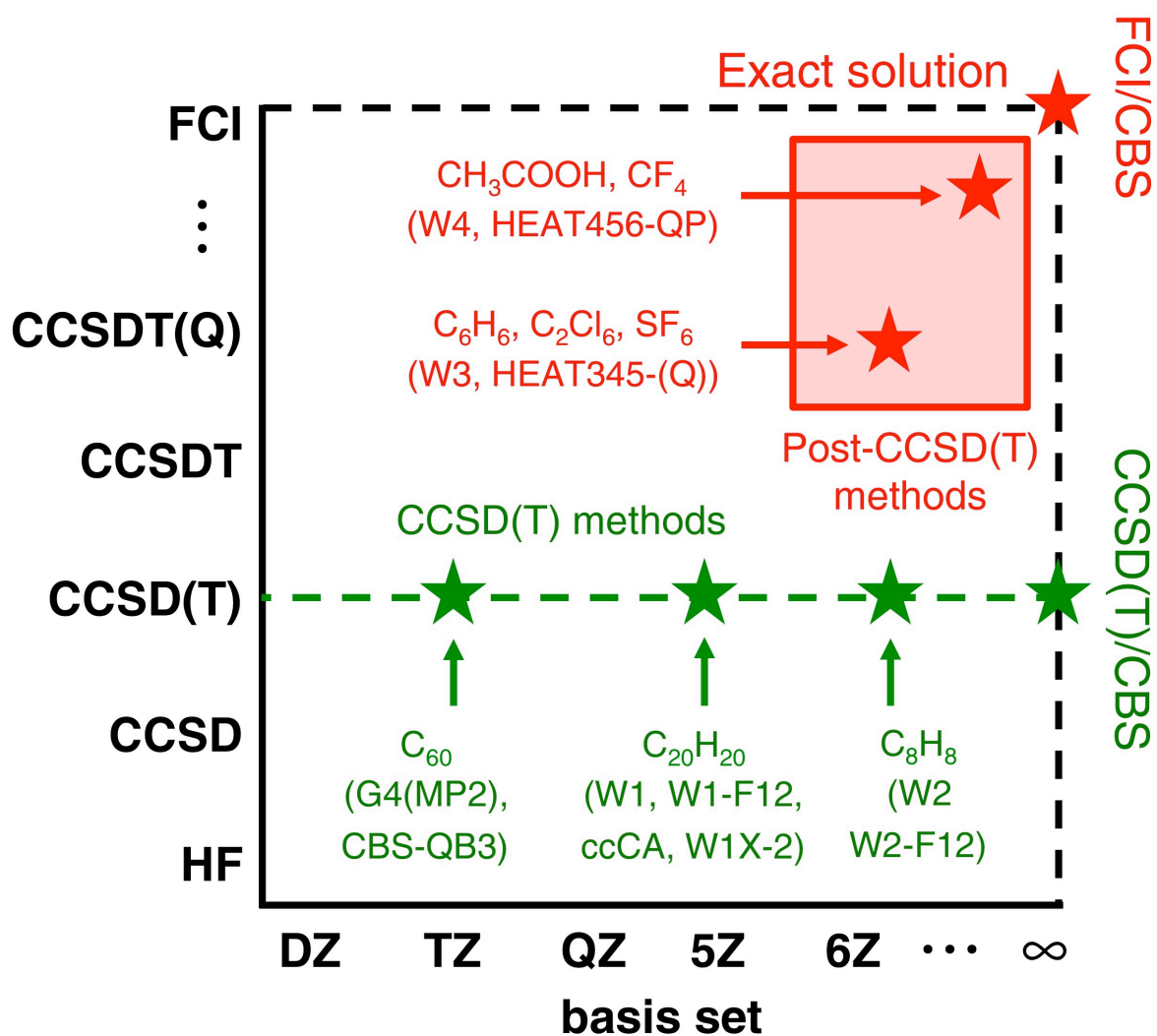


Figure 1. A modified Pople diagram illustrating the electronic energies that CCSD(T) and post-CCSD(T) composite ab initio methods attempt to approximate, and examples of species to which these methods have been applied (examples of popular composite methods are given in parenthesis).

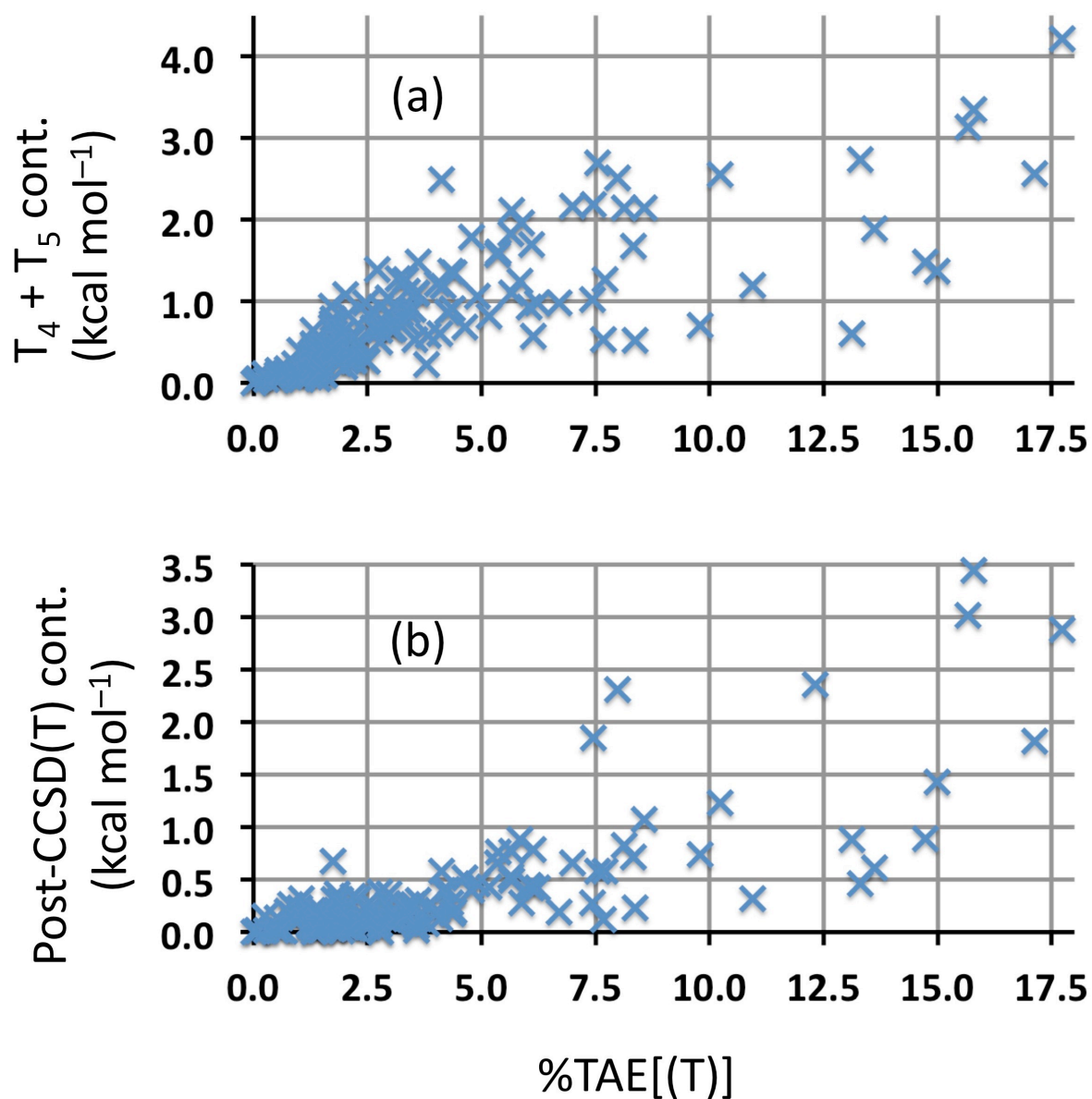


Figure 2. Plots of the %TAE[(T)] diagnostic and the magnitude of the (a) T₄ + T₅ and (b) post-CCSD(T) contributions to the total atomization energies for the 140 species in the W4-11 database.

Tables

Table 1 | Overview of the characteristics of fixed-recipe CCSD(T) and post-CCSD(T) composite ab initio methods

Type	Post-CCSD(T) methods		CCSD(T) methods	
	FCI/CBS	CCSDT(Q)/CBS	CCSD(T)/CBS	CCSD(T)/TZ
Some examples ^a	W4, ³² W4.2, ³² W4.3, ³² W4.4, ²⁸ HEAT-345QP, ³⁰ HEAT-456QP ³¹	W3-F12, ⁵⁵ W3.2lite, ⁵⁴ W4lite, ³² HEAT- 345(Q), ³⁰ HEAT- 456(Q), ³¹ W3X, ⁵⁶ W3X-L ⁵⁷	W1, ⁷² W2, ⁷² W1-F12, ⁵⁵ W2-F12, ⁵⁵ ccCA, ⁸⁰ ccCA- F12, ⁸³ W1X-1, ⁸⁴ W1X-2, ⁸⁴ W2X, ⁵⁷ CCSD(T)+F12+INT ⁶¹	G4, ⁶⁶ G4(MP2), ⁶⁷ G4(MP2)-6X, ⁶⁸ CBS-APNO, ⁶⁹ CBS-QB3 ⁷⁰
Accuracy for TAEs ^b	< 1 kJ mol ⁻¹	~1 kJ mol ⁻¹	~1 kcal mol ⁻¹	> 1 kcal mol ⁻¹
Applicability ^c	CF ₄ , ⁸ CH ₃ COOH ⁸	SF ₆ ⁻ , ⁵¹ C ₆ H ₁₄ , ⁵² C ₆ H ₆ ^{53,54}	C ₂₀ H ₂₀ ⁷⁴	C ₆₀ ⁷¹
Multireference ^d	Yes	Yes ^e	No ^f	No ^f
Empirical parameters	No	No ^g	No ^g	Yes

^aThis is not an exhaustive list of all the fixed-recipe composite procedures that have been developed in recent years, any omissions are by no means intentional. ^b95% confidence intervals for total atomization reactions (see also Table 2 and text). ^cRelatively large systems for which these methods have been recently applied to. ^dApplicability to multireference systems. ^eSome caution is needed for pathologically multireference systems (e.g., O₃ and C_n clusters) and/or second-row systems (e.g., S₄ and SO₃), see text. ^fFor systems with %TAE[(T)] diagnostics larger than ~5% the post-CCSD(T) methods are recommended, see text. ^gSome of these methods employ empirically-motivated scaling factors derived from high-level theoretical calculations.

Table 2 | Overview of the error statistics of post-CCSD(T) composite ab initio methods relative to experimental total atomization energies from the Active Thermochemical Tables (ATcT) thermochemical network (kcal mol⁻¹)

Type	Method	Test set	MSD ^a	MAD ^b	RMSD ^c	95% CI (2σ) ^d	Max Deviations ^e
FCI/CBS	W4.x ^f	I ^g	-0.019	0.039	0.060	0.120	-0.140 (O ₃), +0.110 (C ₂ H ₄)
	W4	I ^g	-0.012	0.065	0.085	0.170	-0.230 (O ₃), +0.190 (HONO)
	W4	II ^h	0.003	0.056	0.072	0.144	-0.130 (H ₂ O ₂), +0.137 (C ₂ H)
	HEAT-456QP	II ^h	-0.057	0.082	0.100	0.200	-0.208 (C ₂ H ₂), +0.072 (OH)
	HEAT-345QP	II ^h	0.002	0.062	0.068	0.135	-0.093 (H ₂ O ₂), +0.105 (OH)
CCSDT(Q)/CBS	W4lite	I ^g	-0.031	0.093	0.125	0.250	-0.330 (Cl ₂), +0.230 (HONO)
	W4lite	II ^h	-0.017	0.071	0.090	0.180	-0.220 (O ₂), +0.120 (HF)
	W3.2	I ^g	-0.077	0.143	0.192	0.383	-0.620 (O ₃), +0.160 (C ₂ H ₆)
	W3.2	II ^h	-0.054	0.112	0.149	0.297	-0.380 (O ₂), +0.160 (H ₂ O)
	W3-F12	I ^g	-0.073	0.150	0.181	0.362	-0.339 (C ₂ H ₂), +0.368 (HONO)
	W3-F12	II ^h	-0.084	0.122	0.168	0.337	-0.339 (C ₂ H ₂), +0.286 (CO ₂)
	HEAT-456(Q)	II ^h	-0.023	0.079	0.101	0.202	-0.208 (C ₂ H ₂), +0.127 (HO ₂)
	HEAT-345(Q)	II ^h	0.036	0.070	0.083	0.166	-0.065 (CH ₂), +0.153 (CO ₂)

^aMean-signed deviation. ^bMean-absolute deviation. ^cRoot-mean-square deviation. ^dThe 95% (2σ) confidence intervals are taken as twice the RMSD. ^eMaximum negative and positive deviations (theory-experiment). ^fFor a subset of 29 molecules for which TAEs from W4.2 theory (or higher) are available (see also footnote g and text). ^gThis set comprises of 35 first- and second-row ATcT atomization energies associated with error bars ≤ 0.05 kcal mol⁻¹, including pathologically multireference systems such as O₃ (see ref. 8 and text). ^hThis set comprises of 18 first-row ATcT TAEs associated with error bars ≤ 0.06 kcal mol⁻¹ (see ref. 31 and text).

Related Articles

Article ID [or Subtopic]	Article title
CMS-079	Higher-Order Coupled-Cluster Theory
CMS-084	Explicitly Correlated Ab Initio Methods
CMS-093	Relativistic Douglas-Kroll-Hess theory
CMS-122	Correlation Consistent Basis Sets
CMS-125	Møller-Plesset Perturbation Theory
CMS-127	Core Correlation Effects
CMS-150	Gn-Theory