Quantum mechanical thermochemical predictions 100 years after the Schrödinger equation

Amir Karton*
School of Science and Technology, University of New England, Armidale, NSW, Australia
*Corresponding author: e-mail address: amir.karton@une.edu.au

Contents

Abstract

Twenty-five years ago, the two main pillars of quantum chemistry—density functional and composite ab initio theories—were recognized with a Nobel Prize in Chemistry awarded to Walter Kohn and John Pople. This recognition sparked intense theoretical developments in both fields. Whereas in 1998, the year the Nobel Prize was awarded, there were only a handful of composite ab initio methods; most notably the Gaussian-*n* methods ($n = 1–3$), CBS methods (e.g., CBS-QCI and CBS-APNO), and the focal-point analysis approach, today there are many more families of such methods, including the Weizmann-*n*, MCCM, HEAT, ccCA, FPD, ATOMIC, INT-MP2-F12, and ChS family of methods, where some of these families include dozens

Annual Reports in Computational Chemistry
ISSN 1574-1400
https://doi.org/10.1016/bs.arcc.2022.09.003

of variants. Overall, there are over 100 contemporary variants of composite ab initio methods to choose from, with many variants implemented as a keyword in popular quantum chemical packages. This situation makes it difficult to choose a proper method for a given chemical system, property, and desired accuracy. This chapter provides an overview of contemporary composite ab initio methods applicable to first- and second-row elements, their main energetic components, and their expected accuracy and applicability. To guide the selection of a suitable method for a given chemical system and desired accuracy, the various methods are classified according to a 'Jacob's Ladder' of composite ab initio methods, from computationally economical methods that are capable of approaching chemical accuracy to computationally demanding methods capable of confident sub-benchmark accuracy.

1. Introduction

Shortly after the Schrödinger equation was developed, arguably the greatest physicist of the 20th century, Paul Dirac, stated that quantum mechanics could be used to understand "the whole of chemistry." *(1)* Such a bold statement would have seemed like a far-reaching dream at a time when the new theory was only applicable to one- and two-electron systems like the helium atom and H_2 molecule *(2)*. Indeed, Dirac quickly hedged this statement by adding "the difficulty is only that the exact application of these laws leads to equations much too complicated to be soluble." This statement is still true today; however, it was Dirac's next sentence that shaped the development of quantum chemistry over the past century "It therefore becomes desirable that approximate practical methods of applying quantum mechanics should be developed, which can lead to an explanation of the main features of complex atomic systems". In this insightful quote, which was written well before the age of computers, let alone supercomputers, Dirac captures quite succinctly two key aspects of contemporary computational quantum chemistry (i) the development of computationally economical quantum chemical theories and (ii) the application of these theories for exploring and predicting the electronic structure of molecules and materials. For most of the 20th century, quantum chemical theories were still applicable to fairly small systems; however, owing to significant developments in quantum chemical theory and supercomputer technology over the past three decades, Dirac's dream has now been realized for complex chemical systems across the Periodic Table *(3–5)*. A major stepping-stone in realizing this goal was the development of quantum chemical theories that are both highly accurate (i.e., capable of thermochemical and kinetic predictions with confident sub-kcal/mol accuracy)

and applicable to medium-sized chemical systems with dozens of atoms. These quantum chemical theories are commonly referred to as *composite ab initio methods* (sometimes also called *ab initio thermochemistry methods* or *compound thermochemistry methods*). These procedures use a series of high-level *ab initio* calculations to obtain accurate thermochemical and kinetic data comparable to experimental data. This field began with the development of the so-called Gaussian-1 (G1) theory by Pople and co-workers in the late 1980s *(6,7)*. Subsequently, major advances in quantum chemical theory and high-performance supercomputer technology have allowed this field to flourish into a widely applied subfield of quantum chemistry. Over the past 30-odd years, a wide range of composite ab initio methods has been developed. Popular examples include the Gaussian-*n* (G*n*) methods *(6–10)* and variants thereof *(11–18)*, complete basis set (CBS) model chemistries *(19–25)*, focal-point analysis (FPA) *(26–30)*, Weizmann-*n* (W*n*) *(31–36)*, W*n*X *(37–39)*, multi-coefficient correlation methods (MCCMs) *(40–45)*, high-accuracy extrapolated ab initio thermochemistry (HEAT) *(46–50)*, correlation consistent composite approach (ccCA) *(51–60)*, Feller–Peterson–Dixon (FPD) *(61–67)*, ab initio thermochemistry using optimal-balance models with isodesmic corrections (ATOMIC) *(68–70)*, interference-corrected explicitly correlated second-order perturbation theory (INT-MP2-F12) *(71)*, and the so-called cheap composite scheme (ChS) *(72,73)* procedures. Today, composite ab initio methods are among the most accurate means for examining chemical processes at the atomic level. These theories allow quantum mechanics not only to explain the chemistry of complex molecules but also to obtain chemical properties (such as reaction energies and barrier heights) with accuracy that rivals or exceeds the most accurate experiments. Thus, these methods are instrumental in (i) investigating transient species that are difficult to study experimentally (e.g., free radicals, transition structures, and short-lived reaction intermediates), (ii) modeling kinetics and mechanisms of challenging chemical reactions, and (iii) analyzing and explaining observed experimental trends. This chapter gives an overview of the various types of composite ab initio methods that have been developed over the past 30-odd years with an emphasis on key theoretical developments, design philosophy, confident sub-chemical accuracy, and strategies for choosing an appropriate method for a given chemical problem and desired level of accuracy. The present review focuses on main-group first- and second-row chemistry, for additional in-depth overviews of composite ab initio methods, the reader is referred to several excellent reviews that have been published over the past

two decades *(61,62,74–91)*. (For comprehensive reviews of coupled-cluster theory see references *(92–95)*, and references therein.)

1.1 Chemical accuracy and theoretical uncertainty

A primary goal of composite ab initio methods is to obtain accurate thermochemical and kinetic properties, if possible, with well-defined error bars. It is helpful to begin by discussing two levels of accuracy that are extensively used in computational thermochemistry, namely "chemical accuracy" and "benchmark accuracy." Chemical accuracy refers to an energy unit of 1.0 kcal $mol^{-1} = 4.184$ kJ mol^{-1}. It is convenient to define this unit of energy since it represents about 1% of the bond dissociation energy (BDE) of typical covalent bonds. Incidentally, $\sim$1 kcal mol^{-1} also represents a typical error bar for experimental tabulations in thermochemical databases such as the National Institute of Standards and Technology (NIST) *(96)*, which were instrumental in calibrating and benchmarking the first-generation composite ab initio methods.

It should be noted that the more recently developed Active Thermochemical Tables (ATcT) approach developed by Ruscic and co-workers is capable of producing thermochemical data with markedly lower error bars *(97,98)*. The development of ATcT played a key role in the development of high-level composite ab initio methods capable of sub-benchmark accuracy, *vide infra*. This illustrates that the development of more accurate next-generation theoretical procedures is often limited by the availability of sufficiently accurate and reliable experimental data needed to evaluate the performance of such theoretical procedures.

Indeed, the term "chemical accuracy" started gaining broad use in computational thermochemistry in the early 1990s alongside the development of the first CCSD(T) or QCISD(T)-based composite ab initio methods *(99)*. This terminology has been instrumental in setting a target accuracy for the early thermochemical methods. However, it should be stressed that chemical accuracy refers to a level of accuracy of 1 kcal mol^{-1}, but it does not specify how this level of accuracy should be quantified. For example, chemical accuracy may refer to mean-absolute deviation (MAD) $\leq$ 1 kcal mol^{-1}, root-mean-square deviation (RMSD) ≤ 1 kcal mol^{-1}, or 95% confidence interval (CI) ≤ 1 kcal mol^{-1} from sufficiently accurate experimental or theoretical data. The lack of a clear definition as to how to quantify chemical accuracy diminishes the value of this terminology.

The quantification of uncertainty in electronic structure calculations has been discussed in detail by Ruscic *(100)*, and we reiterate the recommendation that a 95% CI is the most robust measure of uncertainty in composite ab initio calculations. A 95% CI obtained for a given theoretical procedure, and thermochemical property relative to a sufficiently accurate and large/representative benchmark dataset means that a value calculated following the same procedure should lie within the quoted uncertainty 19 times out of 20 for species outside the dataset. Similarly, Peterson, Feller, and Dixon *(79)*, suggested that to achieve an experimentalist's notion of chemical accuracy, an appropriate definition would be twice the MAD being below 1 kcal mol^{-1}.

It is also important to stress that the performance of any quantum chemical method depends on the chemical property being calculated and the size/diversity of the dataset being used for the performance evaluation. For example, a given quantum chemical procedure may achieve chemical accuracy for isodesmic bond-separation energies but not for total atomization energies (TAEs) *(101,102)*; or achieve chemical accuracy for TAEs of species dominated by dynamical correlation but not TAEs of multireference species *(80,103)*.

The term chemical accuracy is widely used in various contexts to describe the level of accuracy of wavefunction-based methods and even in the context of density functional theory (DFT) calculations. Here we note that the following parameters should be specified in order for the term chemical accuracy to be more meaningful:

- The statistical metric used for defining chemical accuracy (e.g., MAD, RMSD, or 95% CI)
- The composition of the benchmark dataset used for the evaluation (e.g., in terms of the elemental composition or multireference character of the species involved)
- The chemical property that is being considered (e.g., TAEs, conformational energies, reaction barrier heights, or non-covalent interactions)

We note that once the statistical metric used for defining chemical accuracy is specified, it can be converted to a different one using the following guidelines: *(80,100)*

- $\mathrm{MAD} \approx \sqrt{\frac{2}{\pi}} \times \mathrm{RMSD} \approx 0.8 \times \mathrm{RMSD}$ (for a normal error distribution with a small systematic error)
- $95\%\ \mathrm{CI} \approx 2 \times \mathrm{RMSD}$

- 95% CI $\approx 2.5 \times$ MAD (for a normal error distribution with a small systematic error, however, as noted by Ruscic *(100)*, the conversion factor can reach up to 3.5 depending on the error distribution)

In the context of thermochemical and kinetic properties, such as bond dissociation energies, heats of formation, and reaction barrier heights, it is well accepted that chemical accuracy refers to 1 kcal mol^{-1}. However, chemical accuracy may refer to smaller energetic thresholds for less challenging thermochemical properties. By less challenging properties, we mean properties that benefit from a larger degree of systematic error cancelation between reactants and products. For example, it has been noted by Mardirossian and Head-Gordon *(104)* and by Mehta *et al. (105)*, that a value of 0.1 kcal mol^{-1} is more appropriate in the context of nonbonded interactions. Such nonbonded properties may include hydrogen and halogen bonding, dispersion interactions, and conformational isomerizations that do not involve covalent bond breaking *(106)*.

High-level composite ab initio methods that include contributions beyond the CCSD(T) level can obtain thermochemical and kinetic data with confident sub-chemical accuracy. Thus, it is helpful to define another level of accuracy of 1.0 kJ mol$^{-1} \approx 0.239$ kcal mol^{-1}, which is commonly referred to as "benchmark accuracy." It has been found that post-CCSD(T) composite ab initio methods such as the Weizmann-4 (W4) theory are capable of obtaining TAEs 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) *(33,76)*. Notably, this level of accuracy surpasses that of many traditional experimental thermochemical tabulations such as the NIST Chemistry WebBook *(96)* and Computational Chemistry Comparison and Benchmark DataBase (CCCBDB) *(107)*. However, the Active Thermochemical Tables, in general, have substantially higher accuracy. Having said that, it should also be pointed out that post-CCSD(T) composite ab initio methods such as W4 theory are only applicable to relatively small molecules with up to ~ 8 nonhydrogen atoms (e.g., CCl_4, SiF_4, C_6H_6, SF_6, and C_2Cl_6) *(76,80,103,108)*.

Composite ab initio methods are also used for the calculation of spectroscopic properties based on energy derivatives with respect to the nuclear coordinates, such as equilibrium bond distances (r_e), harmonic vibrational frequencies (ω_e), and first-order anharmonic corrections ($\omega_e\chi_e$). Table 1 gives an overview of common definitions for chemical and benchmark accuracies for thermochemical and spectroscopic properties.

Table 1 Common (and suggested) definitions for "chemical" and "benchmark" accuracies for different chemical properties.

Property	Chemical accuracy	Benchmark accuracy
Heats of formation[a]	1.0 kcal $mol^{-1} = 4.2$ kJ mol^{-1}	0.24 kcal $mol^{-1} = 1.0$ kJ mol^{-1}
Weak interactions	0.1 kcal $mol^{-1} = 0.42$ kJ mol^{-1}	0.024 kcal $mol^{-1} = 0.1$ kJ mol^{-1}
Bond distances	0.005 Å	0.001 Å
Vibrational frequencies	5.0 cm^{-1}	1.0 cm^{-1}

[a]Or other chemical properties involving multiple breaking/forming of bonds.

1.2 Limitations of single-point energy ab initio calculations

Table 2 gives the formal computational costs of relevant quantum chemical methods. With current mainstream computer technology, it is possible to run Hartree–Fock (HF) energy calculation for systems with hundreds of non-hydrogen atoms. This also applies to hybrid density functional theory, which involves both exact HF and DFT exchange. Hybrid-DFT has the same computational scaling as HF theory but is far more useful for describing most chemical properties *(104)*. Such calculations are normally carried out in a single step, i.e., via a single-point energy (SPE) calculation. A fundamental limitation of high-level ab initio methods is the exponential increase in computational cost with system size (Table 2). For example, the CCSD(T) method with formal scaling of $\sim N_{bas}{}^{7}$ (compared to $\sim N_{bas}{}^{4}$ for hybrid DFT) is applicable to systems with dozens of non-hydrogen atoms. Likewise, post-CCSD(T) SPE calculations are generally limited to systems with only a handful of non-hydrogen atoms. Furthermore, due to the exceedingly slow basis set convergence of coupled-cluster methods, very large Gaussian basis sets must be employed in such calculations in order to obtain thermochemical data with chemical or benchmark accuracy. Thus, even for very small molecules, it is impractical to approach the full configuration interaction (FCI) complete basis-set limit (CBS) via a single-point energy calculation.

To illustrate the limitations of the SPE approach for approximating the exact solution to the nonrelativistic Schrödinger equation, let us consider the bond dissociation energy of the $C_2(^1\Sigma_g^+)$ diatomic. Hereinafter, the regular and augmented correlation-consistent basis sets are denoted by V*n*Z and

Table 2 Overview of formal computational scaling of coupled-cluster methods.

Definition	Computational scaling[a]	Overall[a]
HF and hybrid DFT	$\sim N_{occ}^{2} N_{virt}^{2} N_{iter}$	$\sim N_{bas}^{4}$
MP2 and DHDFT[b]	$\sim N_{occ}^{2} N_{virt}^{3}$	$\sim N_{bas}^{5}$
CCSD	$\sim N_{occ}^{2} N_{virt}^{4} N_{iter}$	$\sim N_{bas}^{6}$
CCSD(T)	$\sim N_{occ}^{2} N_{virt}^{4} N_{iter} + N_{occ}^{3} N_{virt}^{4}$	$\sim N_{bas}^{7}$
CCSDT	$\sim N_{occ}^{3} N_{virt}^{5} N_{iter}$	$\sim N_{bas}^{8}$
CCSDT(Q)	$\sim N_{occ}^{3} N_{virt}^{5} N_{iter} + N_{occ}^{4} N_{virt}^{5}$	$\sim N_{bas}^{9}$
CCSDTQ	$\sim N_{occ}^{4} N_{virt}^{6} N_{iter}$	$\sim N_{bas}^{10}$
CCSDTQ(5)	$\sim N_{occ}^{4} N_{virt}^{6} N_{iter} + N_{occ}^{5} N_{virt}^{6}$	$\sim N_{bas}^{11}$
CCSDTQ5	$\sim N_{occ}^{5} N_{virt}^{7} N_{iter}$	$\sim N_{bas}^{12}$
CCSDTQ5(6)	$\sim N_{occ}^{5} N_{virt}^{7} N_{iter} + N_{occ}^{6} N_{virt}^{7}$	$\sim N_{bas}^{13}$
CCSDTQ56	$\sim N_{occ}^{6} N_{virt}^{8} N_{iter}$	$\sim N_{bas}^{14}$

[a]Where N_{occ} is the number of occupied orbitals, N_{virt} is the number of unoccupied orbitals, N_{iter} is the number of iterations required to reach convergence, and N_{bas} is the number of basis functions.
[b]DHDFT = double-hybrid density functional theory.

AV*n*Z, respectively (where often AV*n*Z also indicates the omission of diffuse functions from hydrogen atoms), and the notation V{X,Y}Z indicates extrapolation from the VXZ and VYZ basis sets. Calculating the C_2 BDE via a single-point energy CCSDTQ/VTZ calculation results in a BDE that is 6.08 kcal mol^{-1} (!!) below the CCSDTQ/CBS BDE.[107] Increasing the basis set size in the CCSDTQ calculation still results in unacceptably large deviations; namely, the CCSDTQ/V*n*Z BDE underestimates the CCSDTQ/CBS BDE by 2.26 (VQZ), 1.11 (V5Z), and 0.65 (V6Z) kcal mol^{-1}. Thus, even the computationally demanding CCSDTQ/V6Z calculation—which required several weeks to run on a computer node with 512 GB of RAM—is unable to achieve benchmark accuracy. Needless to say, the CCSDTQ/V6Z level of theory is only feasible for light diatomic systems. However, even the CCSDTQ/VTZ level of theory is not practical for molecules with more than five non-hydrogen atoms using current mainstream computer hardware. Therefore, regardless of one's computational resources, SPE calculations are not an effective way of approaching the FCI/CBS limit.

Fig. 1A illustrates the way in which single-point energy calculations approach the exact solution to the nonrelativistic Schrödinger equation. In this approach, both the one-particle and *n*-particle spaces are saturated at the same time. The one-particle space describes the size of the basis set used to express the orbitals in the Hartree–Fock wavefunction, and the *n*-particle space describes the level of excitation included in the cluster operator. Attempting to converge both spaces to completeness simultaneously will indeed approach the exact solution. However, due to the exponential computational scaling of coupled-cluster methods (Table 2), this approach is impractical even for very small molecules. For example, the Hartree–Fock calculation scales as N_{bas}^{4}, the CCSD method scales as iterative N_{bas}^{6}, and the CCSDTQ method scales as iterative N_{bas}^{10} with respect to the number of basis functions.

To illustrate how composite ab initio methods work, let us now consider the following reorganization of the CCSDTQ/V*n*Z energy:

$$\begin{aligned}\text{CCSDTQ/V}n\text{Z} = {} & \text{HF/V}n\text{Z} + [\text{CCSD} - \text{HF}]/\text{V}n\text{Z} \\ & + [\text{CCSD(T)} - \text{CCSD}]/\text{V}n\text{Z} \\ & + [\text{CCSDT} - \text{CCSD(T)}]/\text{V}n\text{Z} \\ & + [\text{CCSDT(Q)} - \text{CCSDT}]/\text{V}n\text{Z} \\ & + [\text{CCSDTQ} - \text{CCSDT(Q)}]/\text{V}n\text{Z} \qquad (1)\end{aligned}$$

Eq. (1) is a simple breakdown of the CCSDTQ/V*n*Z energy into the SCF and coupled-cluster correlation terms. However, it illustrates that using a medium-sized basis set (e.g., VTZ) will result in large basis-set truncation

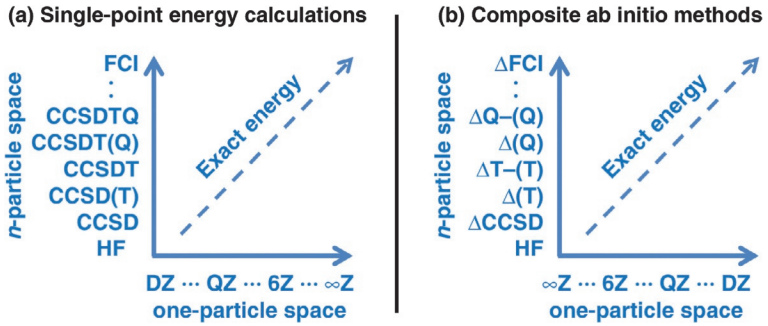

Fig. 1 Modified Pople diagrams illustrating the different relationships between the one- and *n*-particle spaces in (A) single-point energy calculations and (B) composite ab initio methods, in which successively higher cluster expansion terms (ΔCCSD → Δ(T) → ΔT–(T) → Δ(Q) → ΔQ–(Q) → etc.) converge increasingly faster with the basis set size.

errors for the HF and lower-order correlation components. Hence the deviation of 6.08 kcal mol^{-1} is obtained for the $C_2(^1\Sigma_g^+)$ example above. On the other hand, using a large basis set (e.g., V6Z) will make the higher-order correlation components impractical and, as shown above for $C_2(^1\Sigma_g^+)$, still results in a sizable basis set truncation error of 0.65 kcal mol^{-1}. The same arguments apply if we replace the CCSDTQ/V*n*Z energy in Eq. (1) with the FCI/V*n*Z energy.

The underlying premise behind composite ab initio methods is that successively higher cluster expansion terms tend to converge increasingly faster with the basis set since they increasingly reflect nondynamical rather than dynamical correlation *(76)*. Therefore, the computationally more demanding higher-level terms in Eq. (1) can be calculated with smaller and smaller basis sets. This accelerated convergence behavior is the main reason why post-CCSD(T) composite ab initio methods are applicable to much larger systems than high-level SPE calculations (e.g., CCSDTQ/V6Z). This idea is illustrated in Fig. 2B and will be discussed in detail in Section 2.2.

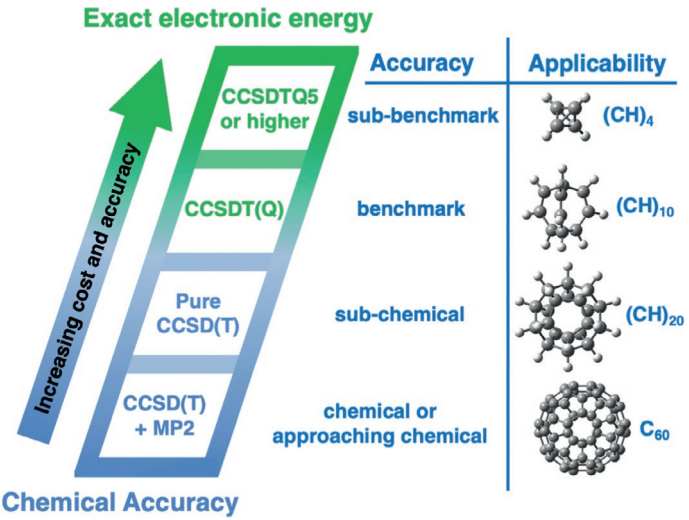

Fig. 2 Jacob's Ladder of composite ab initio methods. Each consecutive rung represents a more rigorous treatment of the one- and/or *n*-particle space, including examples of system sizes to which procedures from each rung can be applied using current mainstream computer hardware.

1.3 The "zoo" of composite ab initio methods

Over the past three decades, there has been a proliferation in the number of developed composite ab initio methods, including the G*n*, CBS, FPA, W*n*, MCCM, HEAT, ccCA, FPD, ATOMIC, INT-MP2-F12, and ChS family of methods, where some of these families include a dozen (or more) of different variants with different applicability and computational cost. Overall, there are over a hundred different composite ab initio methods to choose from, and it has become increasingly difficult to choose the best method for a given chemical system, property, and desired accuracy. When choosing a composite ab initio method, one must consider several key aspects, such as system size, elemental composition, multireference character, electronic state, the chemical property of interest, and the desired accuracy. Additional aspects that may impose additional requirements on the basis sets employed are bond polarity, formal oxidation state, and overall charge. The various composite ab initio methods cover a wide range of accuracies (from chemical to benchmark accuracy), applicability in terms of system size (from five to over 50 non-hydrogen atoms), applicability in terms of elemental composition (from methods applicable only to first-row systems to pan-periodic table methods), and different chemical properties (from atomization energies to properties that depend on energy derivatives such as equilibrium structures and vibrational frequencies). In addition, it should be mentioned that a few composite ab initio methods have been specifically designed for treating excited electronic states and challenging PESs *(109–111)*.

In the context of the present review, it is convenient to classify the composite ab initio methods into four categories:

- Methods that combine second-order Møller–Plesset perturbation theory (MP2) and CCSD(T) calculations
- Methods that rely purely on coupled-cluster calculations up to CCSD(T)
- Methods that rely purely on coupled-cluster calculations up to CCSDT(Q)
- Methods that rely purely on coupled-cluster calculations up to CCSDTQ5 (or higher)

Methods that involve both MP2 and CCSD(T) calculations (hereinafter referred to as hybrid CCSD(T)/MP2 methods) use relatively large basis sets in the MP2 steps and smaller basis sets in the CCSD(T) and higher-order MP*n* calculations. Purely CCSD(T)-based methods use larger basis sets in the CCSD(T) calculations compared to the hybrid CCSD(T)/MP2 methods. Post-CCSD(T) composite approaches may use even larger basis

sets in the CCSD(T) calculations and additionally employ contributions up to the CCSDT(Q) or CCSDTQ5 level. Both the accuracy and computational cost of composite ab initio methods increase in the order:

$$\text{hybrid CCSD(T)/MP2} \rightarrow \text{pure CCSD(T)} \rightarrow \text{pure CCSDT(Q)} \rightarrow \text{pure CCSDTQ5}$$

Fig. 2 depicts a proposed Jacob's Ladder of composite ab initio methods. In this framework, each consecutive rung represents a more rigorous treatment of either the one- or *n*-particle space, along with an increase in the computational cost. The move from rung 1 (hybrid CCSD(T)/MP2 methods) to rung 2 (pure CCSD(T) methods) represents a more rigorous treatment of the one-particle space. The move from rung 2 to rung 3 (CCSDT(Q) methods) represents a more rigorous treatment of the *n*-particle space, which is often accompanied by a more rigorous treatment of the one-particle space at the CCSD(T) level. The move from rung 3 to rung 4 (CCSDTQ5 or higher methods) represents a more rigorous treatment of the *n*-particle space. The methods on the first two rungs are normally capable of chemical accuracy (or approaching chemical accuracy) for TAEs. The methods on the third rung are cable of approaching benchmark accuracy for TAEs. The methods on the fourth rung are cable of sub-benchmark accuracy for TAEs. For example, the Weizmann-4 and HEAT-456QP methods, which are amongst the most accurate methods of the fourth rung, attain RMSDs of 0.072 and 0.100 kcal mol^{-1}. *(76)* respectively, for a set of highly accurate TAEs obtained from the Active Thermochemical Tables thermochemical network *(97,98)*. These RMSDs translate to 95% confidence intervals lower than 1 kJ mol^{-1} *(76)*. The performance of the methods from the middle rung has been evaluated for a large dataset of TAEs relative to CCSDTQ5/CBS data obtained from W4 theory (or higher). For example, for a diverse set of 124 non-multireference TAEs in the W4–11 database, W1U and W1RO theories from the second rung attain RMSDs of 0.57 and 0.65 kcal mol^{-1}, respectively *(80)*. For comparison, methods from the first rung such as G4(MP2) *(112)* and ccCA-PS3 *(55),* attain a MAD of 1.04 and kcal mol^{-1} for the experimental energies of the G3/05 test set. We note that the 95% confidence intervals for the above methods from the first and second rungs of Jacob's Ladder exceed 1 kcal mol^{-1}. Thus, in a more strict sense, these methods do not attain confident chemical accuracy for TAEs (see Section 4 for a more detailed discussion of performance). It is also important to keep in mind that, in a similar manner to Jacob's Ladder of

DFT, Jacob's Ladder of composite ab initio methods is only a general framework for an increase in accuracy and computational cost, and methods from higher rungs are not always guaranteed to be more accurate than methods from a lower rung for any given chemical system and property.

2. Nonrelativistic electronic energy

2.1 Hybrid CCSD(T)/MP2 composite ab initio methods

Hybrid CCSD(T)/MP2 methods are highly successful and cost-effective composite ab initio methods that employ lower-level CCSD(T) calculations in conjunction with large-scale, second-order Møller–Plesset perturbation theory (MP2) calculations. Popular examples of such methods include the G*n*, CBS, and ccCA family of methods. The underlying approximation in many of these methods is the following MP2-based additivity scheme:

$$\begin{aligned}\mathrm{CCSD(T)/Large} \approx\ & \mathrm{CCSD(T)/Small} + \mathrm{MP2/Large} - \mathrm{MP2/Small} \\ & + [\text{additional corrections}]\end{aligned} \tag{2}$$

Here "Small" and "Large" represent different basis set sizes or basis-set extrapolation schemes. This additivity scheme can reduce the computational cost by an order of magnitude relative to the computational cost of pure CCSD(T)-based approaches since it replaces a CCSD/Large or CCSD(T)/Large calculation with an MP2/Large calculation (Table 2). The success of this scheme relies on the similar basis-set convergence behavior of the CCSD(T) and MP2 energies, which was found to be true for thermochemical *(113–115)* and kinetic *(116,117)* properties, as well as weak interactions *(118–123)*.

It is instructive to see what Eq. (2) looks like in two representative composite ab initio approaches, namely G4(MP2) *(112)* and ccCA-PS3 *(55)*. The computationally economical G4(MP2) method employs the Pople-style basis sets for all steps apart from the HF/CBS energy as well as an empirical higher-level correction (HLC) term. The ccCA-PS3 method, on the other hand, employs the correlation-consistent basis sets in conjunction with basis set extrapolations and does not involve an empirical HLC term. G4(MP2) theory uses the following simple and elegant underlying expression for the nonrelativistic electronic energy:

$$\begin{aligned}E[\mathrm{G4(MP2)}] = {} & E[\mathrm{CCSD(T)/6-31G(d)}] \\ & + E[\mathrm{MP2/G3MP2LargeXP}] \\ & - E[\mathrm{MP2/6-31G(d)}] + E[\mathrm{HF/CBS}] \\ & - E[\mathrm{HF/G3MP2LargeXP}] + \mathrm{E(HLC)}\end{aligned} \tag{3}$$

Here, G3MP2LargeXP is an extended version of the Pople 6–311 + G (3df,2p) basis set with additional polarization functions, and HF/CBS indicates extrapolation of the HF energy from truncated versions of the AVTZ and AVQZ basis sets. The E(HLC) term is an empirical "higher level correction" term that depends on the number of paired and unpaired electrons. The HLC empirical parameters are optimized to minimize to mean absolute deviation (MAD) from the experimental determinations in the G3/05 test set *(124)*. Thus, the HLC term compensates for systematic deficiencies in the electronic and nuclear components (and may also include contributions from terms that are not explicitly included in the model, such as core-valence and scalar relativistic corrections). However, it is important to point out that the HLC term cancels out between reactants and products for chemical transformations involving only closed-shell species. Thus, G4(MP2) becomes nonempirical for the calculation of reaction energies of isogyric reactions—that is, reactions in which the number of spin pairs is conserved. The same is true for the calculation of reaction barrier heights in which the number of paired and unpaired electrons is conserved between the transition structure and reactant(s). The bottleneck step in G4(MP2) theory is typically the CCSD(T)/6-31G(d) calculation. Thus, G4(MP2) is one of the computationally most economical composite ab initio methods and can be applied to very large systems, most notably a series of C_{60} and C_{40} isomers *(125–127)*. We also note that some variants of the G4(MP2) procedure (e.g., G4(MP2)-6X) *(13)* use the same energetic components as G4(MP2). Thus, the G4(MP2)-6X energy can be obtained from a G4(MP2) calculation at no additional computational cost.

A different type of hybrid MP2/CCSD(T) composite ab initio method, which does not involve an HLC term, is the ccCA family of methods. For example, the underlying expression for the nonrelativistic electronic energy in the ccCA-PS3 method is:

$$E[\mathrm{ccCA}] = E[\mathrm{HF/CBS}] + E[\mathrm{MP2^{corr}/CBS}] + E[\mathrm{CCSD(T)/VTZ}] - E[\mathrm{MP2/VTZ}] + \Delta E[\mathrm{CV}] \quad (4)$$

Here, the HF and MP2 correlation energies are extrapolated from basis sets of up to AVQZ, and the ΔE[CV] correction is taken as E[MP2(AE)/ACVTZ] − E[MP2(FC)/AVTZ]. The bottleneck step in ccCA-PS3 is typically the CCSD(T)/VTZ calculation. Thus, the ccCA-PS3 method has an intermediate computational cost between W1 theory (with a CCSD/AVQZ bottleneck step) and G4(MP2) (with a CCSD(T)/6-31G (d) bottleneck step).

2.2 Pure coupled-cluster-based composite ab initio methods

Composite ab initio methods based purely on coupled-cluster methods can be divided into two general categories CCSD(T)/CBS methods and post-CCSD(T)/CBS methods. The computationally more economical CCSD(T)/CBS methods can achieve confident chemical accuracy for non-multireference systems and occupy the second rung of our proposed Jacob's Ladder of composite ab initio methods (Fig. 2). These methods are applicable to medium-sized systems such as corannulene ($C_{20}H_{10}$) *(128)*, sumanene ($C_{21}H_{12}$) *(128)*, dodecahedrane ($C_{20}H_{20}$) *(129)*, and carbon clusters (C_{20} and C_{24}) *(130)*. The post-CCSD(T) methods can be divided into CCSDT(Q)/CBS and CCSDTQ5/CBS methods. CCSDT(Q)/CBS methods can normally approach benchmark accuracy and are applicable to systems such as benzene *(36,131)*, and CCSDTQ5/CBS methods can achieve sub-benchmark accuracy and are applicable to smaller systems such as butane and tetrahedrane *(103,106)*.

In purely coupled-cluster-based composite ab initio methods, the coupled-cluster energy is partitioned into the SCF and coupled-cluster correlation components. The notations that are commonly used for the coupled-cluster correlation components are listed in Table 3. The coupled-cluster energies that correspond to rungs 2, 3, and 4 of Jacob's Ladder are given by the following equations:

$$\begin{aligned}\text{Rung 2}: \text{CCSD(T)/CBS} &\approx \text{HF/CBS}_{\text{HF}} + \Delta\text{CCSD/CBS}_{\Delta\text{CCSD}} + \Delta(\text{T})/\text{CBS}_{\Delta(\text{T})} \qquad (5)\\ \text{Rung 3}: \text{CCSDT(Q)/CBS} &\approx \text{HF/CBS}_{\text{HF}} + \Delta\text{CCSD/CBS}_{\Delta\text{CCSD}} + \Delta(\text{T})/\text{CBS}_{\Delta(\text{T})} \\ &\quad + \Delta\text{T}-(\text{T})/\text{CBS}_{\Delta\text{T}-(\text{T})} + \Delta(\text{Q})/\text{CBS}_{\Delta(\text{Q})} \qquad (6)\\ \text{Rung 4}: \text{CCSDTQ5/CBS} &\approx \text{HF/CBS}_{\text{HF}} + \Delta\text{CCSD/CBS}_{\Delta\text{CCSD}} + \Delta(\text{T})/\text{CBS}_{\Delta(\text{T})} \\ &\quad + \Delta\text{T}-(\text{T})/\text{CBS}_{\Delta\text{T}-(\text{T})} + \Delta(\text{Q})/\text{CBS}_{\Delta(\text{Q})} + \Delta\text{Q} \\ &\quad - (\text{Q})/\text{CBS}_{\Delta\text{Q}-(\text{Q})} + \Delta(5)/\text{CBS}_{\Delta(5)} + \Delta 5 \\ &\quad - (5)/\text{CBS}_{\Delta 5-(5)} \qquad (7)\end{aligned}$$

where CBS_{comp} designates a single basis set or basis set extrapolation scheme used for calculating each energetic component depending on the basis set convergence of each component (comp = HF, ΔCCSD, Δ(T), ΔT–(T), etc.). The expansion in Eq. (7) approximates the FCI/CBS energy and can include terms above $\Delta 5\text{–}(5)/\text{CBS}_{\Delta 5-(5)}$ depending on the magnitude of the Δ5–(5) term. This partitioning of the SCF and correlation energy

Table 3 Overview of the coupled-cluster contributions discussed in the present work.

Name	Definition	Abbreviation
Hartree–Fock	HF energy	HF or SCF
Full-iterative connected doubles	CCSD–HF	ΔCCSD
Noniterative connected triples	CCSD(T)–CCSD	Δ(T)
Full-iterative connected triples	CCSDT–CCSD(T)	ΔT–(T)
Noniterative connected quadruples	CCSDT(Q)–CCSDT	Δ(Q)
Full-iterative connected quadruples	CCSDTQ–CCSDT(Q)	ΔQ–(Q)
Connected quadruples as a whole	CCSDTQ–CCSDT	ΔQ
Noniterative quintuples	CCSDTQ(5)–CCSDTQ	Δ(5)
Full-iterative connected quintuples	CCSDTQ5–CCSDTQ(5)	Δ5–(5)
Connected quintuples as a whole	CCSDTQ5–CCSDTQ	Δ5
Noniterative sextuples	CCSDTQ5(6)–CCSDTQ5	Δ(6)
Full-iterative connected sextuples	CCSDTQ56–CCSDTQ5(6)	Δ6–(6)
Connected sextuples as a whole	CCSDTQ56–CCSDTQ5	Δ6
Post-CCSD(T) as a whole	CC*n*–CCSD(T)[a]	Post-CCSD(T)

[a]CC*n* = any post-CCSDT method, e.g., CCSDT(Q), CCSDTQ, CCSDTQ(5), etc.

components is a highly efficient and effective approach for obtaining the CCSD(T)/CBS, CCSDT(Q)/CBS, or FCI/CBS energies.

A few general design features that are important to the success of coupled-cluster-based composite ab initio methods are:

- All the energetic components (HF, ΔCCSD, Δ(T), ΔT–(T), Δ(Q), etc.) are converged separately to a common accuracy level, i.e., the basis set truncation error associated with each of the terms should be roughly the same: $\varepsilon(\mathrm{CBS_{HF}}) \approx \varepsilon(\mathrm{CBS_{\Delta CCSD}}) \approx \varepsilon(\mathrm{CBS_{\Delta(T)}}) \approx \varepsilon(\mathrm{CBS_{\Delta T-(T)}}) \approx \varepsilon(\mathrm{CBS_{\Delta(Q)}}) \approx \cdots$
- Smaller basis sets are needed for calculating higher-order correlation components since they increasingly reflect nondynamical rather than dynamical correlation.
- Each of the energetic components (HF, ΔCCSD, Δ(T), ΔT–(T), Δ(Q), etc.) may exhibit a different basis set convergence behavior and is converged to the CBS limit in the most effective manner (e.g., using

an optimal extrapolation exponent or a scaling factor). The effective extrapolation exponents or scaling factors can be physically or empirically motivated.

- The HF and lower-level correlation components (e.g., ΔCCSD, Δ(T), and ΔT–(T)) are typically extrapolated to the CBS limit, whilst the higher-order correlation components (e.g., Δ(Q) and above) are typically calculated with a single basis set of double-ζ or triple-ζ quality.
- In many contemporary composite ab initio methods, the basis set convergence of the HF, ΔCCSD, and Δ(T) components is accelerated by explicitly correlated, density fitting, and local coupled-cluster techniques keeping computational cost at a minimum.

The above factors contribute to the computational efficiency and accuracy of contemporary composite ab initio methods. Of particular importance to the success of these methods is that contributions from successively higher cluster expansion terms (i.e., ΔCCSD → Δ(T) → ΔT–(T) → Δ(Q) → ΔQ–(Q) → Δ(5) → etc.) tend to converge increasingly faster to the complete basis set limit, since they increasingly reflect static rather than dynamical correlation. This relationship between the *n*- and one-particle spaces is illustrated schematically in Fig. 1B and is the main reason that post-CCSD(T) composite ab initio methods are applicable to molecules with nearly ten non-hydrogen atoms at a realistic computational cost (e.g., C_2Cl_6, SF_6, and C_6H_6) *(76,80,103,108)*.

Table 4 summarizes the basis sets used for extrapolating or calculating the various components in two representative coupled-cluster-based composite ab initio methods W*n* *(33,34,36,76,80,87,90)* and HEAT *(46–50)*. As can be seen, the basis sets used in the W*n* and HEAT methods vary in a systematic manner across the rungs of Jacob's Ladder (rows in Table 4) and correlation components (columns in Table 4). For example, in the original W*n* methods, the HF and ΔCCSD components are extrapolated from the following basis sets AV{T,Q}Z (W1 theory, rung 2), AV{Q,5}Z (W3 theory, rung 3), and AV{5,6}Z (W4 theory, rung 4). As discussed above, smaller basis sets are used for calculating successively higher-order correlation components. For example, in W4 theory, the following basis sets (or basis set extrapolations) are used AV{5,6}Z (ΔCCSD), AV{Q,5}Z (Δ(T)), V{D,T}Z (ΔT–(T)), VTZ (Δ(Q)), VDZ (ΔQ–(Q)), and VDZ(sp) (Δ5) (where VDZ(sp) is a truncated version of the VDZ basis set) *(33)*. We note that the largest basis set used for extrapolating the HF and ΔCCSD components is the same in all cases. From a computational cost perspective, it would make little sense to extrapolate the HF component

Table 4 Overview of the basis sets used for extrapolating or calculating the HF and correlation components in representative variants of the W*n* and HEAT composite ab initio methods.

Name	CC*n*/CBS[a]	HF	ΔCCSD	Δ(T)	ΔT–(T)	Δ(Q)	ΔQ–(Q)	Δ(5) or Δ5	Δ(6)
W1	CCSD(T)	AV{T,Q}Z	AV{T,Q}Z	AV{D,T}Z					
W2	CCSD(T)	AV{Q,5}Z	AV{Q,5}Z	AV{T,Q}Z					
W3	CCSDT(Q)	AV{Q,5}Z	AV{Q,5}Z	AV{T,Q}Z	V{D,T}Z	VDZ			
W4lite	CCSDT(Q)	AV{5,6}Z	AV{5,6}Z	AV{Q,5}Z	V{D,T}Z	VDZ			
W4	CCSDTQ5	AV{5,6}Z	AV{5,6}Z	AV{Q,5}Z	V{D,T}Z	VTZ	VDZ	VDZ(sp)	
W4.3	CCSDTQ56	AV{5,6}Z	AV{5,6}Z	AV{Q,5}Z	V{T,Q}Z	V{T,Q}Z	VTZ	VDZ	VDZ(sp)
W1–F12	CCSD(T)	V{D,T}Z-F12	V{D,T}Z-F12	AV{D,T}Z					
W2–F12	CCSD(T)	V{T,Q}Z-F12	V{T,Q}Z-F12	VTZ-F12					
W3–F12	CCSDT(Q)	V{T,Q}Z-F12	V{T,Q}Z-F12	VTZ-F12	V{D,T}Z	VDZ			
W4–F12	CCSDTQ5	V5Z-F12	V{Q,5}Z-F12	AV{Q,5}Z	V{D,T}Z	VTZ	VDZ	VDZ(sp)	
HEAT-345(Q)	CCSDT(Q)	ACV{T,Q,5}Z	ACV{Q,5}Z	ACV{Q,5}Z	V{T,Q}Z	VDZ			
HEAT-456QP	CCSDTQ5	ACV{Q,5,6}Z	ACV{5,6}Z	ACV{5,6}Z	V{T,Q}Z	VDZ	VDZ	VDZ	
diet-HEAT	CCSDT(Q)	AV{T,Q,5}Z	AV{Q,5}Z	AV{Q,5}Z	VTZ	VDZ			
diet-HEAT-F12	CCSDT(Q)	VQZ-F12	V{T,Q}Z-F12	V{T,Q}Z-F12	VTZ(spd)	VDZ			

[a]CC*n* = Coupled-cluster excitation level being approximated.

from smaller basis sets than the ΔCCSD component, however in some cases (e.g., W4–F12 and diet-HEAT-F12) there is no need to extrapolate the HF component, and it is simply calculated using the larger basis set.

A notable difference between the W*n* and HEAT theories is that in the W*n* methods the Δ(T) and the ΔCCSD correlation components are extrapolated separately to the complete basis set limit, whereas in the HEAT methods the ΔCCSD(T) component is extrapolated to the CBS limit as a whole. In the W*n* methods, the Δ(T) component is typically extrapolated from basis sets with one cardinal number smaller than those used for extrapolating the ΔCCSD component. This separation makes the CCSD(T)/CBS W*n* methods computationally more economical. For example, W1 and W1–F12 theories have been applied to systems as large as arginine ($C_6H_{14}N_4O_2$) *(132),* terphenyl ($C_{18}H_{14}$) *(133),* corannulene ($C_{20}H_{10}$) *(128),* dodecahedrane ($C_{20}H_{20}$) *(129),* and sumanene ($C_{21}H_{12}$) *(125)*. Following the same trend of the HF, ΔCCSD, and Δ(T) components, the largest basis set used for extrapolating or calculating the ΔT–(T) component is smaller by 1–2 cardinal numbers than the largest basis set used for extrapolating the Δ(T) component. In all the CCSDT(Q)/CBS methods in Table 4, the Δ(Q) component is calculated with the VDZ basis set. In the CCSDTQ5/CBS methods in Table 4, the ΔQ–(Q) component is calculated with the VDZ basis set (except for W4.3 theory). We note that the principal bottleneck in applying post-CCSD(T)/CBS methods to large systems is typically the evaluation of the ΔT–(T) term in CCSDT(Q)/CBS methods and the ΔQ–(Q) term in CCSDTQ5/CBS methods.

Another important distinction between the W*n* and HEAT methods is the treatment of the core electrons in the CCSD(T) calculations. In the HEAT methods, all electrons are correlated in the CCSD(T) calculations. This means that all-electron CCSD(T)/ACV5Z (HEAT-345(Q)) and CCSD(T)/ACV6Z (HEAT-456QP) calculations are performed. Importantly, this eliminates the error associated with the CCSD(T)-based core-valence correction; however, it significantly increases the computational cost of the CCSD(T) calculations, in particular for molecules containing second-row atoms. In the W*n* methods, on the other hand, the ΔCCSD and Δ(T) terms are obtained within the frozen-core approximation. That is, the inner-shell orbitals (1*s* for first-row atoms, and 1*s*, 2*s*, and 2*p* for second-row atoms) are constrained to be doubly occupied in all configurations. In practice, this partitioning between the inner- and valence-shell electrons makes W*n* theories applicable to molecules with multiple second-row atoms at a realistic computational cost. For example,

W4lite theory has been applied to molecules such as S_8 *(134)*, and W1–F12 theory has been applied to systems as large as P_4S_{10} *(135)*. To account for inner-shell correlation, the W*n* theories include a core-valence (CV) correction term. The CV correction is obtained at the CCSD(T)/APWCV{T,Q}Z level in W2, W3, and W4 theories. It has been found that this level of theory provides an excellent balance between accuracy and computational cost with a root-mean-square deviation (RMSD) of merely 0.03 kcal mol^{-1} for the diverse set of 200 small first- and second-row molecules in the W4–17 database *(103,136)*. In high-level theories such as W4, this RMSD is comparable to the errors in the valence parts and to post-CCSD(T) contributions to the CV component. For example, the ΔT–(T) correction to the CV component increases the atomization energies of HO_3 and O_3 by 0.02 and 0.03 kcal mol^{-1}, respectively *(33)*.

2.3 Correlation contributions beyond the CCSD(T) level

Post-CCSD(T) correlation contributions are of key importance for achieving chemical accuracy for multireference systems and for achieving sub-benchmark accuracy for non-multireference systems. It is therefore useful to gain an understanding of the magnitude of the post-CCSD(T) correlation contributions relative to the CCSD(T) energy. For this purpose, we consider the W4–17 database of highly accurate TAEs of 200 organic and inorganic species *(36)*. Most of the TAEs in the W4–17 database have been obtained at the CCSDTQ5/CBS level of theory from W4 theory. For a set of 46 small molecules (e.g., acetylene, CH_4, and ClO), the TAEs have been obtained at the CCSDTQ56/CBS level of theory from W4.3 and W4.4 theories. Whereas for a subset of 33 larger molecules (e.g., benzene, SF_6, and C_2Cl_6) the TAEs have been obtained at the CCSDT(Q)/CBS level of theory from W4lite theory. Overall, the W4–17 dataset includes molecules with up to eight non-hydrogen atoms, which cover a broad spectrum of bonding situations, electronic states, and multireference character.

Table 5 gives an overview of the spread of the electronic SCF, ΔCCSD, Δ(T), ΔT–(T), ΔQ, Δ5, and Δ6, contributions for the W4–17 database in terms of the mean, standard deviation, and min/max values. The largest contribution to the TAEs is normally obtained at the SCF level. The mean SCF contribution to the TAEs is 270.6 kcal mol^{-1} with a standard deviation of 235.5 kcal mol^{-1}. However, for the larger molecules, the SCF contribution exceeds well over 1000 kcal mol^{-1}, with the largest contribution of 1239.8 kcal mol^{-1} for pentane. It is interesting to note that highly

Table 5 Overview of the magnitude (in kcal mol^{-1}) and signs of the electronic energetic contributions from W4.*x* theory for the set of 200 TAEs in the W4–17 database.

Component	Mean	SD[a]	Largest
SCF	270.63	235.46	1239.8 (C_5H_{12})
ΔCCSD	123.73	62.12	334.4 (C_5H_{12})
Δ(T)	11.92	6.99	42.8 (N_2O_4)
ΔT–(T)	−0.85	0.68	−3.3 (P_4)
ΔQ	0.97	0.77	4.8 (S_4)
Δ5	0.06	0.07	0.4 (O_3)
Δ6	0.01	0.01	0.1 (C_2)

[a]Mean = mean value; SD = standard deviation; Largest = largest positive or negative value.

multireference systems such as BN, O_3, FO, F_2, F_2O, FO_2, F_2O_2, ClO_2, ClO_3, ClF_3, and ClF_5 are unbound at the SCF level, i.e., the SCF contribution to the TAE is negative. Single and double excitations from the Hartree–Fock configuration constitute the largest contribution to the correlation energy. The mean ΔCCSD contribution for the W4–17 set is 123.7 kcal mol^{-1} with a standard deviation of 62.1 kcal mol^{-1}. The Δ(T) contribution to the TAEs is still large, with a mean and standard deviation of 11.9 and 7.0 kcal mol^{-1}, respectively. For over 50% of the species in the W4–17 database, the Δ(T) contribution exceeds 10.0 kcal mol^{-1}, and the maximum Δ(T) contribution reaches 42.8 kcal mol^{-1} for N_2O_4. These statistical measures illustrate why all the composite ab initio methods on Jacob's Ladder (Fig. 2) must explicitly include the Δ(T) contribution.

It is well established that the higher-order triples contributions (ΔT–(T)) tend to reduce the TAEs. Indeed, 94% of the ΔT–(T) contributions are negative and the positive contributions are close to zero, i.e., they are smaller than +0.3 kcal mol^{-1}. The mean and standard deviation of the ΔT–(T) contribution are −0.85 and 0.68 kcal mol^{-1}, respectively. However, this contribution can reach a maximum negative value of −3.3 kcal mol^{-1} for P_4. The quadruple excitations, on the other hand, universally increase the TAEs (i.e., all the ΔQ contributions are positive). The mean value of the ΔQ contributions is 0.97 kcal mol^{-1} with a standard deviation of 0.77 kcal mol^{-1}. Thus, the positive ΔQ contributions have a

similar magnitude to the (mostly) negative ΔT–(T) contributions. For this reason, the complete neglect of contributions beyond the CCSD(T) level is one of the most successful yet computationally economical approaches in quantum chemistry, and the CCSD(T) method is rightly referred to as the gold standard of quantum chemistry. Nevertheless, as we will see in the following paragraph, this gold standard is only applicable in the accuracy range of ±1 kcal mol^{-1}. In the context of sub-kcal/mol accuracies, the CCSDT(Q) (or CCSDTQ) method is more appropriately referred to as the gold standard.

Inspection of the Δ5 contributions to the TAEs in the W4–17 database reveals that these contributions are typically smaller than 0.1 kcal mol^{-1} for all but highly multireference systems. For multireference systems, however, the Δ5 contributions are still significant at the sub-kcal/mol level. For example, they range between 0.2 and 0.4 kcal mol^{-1} for systems like SO_3, ClO_2, FO_2, NCCN, F_2O_2, S_4, C_2, and O_3. The Δ6 contributions to the TAEs are very small and certainly negligible for all but highly multireference systems. The largest Δ6 contributions are 0.04 and 0.06 kcal mol^{-1} for BN and C_2, respectively *(137,138)*. The Δ7 contributions to the TAEs are practically nil being 0.003 kcal mol^{-1} for C_2 *(137)*.

The W4–17 database includes 200 CCSDT(Q), CCSDTQ5, and CCSDTQ56 TAEs for molecules with up to eight non-hydrogen atoms, which cover a broad spectrum of bonding situations, electronic states, and multireference character. As such it is an excellent resource for quantitative evaluation of the accuracy that can be expected from CCSD(T)-based methods on the first two rungs of Jacob's Ladder. Fig. 3 gives an overview of the post-CCSD(T) contributions to the 200 TAEs in the W4–17 database. Post-CCSD(T) contributions to the TAEs tend to be evenly distributed between positive and negative values, albeit there are slightly more positive than negative values. Overall, 58% of the post-CCSD(T) contributions to the TAEs are positive and 42% are negative. Importantly, the highly multireference systems all have large and positive post-CCSD(T) contributions ranging between 1.0 and 3.5 kcal mol^{-1}. For these systems, the positive ΔQ contributions are significantly larger than the negative ΔT–(T) contributions. In particular, for eight systems, the overall post-CCSD(T) contributions range between 1 and 2 kcal mol^{-1}, namely 1.1 (ClF_5, NO_2), 1.2 (S_3), 1.3 (N_2O_4), 1.4 (B_2, ClNO), and 1.8 (F_2O_2, cis-HO_3) kcal mol^{-1}. For five systems, the post-CCSD(T) contributions to the TAEs range between 2.0 and 3.5 kcal mol^{-1}, namely 2.3 (trans-HO_3), 2.4 (S_4), 2.9 (O_3), 3.0 (FO_2), and 3.5 (ClO_2) kcal mol^{-1}.

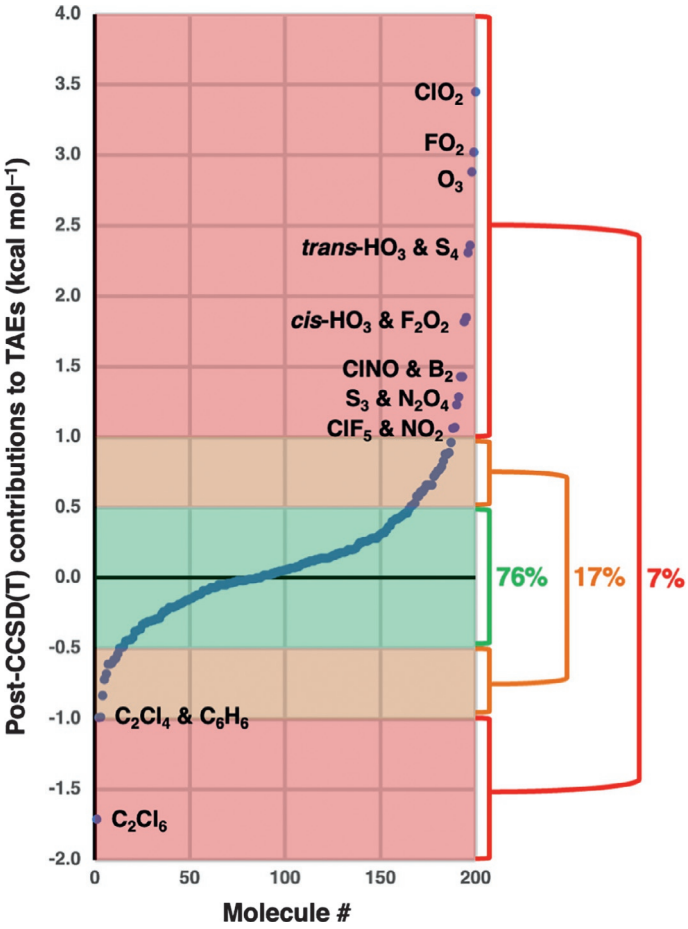

Fig. 3 Overview of post-CCSD(T) contributions to the 200 total atomization energies in the W4–17 database (in kcal mol^{-1}).

The post-CCSD(T) contributions for the lion's share of the TAEs (93%) are confined between ±1.0 kcal mol^{-1}. Furthermore, for 76% of the TAEs, the post-CCSD(T) contributions are confined between ±0.5 kcal mol^{-1}. Thus, it is fair to say that the CCSD(T)/CBS level of theory is indeed, on average, a "gold standard" for systems dominated by mild-to-moderate multireference effects. Let us examine more closely the systems with relatively large post-CCSD(T) contributions ranging between −0.5 and −1.0 and between +0.5 and +1.0 kcal mol^{-1}. This set consists of 17% of the systems in the W4–17 database. The subset with positive post-CCSD(T) contributions ranging between +0.5 and +1.0 kcal mol^{-1}

includes molecules that are characterized by moderate-to-severe multireference effects (e.g., BN, CN, NCCN, O_2, HO_2, F_2, N_2O, F_2O, P_2, ClO, Cl_2O_2, and ClO_3). On the other hand, the subset of molecules with negative post-CCSD(T) contributions ranging between −0.5 and −1.0 kcal mol^{-1} includes mostly hydrocarbons that are not normally associated with strong multireference effects (e.g., cyclobutene, *n*-pentane, furan, 1,3-dithiotane, tetrahedrane, cyclopentadiene, pyrrole, and thiophene). We note that tetrachloroethylene and benzene both have large negative post-CCSD(T) contributions of −0.99 kcal mol^{-1} and that hexachloroethane has a post-CCSD(T) contribution of −1.7 kcal mol^{-1}. It has been previously noted that medium-sized hydrocarbons are associated with post-CCSD(T) contributions that can exceed half a kcal mol^{-1} even though they are clearly dominated by a single reference configuration *(84,139)*. Examining the ΔT–(T) and ΔQ contributions for the above medium-sized hydrocarbons, we find that the negative higher-order triples (ΔT–(T)) contributions are larger than the positive connected quadruple contributions. This is in contrast to systems that are dominated by multireference effects, for which the positive ΔQ contribution outweighs the negative ΔT–(T) contribution. There are some indications that this effect correlates with the size of the system. For example, for the series of saturated *n*-alkanes post-CCSD(T) contributions increase linearly with the size of the system, namely they amount to −0.13 (ethane), −0.28 (propane), −0.40 (*n*-butane), −0.54 (*n*-pentane), and −0.65 (*n*-hexane) kcal mol^{-1} *(139)*. We note that the squared correlation coefficient between the post-CCSD(T) contributions and the number of carbons in these alkanes is $R^2 = 0.9991$.

The above discussion illustrates that multireference effects are not the only factor affecting the magnitude of post-CCSD(T) contributions and that the sign of the overall post-CCSD(T) contribution to the TAEs may indicate whether multireference or size effects are dominant. Systems with relatively large positive post-CCSD(T) contributions (+0.5 to +1.0 kcal mol^{-1}) are indicative of multireference character, whereas systems with relatively large negative post-CCSD(T) contributions (−0.5 to −1.0 kcal mol^{-1}) may be indicative of size effects.

3. Secondary energetic corrections

3.1 The 3D Pople diagram of composite ab initio methods

Fig. 1 shows how the one-particle and *n*-particle spaces converge to the exact nonrelativistic electronic energy. However, energies and chemical

properties calculated on the nonrelativistic electronic potential energy surface are not directly comparable to those obtained from experiments. To reproduce accurate experimental energetic and spectroscopic properties, secondary energetic contributions must be considered. 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, conformational corrections to the enthalpy for floppy molecules *(132,140)*, or tunneling contributions for reaction barrier heights involving hydrogen transfer (or heavy-atom transfers at low temperatures) *(141)*. Fig. 4 gives a complete overview of the components involved in composite ab initio methods. The front face of the three-dimensional Pople diagram (red arrows) represents the two-dimensional convergence of the one-particle and *n*-particle spaces, whereas the third dimension (green arrow) represents any secondary energetic contributions that are needed for a meaningful comparison with experimentally observable energetic and spectroscopic quantities. In principle, any secondary energetic 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.

The nonrelativistic electronic energy typically accounts for about 95% of the relativistic, all-electron, ZPVE-inclusive TAE. For a detailed

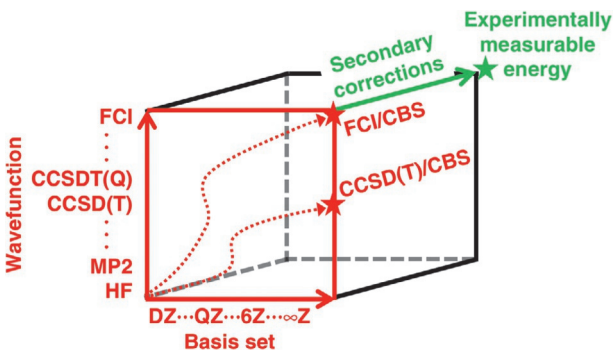

Fig. 4 A three-dimensional Pople diagram illustrating the components comprising composite ab initio methods. The red axes represent the two-dimensional convergence of the nonrelativistic electronic energy. The green axis represents any additional energetic contributions (e.g., scalar relativistic, spin-orbit, zero-point vibrational energy, and Born–Oppenheimer corrections) that are needed for a meaningful comparison with experimentally observable quantities.

discussion of the various secondary energetic contributions, see Refs. *(62,67,76,81,86–89,92)*. Here, we will focus on the relative magnitudes of the various secondary energetic contributions in the highly diverse W4–17 database. This benchmark dataset includes 200 all-electron, relativistic, ZPVE-inclusive, and DBOC-inclusive TAEs obtained mostly at the CCSDTQ5/CBS level from W4 theory. The W4–17 dataset includes first- and second-row molecules with up to eight non-hydrogen atoms and covers a broad spectrum of bonding situations and electronic states *(36)*. Table 6 gives an overview of the magnitude and spread of the secondary energetic contributions for the entire W4–17 database in terms of the mean, standard deviation, and maximum values.

3.2 The core-valence correction

We begin by noting that some lower-level composite ab initio methods (e.g., G4(MP2)) do not include a CV term and rely on fortuitous error cancellation *(112)*, whilst in some high-level methods (e.g., HEAT), a CV term is not needed since the ΔCCSD and Δ(T) components are calculated with all-electrons correlated. Yet, most composite ab initio methods (e.g., CBS-APNO, G4, ccCA, W*n*, W*n*-F12, and FPD) obtain the CCSD(T)/CBS energy with only the valence electrons correlated (CCSD(T)$_{val}$) and include a core-valence correction ΔCV to approximate the all-electron CCSD(T) energy (CCSD(T)$_{all}$):

$$\mathrm{CCSD(T)_{all}/CBS} \approx \mathrm{CCSD(T)_{val}/CBS} + \Delta\mathrm{CV} \tag{8}$$

Table 6 Overview of the magnitude (in kcal mol^{-1}) and signs of the electronic and secondary energetic contributions from W4 theory for a set of 200 TAEs of first- and second-row molecules in the W4–17 database.

Component	Mean	SD[a]	Largest
Core-valence	1.56	1.48	7.4 (C_6H_6)
Scalar relativ.	−0.55	0.44	−3.2 (SF_6)
Spin-orbit	−0.74	0.78	−5.2 (C_2Cl_6)
DBOC	0.06	0.07	0.3 (C_5H_{12})
ZPVE	17.61	16.97	99.5 (C_5H_{12})

[a]Mean = mean value; SD = standard deviation; Largest = largest positive or negative value.

This approach significantly reduces the computational cost of the demanding CCSD(T)/CBS calculations and is essential for composite ab initio methods that are applicable to molecules containing several second-row (or heavier) elements. In most composite ab initio methods that belong to the first rung of Jacob's Ladder (e.g., G3, G4, ccCA-PS3, and W1X-*n*) the ΔCV term is calculated at the MP2 level ($\Delta CV = MP2_{all} - MP2_{val}$). In methods that belong to the second rung, the ΔCV term is usually calculated at the CCSD(T) level ($\Delta CV = CCSD(T)_{all} - CCSD(T)_{val}$). For example, in the original W1 and W2 theories, the ΔCV term is calculated at the CCSD(T)/MTsmall level of theory, where MTsmall is a completely decontracted VTZ basis set with additional tight 2d1f functions *(31)*. In methods that belong to the third and fourth rungs, the CV term is usually obtained at the CCSD(T)/CBS level of theory ($\Delta CV = CCSD(T)_{all}/CBS - CCSD(T)_{val}/CBS$). For example, in W4lite and W4 theories, the CV correction is extrapolated from the APWCV{T,Q}Z basis set pair *(33)*. Finally, in some high-level methods on the fourth rung (e.g., W4.2 and W4.3), the CV correction is obtained at the CCSDT level *(33)*, and in W4.4, it is obtained at the CCSDT(Q) level *(34)*. For a detailed discussion of the basis set and method dependencies of the core-valence (and core-core) contributions to TAEs see Ref. *(139)*.

The core-valence corrections in the W4–17 database have been calculated at the CCSD(T)/APWCV{T,Q}Z level of theory (with post-CCSD(T) contributions included for some of the smaller systems). Except for a small number of highly polar systems (e.g., ClF_5, SF_6, $AlCl_3$) for which the CV correction is repulsive, the CV term is attractive for nearly all species. The largest CV corrections in the W4–17 database are obtained for medium-sized hydrocarbon/heteroatom first-row molecules, namely 4.4 (cyclobutadiene), 4.6 (cyclobutene and cyclobutene), 4.8 (*n*-butane and *trans*-butadiene), 5.1 (tetrahedrane and furan), 5.6 (pyrrole), 5.8 (borole), 6.0 (*n*-pentane and cyclopentadiene), and 7.4 (benzene) kcal mol^{-1}. Thus, this term clearly cannot be neglected and has to be treated at a sufficiently high level of theory for quantitative chemical accuracy *(36,136,142)*.

It is instructive to examine how the CV correction varies with molecular size for a systematic series of hydrocarbons of increasing size. Fig. 5 gives an overview of the CV contributions for two such series (i) small straight-chain alkanes with up to five carbon atoms and (ii) $(CH)_n$ hydrocarbon cages with up to 20 carbon atoms. The CV corrections for the straight-chain alkanes are taken from W4 theory *(103)*, and those for the $(CH)_n$ hydrocarbon cages are taken from W1–F12 theory *(129)*. For the homologous series

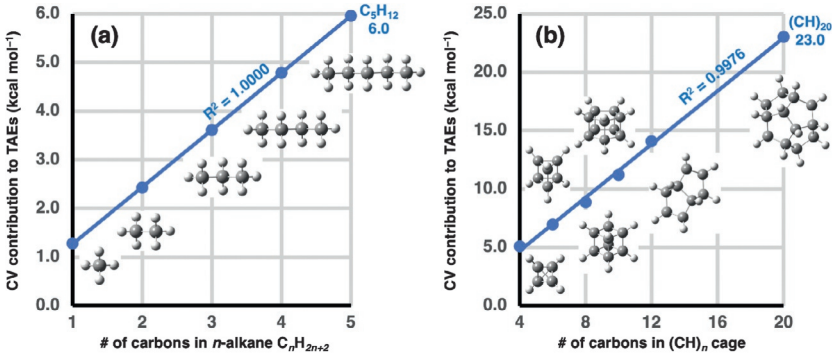

Fig. 5 Overview of core-valence contributions to TAEs (in kcal mol^{-1}) for a series of (A) linear alkanes (methane, ethane, propane, *n*-butane, *n*-pentane) from W4 theory and (B) platonic/prismatic $(CH)_n$ hydrocarbon cages: tetrahedrane $(CH)_4$, triprismane $(CH)_6$, cubane $(CH)_8$, pentaprismane $(CH)_{10}$, octahedrane $(CH)_{12}$, and dodecahedrane $(CH)_{20}$ from W1–F12 theory.

of straight-chain alkanes (methane, ethane, propane, *n*-butane, and *n*-pentane), there is a perfect linear correlation between the number of carbons in the alkane and the magnitude of the CV correction with a squared correlation coefficient of $R^2 = 1.0000$ (Fig. 5A). For methane, the CV correction already exceeds 1 kcal mol^{-1}, i.e., it is 1.3 kcal mol^{-1} at the CCSD(T)/APWCV{T,Q}Z level of theory, and each additional CH_2 group increases the CV contribution by ~1.2 kcal mol^{-1} up to a contribution of 6.0 kcal mol^{-1} for *n*-pentane. The nearly constant increase in the magnitude of the CV correction with the number of CH_2 groups in linear alkanes has been previously noted by Dixon and co-workers, where a slightly lower increase of ~1.1 kcal mol^{-1} per CH_2 group was obtained at the CCSD(T)/PWCVTZ level of theory *(143)*.

For the series $(CH)_n$ hydrocarbon cages (Fig. 5B), we still obtain an almost perfect linear correlation ($R^2 = 0.9976$) between the magnitude of the CV correction and the number of carbons in the $(CH)_n$ cage. However, it is important to note that this is not a true homologous series. Namely, this series is composed of three platonic hydrocarbons (tetrahedrane $(CH)_4$, cubane $(CH)_8$, and dodecahedrane $(CH)_{20}$), two prismatic hydrocarbons (triprismane $(CH)_6$ and pentaprismane $(CH)_{10}$), and one truncated tetrahedrane (octahedrane $(CH)_{12}$). We note that the W1–F12 CV correction for tetrahedrane (5.09 kcal mol^{-1}) *(129)* is identical to that obtained at the W4 level (5.08 kcal mol^{-1}) *(36)*. The CV correction for triprismane (7.0 kcal mol^{-1}) is on the same order of magnitude as that

for the benzene isomer (7.4 kcal mol^{-1}). Similarly to the case of the linear alkanes, with each addition of a CH unit, the CV correction increases by roughly 1.2 kcal mol^{-1}, up to a contribution of 23.0 kcal mol^{-1} for dodecahedrane $(CH)_{20}$.

3.3 Relativistic corrections

Table 6 gives an overview the magnitude of the scalar relativistic effects in the W4–17 database. All the scalar relativistic corrections in the W4–17 database are calculated using the second-order Douglas–Kroll–Hess (DKH) approximation *(144,145)*, which has been shown to yield results in close agreement with the full relativistic treatment for first- and second-row systems *(146,147)*. The scalar relativistic contributions in the W4–17 database are calculated at the CCSD(T)/AVQZ-DK level of theory. The scalar relativistic corrections to the TAEs are universally repulsive. Relatively small contributions below 1 kJ mol^{-1} are obtained for diatomic molecules (e.g., C_2, N_2, O_2, F_2, Cl_2, CO, NO, FO, and ClO) and small hydrides (e.g., BH, BH_3, B_2H_6, CH, CH_2, CH_3, CH_4, NH, NH_2, and NH_3), but also for oxygen-rich species like ozone and halogen oxides (e.g., F_2O, Cl_2O, FO_2, ClO_2, and F_2O_2). The largest contributions in the W4–17 database are obtained for polyhalogenated compounds with a central second-row atom, for example −3.2 (SF_6), −2.7 ($HClO_4$), −2.6 (PF_5), −1.9 (SO_3, SiF_4), −1.6 (ClO_3), −1.3 ($HClO_3$, AlF_3, $AlCl_3$), and −1.0 (PF_3) kcal mol^{-1}; as well as for fluorocarbon and chlorocarbon compounds −1.3 (C_2F_6) and −1.1 (C_2Cl_6, C_2Cl_4, C_2F_4, cis/trans-$C_2F_2Cl_2$) kcal mol^{-1}. Fig. 6 gives an overview of the scalar relativistic correction for the series of *n*-alkanes (up to *n*-pentane) and hydrocarbon cages (up to dodecahedrane). As is the case for the CV correction (Fig. 5), for both series, there is a near-perfect linear correlation between the scalar relativistic correction and the number of carbons in the system. The squared correlation coefficient is 0.9999 for the alkanes and 0.9995 for the hydrocarbon cages (Fig. 6). For methane, the alkanes relativistic corrections range between −0.2 (methane) and −1.0 (*n*-pentane) kcal mol^{-1}, whereas for the $(CH)_n$ cages, they range between −0.8 (tetrahedrane) and −3.7 (dodecahedrane) kcal mol^{-1}. Again, it should be noted that the nearly constant increase in the magnitude of the relativistic correction with the number of CH_2 groups in linear alkanes has been previously noted by Dixon and co-workers *(143)*.

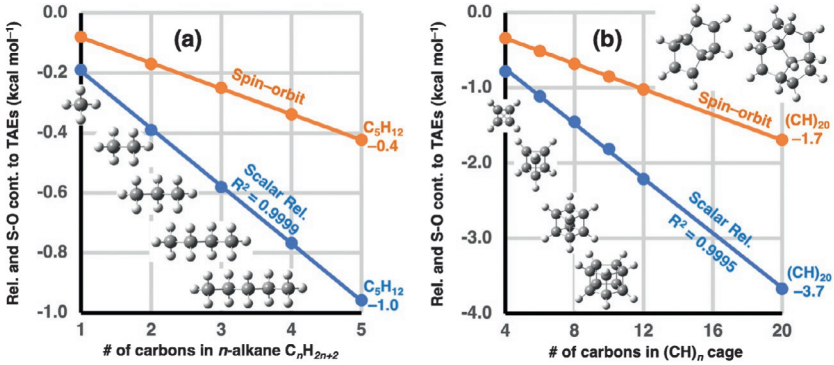

Fig. 6 Overview of scalar relativistic and first-order spin-orbit contributions to TAEs (in kcal mol^{-1}) for a series of (A) linear alkanes (methane, ethane, propane, *n*-butane, *n*-pentane) from W4 theory; and (B) platonic/prismatic $(CH)_n$ hydrocarbon cages (tetrahedrane $(CH)_4$, triprismane $(CH)_6$, cubane $(CH)_8$, pentaprismane $(CH)_{10}$, octahedrane $(CH)_{12}$, and dodecahedrane $(CH)_{20}$) from W1-F12 theory.

With the exception of composite ab initio methods developed specifically for treating heavy main-group, transition-metal, and f-block systems, most methods consider only first-order atomic and molecular spin-orbit corrections. These corrections are nonzero for radicals in a degenerate ground state and can make nontrivial contributions to the molecular binding energies (Table 6). The atomic spin-orbit correction for first- and second-row elements amount to 0.029 (B), 0.085 (C), 0.223 (O), 0.385 (F), 0.214 (Al), 0.428 (Si), 0.560 (S), and 0.841 (Cl) kcal mol^{-1} *(33)*. Thus, the largest spin-orbit corrections in the W4–17 database are obtained for compounds with multiple oxygen, fluorine, and second-row atoms. Prominent examples with spin-orbit correction in excess of 1 kcal mol^{-1} include C_2Cl_6 (−5.2), C_2Cl_4 (−3.5), SF_6 (−2.9), ClF_5 (−2.8), $AlCl_3$ (−2.7), C_2F_6 (−2.5), S_4 (−2.2), Cl_2O_2 (−2.1), SiF_4 (−2.0), Cl_2 (−1.7), CF_4 (−1.6), ClO_3 (−1.5), SO_3 (−1.2), and S_2 (−1.1 kcal mol^{-1}). For closed-shell hydrocarbons, the first-order spin-orbit correction to the TAE increases more slowly with the molecular size; however, it clearly cannot be neglected at the chemical accuracy level. For example, it amounts to just over 0.5 kcal mol^{-1} for benzene, and just over 1 kcal mol^{-1} for octahedrane. Molecular spin-orbit corrections can have nontrivial contributions even for first- and second-row radicals and have to be taken into account for quantitative chemical accuracy. For example, they amount to 0.11 (CF), 0.18 (NO), 0.20 (OH, SiH), 0.23 (SiF), 0.28 (OF), 0.45 (ClO), and 0.54 (HS) kcal mol^{-1} *(33)*.

3.4 The DBOC correction

Deviations from the Born–Oppenheimer approximation affect the total atomization energies at the sub-kcal mol^{-1} level even for small and medium-sized hydrocarbons. For example, for our series of hydrocarbon cages, the diagonal Born–Oppenheimer corrections (DBOCs) are 0.23 (tetrahedrane), 0.28 (triprismane), 0.33 (cubane), 0.41 (pentaprismane), 0.54 (octahedrane), 0.78 (dodecahedrane) kcal mol^{-1} *(129)*. These values are calculated at the HF/VTZ level of theory. However, it should be pointed out that for systems with many hydrogens, correlation contributions to the DBOC can reduce the HF DBOC contribution by up to 50% *(34,48,106,128,132,133,148–150)*. For example, for the hydrocarbon cages, correlation contributions calculated at the ΔCCSD/VDZ level reduce the DBOC by −0.08 (tetrahedrane), −0.11 (triprismane), −0.15 (cubane), −0.18 (pentaprismane), −0.22 (octahedrane), and −0.36 (dodecahedrane) kcal mol^{-1} *(129)*. This point is further illustrated by examining a larger set of polycyclic aromatic hydrocarbons (PAHs) with up to 18 carbon atoms *(133)*. The considered set of 20 PAHs includes a diverse range of systems such as benzene, indene, naphthalene, biphenylene, biphenyl, anthracene, pyracene, pyrene, chrysene, and terphenyl. Fig. 7 plots

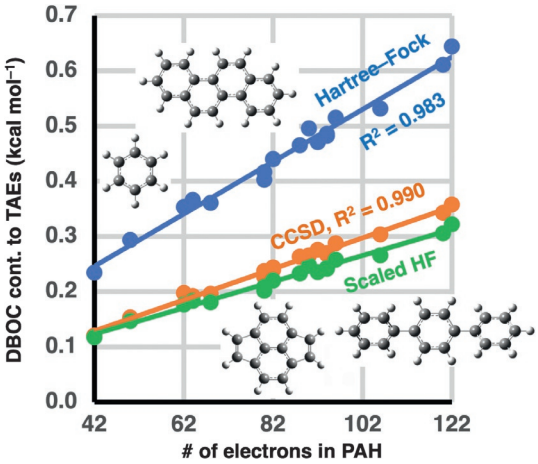

Fig. 7 Overview of the DBOC contribution to the TAEs calculated at the HF (blue line) and CCSD (orange line) levels for a series of 20 PAHs (in kcal mol^{-1}). A number of representative PAHs are shown in the figure (see Ref. *(129)* for further details). The HF DBOC contribution scaled by 0.5 (green line) is in excellent agreement with the CCSD DBOC contribution (see text).

the DBOC to the TAEs calculated at the HF/VTZ and CCSD levels (where CCSD = HF/VTZ + ΔCCSD and ΔCCSD = CCSD/VDZ − HF/VDZ). The ΔCCSD correlation contribution reduces the HF/VTZ DBOC by amounts ranging from 41% (pyracyclene) to 49% (benzene). At the Hartree–Fock level, the DBOC contributions to the TAEs range between 0.23 (benzene) and 0.64 (terphenyl) kcal mol^{-1}. However, at the CCSD level, the DBOC contributions to the TAEs range between 0.12 (benzene) and 0.36 (terphenyl) kcal mol^{-1}.

The above discussion illustrates that DBOC contributions clearly have to be considered in methods in the upper two rungs of Jacob's Ladder (e.g., HEAT, W*n*, and W*n*-F12 ($n = 3, 4$)). Since DBOC calculations are not computationally demanding (at least at the HF level), they are sometimes considered in methods on the second rung of Jacob's Ladder (e.g., W*n*-F12, $n = 1, 2$). However, we recommend that DBOC contributions calculated at the HF level should be scaled by a factor of 0.5. For the above sets of PAHs and $(CH)_n$ hydrocarbon cages, scaling the HF DBOC contribution by 0.5 results in an RMSD of merely 0.03 kcal mol^{-1} relative to the CCSD DBOC contribution (cf. an RMSD of 0.20 kcal mol^{-1} for the unscaled HF DBOC contribution).

Fig. 7 also shows that there is a linear correlation between the DBOC values and the number of electrons in the PAHs. In particular, we obtain squared correlation coefficients of $R^2 = 0.983$ and 0.990 at the HF and CCSD levels, respectively. We note that a perfect linear correlation is not expected since the examined set of PAHs is not a homologous series. Namely, it includes nonaromatic 4- and 5-membered rings (e.g., biphenylene and fluorene) as well as aromatic rings connected via C–C and $–CH_2–$ linkers (e.g., biphenyl, diphenylmethane, and terphenyl) (see Ref. *(129)* for further details).

4. Overview of accuracy and concluding remarks

We conclude this chapter with an overview of the accuracy of several composite ab initio methods from each rung of the composite correlated molecular orbital theory Jacob's Ladder. Fig. 8A gives the RMSDs and 95% confidence intervals for hybrid CCSD(T)/MP*n* methods (rung 1) and pure CCSD(T) methods (rung 2) for the set of 183 TAEs for non-multireference first- and second-row species in the W4–17 database *(103)*. For most of these systems, the reference TAEs are calculated at the CCSDTQ5/CBS level of theory (from W4 and W4.2 theories). For a subset

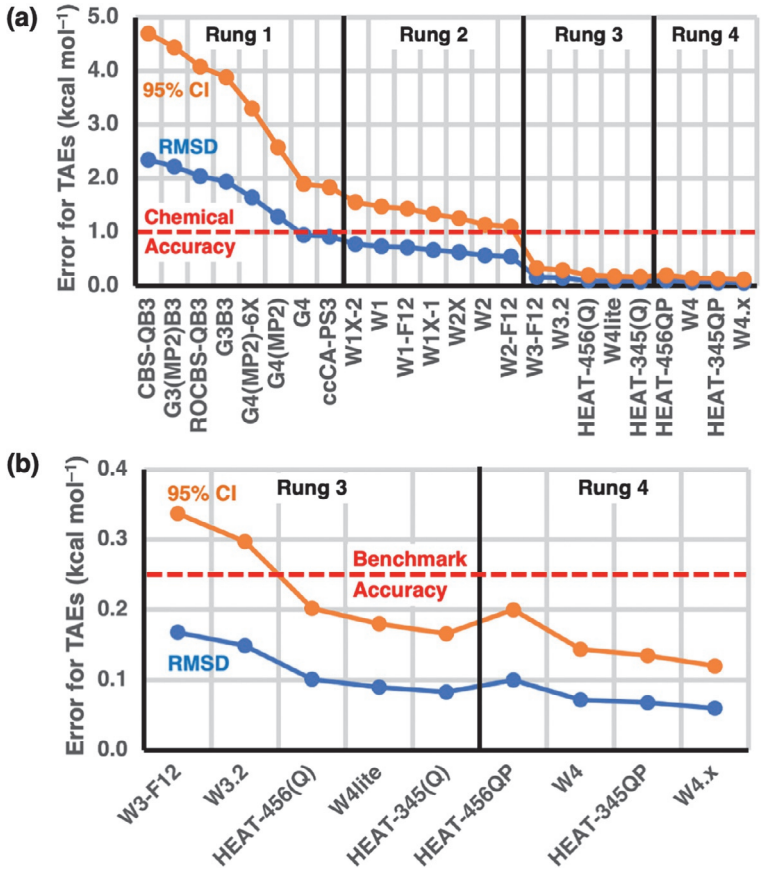

Fig. 8 (A) Overview of the error statistics for total atomization energies (TAEs) for composite ab initio methods across the rungs of Jacob's Ladder of composite correlated molecular orbital theory. The performance of CCSD(T)-based methods (rungs 1 and 2) is evaluated relative to the W4–17 database. The performance of post-CCSD(T) methods (rungs 3 and 4) is evaluated relative to a smaller set of highly accurate experimental TAEs from ATcT (see Ref. *(76)* for further details). Both 95% confidence intervals (95% CIs) and root-mean-square deviations (RMSDs) are given in kcal mol^{-1}. (B) Enlarged view of performance for post-CCSD(T) methods (rungs 3 and 4).

of small molecules (e.g., C_2H_2, HCN, SiH_4, FO, and Cl_2), the reference TAEs are calculated at the CCSDTQ56/CBS level of theory (from W4.3 and W4.4 theories). For a subset of larger molecules (e.g., C_5H_{12}, C_6H_6, CH_3COOH, $HClO_4$, and SF_6), the reference TAEs are calculated at the CCSDT(Q)/CBS level of theory (from W4lite theory). The TAEs in the W4–17 dataset are associated with a 3σ confidence interval of 1 kJ mol^{-1}

and, therefore, should be sufficiently accurate for benchmarking the performance of CCSD(T)-based composite ab initio methods. In terms of chemical diversity, the W4–17 database includes both organic and inorganic species involving single and multiple bonds with varying degrees of covalent and ionic characters. The organic systems include hydrocarbon and halogenated alkanes/alkenes/alkynes, arenes, aromatic heterocycles, non-aromatic heterocycles, alcohols, aldehydes, ketones, anhydrides, carboxylic acids, amines, imines, and nitriles. The inorganic species include halogenated species, boranes, oxides, acids, hydrides, and pure atomic clusters. The set of species used for evaluating the CCSD(T)-based methods spans the gamut from systems dominated by a single reference configuration (e.g., CH_4, CH_3OH, CH_3NH_2) to systems that exhibit appreciable non-dynamical correlation effects (e.g., O_2, SO_3, N_2O_4), however, it excludes systems exhibiting pathological non-dynamical correlation effects (e.g., C_2, O_3, F_2O_2).

Fig. 8 depicts both the RMSDs and 95% CIs to illustrate the significant difference in establishing chemical and benchmark accuracies using these two statistical metrics. We start by noting that, as expected, there is a clear improvement in performance along the rungs of Jacob's Ladder of composite ab initio methods (Fig. 8A). Apart from G4 and ccCA-PS3 theories, the considered hybrid CCSD(T)/MP*n* methods (rung 1) attain RMSDs larger than 1 kcal mol^{-1} for TAEs. For three of the methods (CBS-QB3, G3(MP2)B3, and ROCBS-QB3), the RMSDs are larger than 2 kcal mol^{-1}. However, we note that this level of accuracy is still much better than that obtained for computationally demanding MP2-based and DHDFT methods in conjunction with a quadruple-ζ basis set. For example, the following RMSDs are obtained for a number of representative methods 2.3 (B2GP-PLYP), 2.6 (ωB97X-2 (TQZ)), 2.8 (PWPB95), 3.4 (B2-PLYP), 3.6 (SCS-MP3), and 8.1 (SCS-MP2) kcal mol^{-1} *(103)*. The G3B3, G4(MP2)-6X, and G4(MP2) methods attain RMSDs between 1 and 2 kcal mol^{-1}, and the computationally more demanding G4 and ccCA-PS3 methods attain RMSDs just below 1 kcal mol^{-1}.

The pure CCSD(T)-based methods from the second rung of Jacob's Ladder attain RMSDs that are well below the 1 kcal mol^{-1} mark. For example, the following RMSDs are obtained for representative methods 0.72 (W1-F12), 0.63 (W2X), and 0.55 (W2-F12) kcal mol^{-1}. Nevertheless, these RMSDs still translate to 95% CIs that are above the 1 kcal mol^{-1} chemical accuracy threshold (Fig. 8A).

There is a clear drop in the RMSDs and 95% CIs when moving from the CCSD(T)/CBS methods (rung 2) to the CCSDT(Q)/CBS methods (rung 3) (Fig. 8A). It is important to stress that the theoretical TAEs from the W4–17 database can no longer be used for assessing the post-CCSD(T) composite ab initio methods on rungs 3 and 4. Here, these methods are assessed against highly accurate experimental data from the Active Thermochemical Tables. Fig. 8 gives the RMSDs and 95% CIs for the third-rung methods against a set of 18 first-row ATcT TAEs associated with error bars $\leq$0.06 kcal mol^{-1}. This set comprising only first-row systems is used here so that the HEAT and W*n* methods could be compared on an even keel. However, we note that similar error statistics are obtained for the W*n* methods against a larger set of first- and second-row ATcT atomization energies associated with error bars $\leq$0.05 kcal mol^{-1} (see Ref. *(78)* for further details). The RMSDs for the third-rung methods are 0.168 (W3-F12), 0.149 (W3.2), 0.101 (HEAT-456(Q)), 0.090 (W4lite), and 0.083 (HEAT-345(Q)) kcal mol^{-1}. These RMSDs translate to 95% CIs ranging between 0.337 (W3-F12) and 0.166 (HEAT-345(Q)) kcal mol^{-1}. Thus, as shown in Fig. 8B, nearly all the third rung methods attain benchmark accuracy in terms of the 95% CIs, and all of them attain benchmark accuracy in terms of the RMSDs.

Finally, let us move to the post-CCSDT(Q)/CBS methods on the fourth rung. These methods attain RMSDs $\leq$ 0.1 kcal mol^{-1} and 95% CIs $\leq$ 0.2 kcal mol^{-1}. For example, we obtain the following RMSDs 0.100 (HEAT-456QP), 0.072 (W4), 0.068 (HEAT-345QP), and 0.060 (W4.*x*) kcal mol^{-1}, which translate to 95% CIs of 0.200 (HEAT-456QP), 0.144 (W4), 0.135 (HEAT-345QP), and 0.120 (W4.*x*) kcal mol^{-1}.

Finally, it should be emphasized that the improvement in performance along the rungs of Jacob's Ladder comes with a significant increase in computational cost. For example, methods from the first rung, such as G4(MP2), have been applied to systems as large as C_{60} *(125)*, and methods from the second rung have been applied to systems as large as the C_{24} carbon clusters *(130)*, dodecahedrane $(CH)_{20}$ *(129)*, and PAHs with up to 18 carbons *(133)*. Methods from the third rung can generally be applied to systems with up to ~10 non-hydrogen atoms with current mainstream technology. For example, W3lite theory has been applied to barbaralane (C_9H_{10}) *(141)* and phosphorus sulfide isomers (P_4S_4) *(135)*; and W4lite theory has been applied to C_2X_6 (X = F, Cl) *(103)*, the SF_6^- anion *(108)*, cyclic S_8 *(134)*, and benzene *(36)*. The largest systems methods from the fourth rung have

been applied to normally include highly symmetric species with up to five non-hydrogen atoms. For example, W4 theory has been applied to CCl_4, SiF_4, cyanogen $(CN)_2$, and tetrahedrane $(CH)_4$ *(80,103)*.

Acknowledgments

I gratefully thank Professors Gershom (Jan M. L.) Martin (Weizmann Institute of Science), Leo Radom (University of Sydney), and Krishnan Raghavachari (Indiana University, Bloomington) for many fruitful discussions and insights on this topic over the past two decades.

References

1. Dirac, P. A. M. Quantum Mechanics of Many Electron Systems. *Proc. R. Soc. A* **1929**, *123*, 714.
2. Pauling, L. S. The Application of the Quantum Mechanics to the Structure of the Hydrogen Molecule and Hydrogen Molecule-Ion and to Related Problems. *Chem. Rev.* **1928**, *5*, 173.
3. Narayanan, B.; Redfern, P. C.; Assary, R. S.; Curtiss, L. A. Accurate Quantum Chemical Energies for 133000 Organic Molecules. *Chem. Sci.* **2019**, *10*, 7449.
4. Ratcliff, L. E.; Mohr, S.; Huhs, G.; Deutsch, T.; Masella, M.; Genovese, L. Challenges in Large Scale Quantum Mechanical Calculations. *WIREs Comput. Mol. Sci.* **2017**, *7*, e1290.
5. Riplinger, C.; Sandhoefer, B.; Hansen, A.; Neese, F. Natural Triple Excitations in Local Coupled Cluster Calculations with Pair Natural Orbitals. *J. Chem. Phys.* **2013**, *139*, 134101.
6. Pople, J. A.; Head-Gordon, M.; Fox, D. J.; Raghavachari, K.; Curtiss, L. A. Gaussian-1 Theory: A General Procedure for Prediction of Molecular-Energies. *J. Chem. Phys.* **1989**, *90*, 5622.
7. Curtiss, L. A.; Jones, C.; Trucks, G. W.; Raghavachari, K.; Pople, J. A. Gaussian-1 Theory of Molecular-Energies for 2nd-Row Compounds. *J. Chem. Phys.* **1990**, *93*, 2537.
8. Curtiss, L. A.; Raghavachari, K.; Trucks, G. W.; Pople, J. A. Gaussian-2 Theory for Molecular Energies of First- and Second-Row Compounds. *J. Chem. Phys.* **1991**, *94*, 7221.
9. Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Pople, J. A. Gaussian-3 (G3) Theory for Molecules Containing first And Second-Row Atoms. *J. Chem. Phys.* **1998**, *109*, 7764.
10. Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. Gaussian-4 Theory. *J. Chem. Phys.* **2007**, *126*, 084108.
11. Henry, D. J.; Sullivan, M. B.; Radom, L. G3-RAD and G3X-RAD: Modified Gaussian-3 (G3) and Gaussian-3X (G3X) Procedures for Radical Thermochemistry. *J. Chem. Phys.* **2003**, *118*, 4849.
12. Chan, B.; Coote, M. L.; Radom, L. G4-SP, G4(MP2)-SP, G4-sc, and G4(MP2)-sc: Modifications to G4 and G4(MP2) for the Treatment of Medium-Sized Radicals. *J. Chem. Theory Comput.* **2010**, *6*, 2647.
13. Chan, B.; Deng, J.; Radom, L. G4(MP2)-6X: A Cost-Effective Improvement to G4(MP2). *J. Chem. Theory Comput.* **2011**, 7, 112.
14. Karton, A.; O'Reilly, R. J.; Chan, B.; Radom, L. Determination of Barrier Heights for Proton Exchange in Small Water, Ammonia, and Hydrogen Fluoride Clusters with G4(MP2)-type, MPn, and SCS-MPn Procedures—A Caveat. *J. Chem. Theory Comput.* **2012**, *8*, 3128.

15. da Silva, G. G3X-K theory: A Composite Theoretical Method for Thermochemical Kinetics. *Chem. Phys. Lett.* **2013**, *558*, 109.
16. Chan, B.; Karton, A.; Raghavachari, K.; Radom, L. Restricted Open-Shell G4(MP2)-Type Procedures. *J. Phys. Chem. A* **2016**, *120*, 9299.
17. Chan, B.; Karton, A.; Raghavachari, K. G4(MP2)-XK: A Variant of the G4(MP2)-6X Composite Method with Expanded Applicability for Main Group Elements up to Radon. *J. Chem. Theory Comput.* **2019**, *15*, 4478.
18. Semidalas, E.; Martin, J. M. L. Canonical and DLPNO-Based G4(MP2)XK-Inspired Composite Wavefunction Methods Parametrized Against the GMTKN55 Training Set: Are They More Accurate and/or Robust Than Double-Hybrid DFT? *J. Chem. Theory Comput.* **2020**, *16*, 4238.
19. Petersson, G. A.; Bennett, A.; Tensfeldt, T. G.; Al-Laham, M. A.; Shirley, W. A.; Mantzaris, J. The Total Energies of Closed-Shell Atoms and Hydrides of the First-Row Elements. *J. Chem. Phys.* **1988**, *89*, 2193.
20. Petersson, G. A.; Al-Laham, M. A. A complete basis set model chemistry. II. Open-shell systems and the total energies of the first-row atoms. *J. Chem. Phys.* **1991**, *94*, 6081.
21. Petersson, G. A.; Bennett, A.; Tensfeldt, T. G. A Complete Basis Set Model Chemistry. III. The Complete Basis Set-Quadratic Configuration Interaction Family of Methods. *J. Chem. Phys.* **1991**, *94*, 6091.
22. Montgomery, J. A., Jr.; Ochterski, J. W.; Petersson, G. A. A Complete Basis Set Model Chemistry. IV. An Improved Atomic Pair Natural Orbital Method. *J. Chem. Phys.* **1994**, *101*, 5900.
23. Ochterski, J. W.; Petersson, G. A.; Montgomery, J. A. Jr. A Complete Basis Set Model Chemistry. V. Extensions to Six or More Heavy Atoms; J. Chem. Phys. 1996, 104, 2598.
24. Montgomery, J. A., Jr.; Frisch, M. J.; Ochterski, J. W.; Petersson, G. A. A Complete Basis Set Model Chemistry. VI. Use of Density Functional Geometries and Frequencies. *J. Chem. Phys.* **1999**, *110*, 2822.
25. Wood, G. P. F.; Radom, L.; Petersson, G. A.; Barnes, E. C.; Frisch, M. J.; Montgomery, J. A., Jr. A Restricted-Open-Shell Complete-Basis-Set Model Chemistry. *J. Chem. Phys.* **2006**, *125*, 094106.
26. Allen, W. D.; East, A. L. L.; Császár, A. G. Ab Initio Anharmonic Vibrational Analyses of Non-Rigid Molecules. In *Structures and Conformations of Non-Rigid Molecules*; Laane, J., Dakkouri, M., van der Veken, B., Oberhammer, H., Eds.; Kluwer: Dordrecht, 1993; p. 343.
27. East, A. L. L.; Allen, W. D. The Heat of Formation of NCO. *J. Chem. Phys.* **1993**, *99*, 4638.
28. Klippenstein, S. J.; East, A. L. L.; Allen, W. D. A High Level Ab Initio Map and Direct Statistical Treatment of the Fragmentation of Singlet Ketene. *J. Chem. Phys.* **1996**, *105*, 118.
29. Császár, A. G.; Allen, W. D.; Schaefer, H. F. In Pursuit of the Ab Initio Limit for Conformational Energy Prototypes. *J. Chem. Phys.* **1998**, *108*, 9751.
30. Schuurman, M. S.; Muir, S. R.; Allen, W. D.; Schaefer, H. F. Toward Subchemical Accuracy in Computational Thermochemistry: Focal Point Analysis of the Heat of Formation of NCO and [H,N,C,O] Isomers. *J. Chem. Phys.* **2004**, *120*, 11586.
31. Martin, J. M. L.; de Oliveira, G. Towards Standard Methods for Benchmark Quality Ab Initio Thermochemistry—W1 and W2 Theory. *J. Chem. Phys.* **1843**, *1999*, 111.
32. Boese, A. D.; Oren, M.; Atasoylu, O.; Martin, J. M. L.; Kallay, M.; Gauss, J. W3 Theory: Robust Computational Thermochemistry in the kJ/mol Accuracy Range. *J. Chem. Phys.* **2004**, *120*, 4129.

33. Karton, A.; Rabinovich, E.; Martin, J. M. L.; Ruscic, B. W4 Theory for Computational Thermochemistry: In Pursuit of Confident sub-kJ/mol Predictions. *J. Chem. Phys.* **2006**, *125*, 144108.
34. Karton, A.; Taylor, P. R.; Martin, J. M. L. Basis Set Convergence of Post-CCSD Contributions to Molecular Atomization Energies. *J. Chem. Phys.* **2007**, *127*, 064104.
35. Karton, A.; Martin, J. M. L. Explicitly Correlated Wn Theory: W1-F12 and W2-F12. *J. Chem. Phys.* **2012**, *136*, 124114.
36. Sylvetsky, N.; Peterson, K. A.; Karton, A.; Martin, J. M. L. Toward a W4-F12 Approach: Can Explicitly Correlated and Orbital-Based Ab Initio CCSD(T) Limits Be Reconciled? *J. Chem. Phys.* **2016**, *144*, 214101.
37. 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.
38. Chan, B.; Radom, L. W3X: A Cost-Effective Post-CCSD (T) Composite Procedure. *J. Chem. Theory Comput.* **2013**, *9*, 4769.
39. 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.
40. Fast, P. L.; Sanchez, M. L.; Truhlar, D. G. Multi-Coefficient Gaussian-3 Method for Calculating Potential Energy Surfaces. *Chem. Phys. Lett.* **1999**, *306*, 407.
41. Fast, P. L.; Corchado, J. C.; Sanchez, M. L.; Truhlar, D. G. Multi-Coefficient Correlation Method for Quantum Chemistry. *J. Phys. Chem.* **1999**, *103*, 5129.
42. Fast, P. L.; Truhlar, D. G. MC-QCISD: Multi-Coefficient Correlation Method Based on Quadratic Configuration Interaction with Single and Double Excitations. *J. Phys. Chem. A* **2000**, *104*, 6111.
43. Lynch, B. J.; Truhlar, D. G. Robust and Affordable Multicoefficient Methods for Thermochemistry and Thermochemical Kinetics: The MCCM/3 Suite and SAC/3. *J. Phys. Chem. A* **2003**, *107*, 3898.
44. Lynch, B. J.; Zhao, Y.; Truhlar, D. G. The 6-31B(d) Basis Set and the BMC-QCISD and BMC-CCSD Multicoefficient Correlation Methods. *J. Phys. Chem. A* **2005**, *109*, 1643.
45. Zhang, W.; Kong, X.; Liu, S.; Zhao, Y. Multi-Coefficients Correlation Methods. *WIREs Comput. Mol. Sci.* **2020**, *10*, e1474.
46. Tajti, A.; Szalay, P. G.; Császár, A. G.; Kállay, M.; Gauss, J.; Valeev, E. F.; Flowers, B. A.; Vázquez, J.; Stanton, J. F. HEAT: High Accuracy Extrapolated Ab Initio Thermochemistry. *J. Chem. Phys.* **2004**, *121*, 11599.
47. Bomble, Y. J.; Vázquez, J.; Kállay, M.; Michauk, C.; Szalay, P. G.; Császár, A. G.; Gauss, J.; Stanton, J. F. High-Accuracy Extrapolated Ab Initio Thermochemistry. II. Minor Improvements to the Protocol and a Vital Simplification. *J. Chem. Phys.* **2006**, *125*, 064108.
48. Harding, M. E.; Vázquez, J.; Ruscic, B.; Wilson, A. K.; Gauss, J.; Stanton, J. F. High-Accuracy Extrapolated Ab Initio Thermochemistry. III. Additional Improvements and Overview. *J. Chem. Phys.* **2008**, *128*, 114111.
49. Thorpei, J. H.; Lopez, C. A.; Nguyen, T. L.; Baraban, J. H.; Bross, D. H.; Ruscic, B.; Stanton, J. F. High-Accuracy Extrapolated Ab Initio Thermochemistry. IV. A Modified Recipe for Computational Efficiency. *J. Chem. Phys.* **2019**, *150*, 224102.
50. Thorpe, J. H.; Kilburn, J. L.; Feller, D.; Changala, P. B.; Bross, D. H.; Ruscic, B.; Stanton, J. F. Elaborated Thermochemical Treatment of HF, CO, N_2, and H_2O: Insight into HEAT and Its Extensions. *J. Chem. Phys.* **2021**, *155*, 184109.
51. DeYonker, N. J.; Cundari, T. R.; Wilson, A. K. The Correlation Consistent Composite Approach (ccCA): An Alternative to the Gaussian-*n* Methods. *J. Chem. Phys.* **2006**, *124*, 114104.
52. DeYonker, N. J.; Peterson, K. A.; Steyl, G.; Wilson, A. K.; Cundari, T. R. Quantitative Computational Thermochemistry of Transition Metal Species. *J. Phys. Chem. A* **2007**, *111*, 11269.

53. DeYonker, N. J.; Williams, T. G.; Imel, A. E.; Cundari, T. R.; Wilson, A. K. Accurate Thermochemistry for Transition Metal Complexes from First-Principles Calculations. *J. Chem. Phys.* **2009**, *131*, 024106.
54. Mintz, B.; Williams, T. G.; Howard, L.; Wilson, A. K. Computation of Potential Energy Surfaces with the Multireference Correlation Consistent Composite Approach. *J. Chem. Phys.* **2009**, *130*, 234104.
55. DeYonker, N. J.; Wilson, B. R.; Pierpont, A. W.; Cundari, T. R.; Wilson, A. K. Towards the Intrinsic Error of the Correlation Consistent Composite Approach (ccCA). *Mol. Phys.* **2009**, *107*, 1107.
56. Prascher, B. P.; Lai, J. D.; Wilson, A. K. The Resolution of the Identity Approximation Applied to the Correlation Consistent Composite Approach. *J. Chem. Phys.* **2009**, *131*, 044130.
57. Laury, M. L.; DeYonker, N. J.; Jiang, W.; Wilson, A. K. A Pseudopotential-Based Composite Method: The Relativistic Pseudopotential Correlation Consistent Composite Approach for Molecules Containing 4d Transition Metals (Y–Cd). *J. Chem. Phys.* **2011**, *135*, 214103.
58. Laury, M. L.; Wilson, A. K. Examining the Heavy p-Block with a Pseudopotential-Based Composite Method: Atomic and Molecular Applications of rp-ccCA. *J. Chem. Phys.* **2012**, *137*, 214111.
59. Mahler, A.; Wilson, A. K. Explicitly Correlated Methods within the ccCA Methodology. *J. Chem. Theory Comput.* **2013**, *9*, 1402.
60. Welch, B. K.; Almeida, N. M. S.; Wilson, A. K. Super ccCA (s-ccCA): An Approach for Accurate Transition Metal Thermochemistry. *Mol. Phys.* **2021**, *119*, e1963001.
61. Dixon, D. A.; Feller, D.; Peterson, K. A. A Practical Guide to Reliable First Principles Computational Thermochemistry Predictions Across the Periodic Table. *Annu. Rep. Comput. Chem.* **2012**, *8*, 1.
62. Feller, D.; Peterson, K. A.; Dixon, D. The Impact of Larger Basis Sets and Explicitly Correlated Coupled Cluster Theory on the Feller–Peterson–Dixon Composite Method. *Annu. Rep. Comput. Chem.* **2016**, *12*, 47.
63. Feller, D.; Peterson, K. A.; Dixon, D. A. Refined Theoretical Estimates of the Atomization Energies and Molecular Structures of Selected Small Oxygen Fluorides. *J. Phys. Chem. A* **2010**, *114*, 613.
64. Feller, D.; Peterson, K. A.; Dixon, D. A. Ab Initio Coupled Cluster Determination of the Heats of Formation of $C_2H_2F_2$, C_2F_2, and C_2F_4. *J. Phys. Chem. A* **2011**, *115*, 1440.
65. Feller, D.; Peterson, K. A.; Dixon, D. A. Further Benchmarks of a Composite, Convergent, Statistically Calibrated Coupled Cluster-Based Approach for Thermochemical and Spectroscopic Studies. *Mol. Phys.* **2012**, *110*, 2381.
66. Feller, D.; Peterson, K. A.; Ruscic, B. Improved Accuracy Benchmarks of Small Molecules Using Correlation Consistent Basis Sets. *Theor. Chem. Accounts* **2014**, *133*, 1407.
67. Feller, D. Estimating the Intrinsic Limit of the Feller-Peterson-Dixon Composite Approach When Applied to Adiabatic Ionization Potentials in Atoms and Small Molecules. *J. Chem. Phys.* **2017**, *147*, 034103.
68. Bakowies, D. Ab Initio Thermochemistry Using Optimal-Balance Models with Isodesmic Corrections: The ATOMIC Protocol. *J. Chem. Phys.* **2009**, *130*, 144113.
69. Bakowies, D. Estimating Systematic Error and Uncertainty in Ab Initio Thermochemistry. I. Atomization Energies of Hydrocarbons in the ATOMIC(hc) Protocol. *J. Chem. Theory Comput.* **2019**, *15*, 5230.
70. Bakowies, D. Estimating Systematic Error and Uncertainty in Ab Initio Thermochemistry: II. ATOMIC(hc) Enthalpies of Formation for a Large Set of Hydrocarbons. *J. Chem. Theory Comput.* **2020**, *16*, 399.

71. Vogiatzis, K. D.; Haunschild, R.; Klopper, W. Accurate Atomization Energies from Combining Coupled-Cluster Computations with Interference-Corrected Explicitly Correlated Second-Order Perturbation Theory. *Theor. Chem. Accounts* **2014**, *133*, 1446.
72. Alessandrini, S.; Barone, V.; Puzzarini, C. Extension of the "Cheap" Composite Approach to Noncovalent Interactions: The jun-ChS Scheme. *J. Chem. Theory Comput.* **2020**, *16*, 988.
73. Lupi, J.; Alessandrini, S.; Puzzarini, C.; Barone, V. junChS and junChS-F12 Models: Parameter-free Efficient yet Accurate Composite Schemes for Energies and Structures of Noncovalent Complexes. *J. Chem. Theory Comput.* **2021**, *17*, 6974.
74. Patel, P.; Melin, T. R. L.; North, S. C.; Wilson, A. K. Ab Initio Composite Methodologies: Their Significance for the Chemistry Community. *Annu. Rep. Comput. Chem.* **2021**, *17*, 113.
75. Chan, B. How to Computationally Calculate Thermochemical Properties Objectively, Accurately, and as Economically as Possible. *Pure Appl. Chem.* **2017**, *89*, 699.
76. Karton, A. A computational Chemist's Guide to Accurate Thermochemistry for Organic Molecules. *WIREs Comput. Mol. Sci.* **2016**, *6*, 292.
77. Peterson, C.; Penchoff, D. A.; Wilson, A. K. Prediction of Thermochemical Properties Across The Periodic Table: A Review of the Correlation Consistent Composite Approach (ccCA) Strategies and Applications. *Annu. Rep. Comput. Chem.* **2016**, *12*, 3.
78. Jiang, W.; Wilson, A. K. Ab Initio Composite Approaches: Potential Energy Surfaces and Excited Electronic States. *Annu. Rep. Comput. Chem.* **2012**, *8*, 29.
79. Peterson, K. A.; Feller, D.; Dixon, D. A. Chemical Accuracy in Ab Initio Thermochemistry and Spectroscopy: Current Strategies and Future Challenges. *Theor. Chem. Accounts* **2012**, *131*, 1079.
80. Karton, A.; Daon, S.; Martin, J. M. L. W4-11: A High-Confidence Benchmark Dataset for Computational Thermochemistry Derived from First-Principles W4 Data. *Chem. Phys. Lett.* **2011**, *510*, 165.
81. Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. G*n* Theory. *WIREs Comput. Mol. Sci.* **2011**, *1*, 810.
82. Klopper, W.; Bachorz, R. A.; Hättig, C.; Tew, D. P. Accurate Computational Thermochemistry from Explicitly Correlated Coupled-Cluster Theory. *Theor. Chem. Accounts* **2010**, *126*, 289.
83. DeYonker, N.; Cundari, T. R.; Wilson, A. K. In *Advances in the Theory of Atomic and Molecular Systems (Progress in Theoretical Chemistry and Physics, Vol. 19)*; Piecuch, P., Maruani, J., Delgado-Barrio, G., Wilson, S., Eds.; Springer: Netherlands, Dordrecht, 2009; pp. 197–224.
84. Feller, D.; Peterson, K. A.; Dixon, D. A. A Survey of Factors Contributing to Accurate Theoretical Predictions of Atomization Energies and Molecular Structures. *J. Chem. Phys.* **2008**, *129*, 204105.
85. Helgaker, T.; Klopper, W.; Tew, D. P. Quantitative Quantum Chemistry. *Mol. Phys.* **2008**, *106*, 2107.
86. Fabian, W. M. F. Accurate Thermochemistry from Quantum Chemical Calculations? *Monatsch. Chem.* **2008**, *139*, 309.
87. Martin, J. M. L. Computational Thermochemistry: A Brief Overview of Quantum Mechanical Approaches. *Annu. Rep. Comput. Chem.* **2005**, *1*, 31.
88. Friesner, R. A. Ab Initio Quantum Chemistry: Methodology and Applications. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 6648.
89. Helgaker, T.; Klopper, W.; Bak, K. L.; Halkier, A.; Jørgensen, P.; Olsen, J. Highly Accurate Ab Initio Computation of Thermochemical Data. Understanding Chemical Reactivity. In *Quantum-Mechanical Prediction of Thermochemical Data*; Cioslowski, J., Ed.; vol. 22; Kluwer: Dordrecht, 2001; pp. 1–30.

90. Martin, J. M. L.; Parthiban, S. W1 and W2 Theory and Their Variants: Thermochemistry in the kJ/mol Accuracy Range. Understanding Chemical Reactivity. In *Quantum-Mechanical Prediction of Thermochemical Data*; Cioslowski, J., Ed.; vol. 22; Kluwer: Dordrecht, 2001; pp. 31–65.
91. Dunning, T. H. A Road Map for the Calculation of Molecular Binding Energies. *J. Phys. Chem. A* **2000**, *104*, 9062.
92. Crawford, T. D.; Kumar, A.; Bazanté, A. P.; Remigio, R. D. Reduced-Scaling Coupled Cluster Response Theory: Challenges and Opportunities. *WIREs Comput. Mol. Sci.* **2019**, *9*, e1406.
93. Bartlett, R. J. The Coupled-Cluster Revolution. *Mol. Phys.* **2010**, *108*, 2905.
94. Shavitt, I.; Bartlett, R. J. *Many-Body Methods in Chemistry and Physics: Many-Body Perturbation Theory and Coupled-Cluster Theory*; Cambridge University Press: Cambridge, 2009.
95. Bartlett, R. J.; Musial, M. Coupled-Cluster Theory in Quantum Chemistry. *Rev. Mod. Phys.* **2007**, *79*, 291.
96. Afeefy, H. Y.; Liebman, J. F.; Stein, S. E. In NIST Chemistry WebBook, NIST Standard Reference Database Number 69; Linstrom, P. J.; Mallard, W. G. Eds.; National Institute of Standards and Technology: Gaithersburg, MD, see http://webbook.nist.gov.
97. Ruscic, B.; Pinzon, R. E.; Morton, M. L.; von Laszewski, G.; Bittner, S. J.; Nijsure, S. G.; Amin, K. A.; Minkoff, M.; Wagner, A. F. Introduction to Active Thermochemical Tables: Several "Key" Enthalpies of Formation Revisited. *J. Phys. Chem. A* **2004**, *108*, 9979.
98. Ruscic, B.; Pinzon, R. E.; von Laszewski, G.; Kodeboyina, D.; Burcat, A.; Leahy, D.; Montoya, D.; Wagner, A. F. Active Thermochemical Tables: Thermochemistry for the 21st Century. *J. Phys. Conf. Ser.* **2005**, *16*, 561.
99. Bauschlicher, C. W.; Langhoff, S. R. Quantum Mechanical Calculations to Chemical Accuracy. *Science* **1991**, *254*, 394.
100. Ruscic, B. Uncertainty Quantification in Thermochemistry, Benchmarking Electronic Structure Computations, and Active Thermochemical Tables. *Int. J. Quantum Chem.* **2014**, *114*, 1097.
101. Wheeler, S. E. Homodesmotic Reactions for Thermochemistry. *WIREs Comput. Mol. Sci.* **2012**, *2*, 204.
102. Chan, B.; Collins, E.; Raghavachari, K. Applications of Isodesmic-Type Reactions for Computational Thermochemistry. *WIREs Comput. Mol. Sci.* **2021**, *11*, e1501.
103. Karton, A.; Sylvetsky, N.; Martin, J. M. L. W4-17: A Diverse and High-Confidence Dataset of Atomization Energies for Benchmarking High-Level Electronic Structure Methods. *J. Comput. Chem.* **2017**, *38*, 2063.
104. Mardirossian, N.; Head-Gordon, M. Thirty Years of Density Functional Theory in Computational Chemistry: An Overview and Extensive Assessment of 200 Density Functionals. *Mol. Phys.* **2017**, *115*, 2315.
105. Mehta, N.; Fellowes, T.; White, J. M.; Goerigk, L. CHAL336 Benchmark Set: How Well Do Quantum-Chemical Methods Describe Chalcogen-Bonding Interactions? *J. Chem. Theory Comput.* **2021**, *17*, 2783.
106. Karton, A.; Gruzman, D.; Martin, J. M. L. 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.
107. Johnson, R. D., III. *Computational Chemistry Comparison and Benchmark DataBase, Release 21 Standard Reference Database 101*. http://cccbdb.nist.gov.
108. Karton, A.; Martin, J. M. L. Comment on "Revised electron Affinity of SF6 from Kinetic Data" [J. Chem. Phys. 136, 121102 2012]. *J. Chem. Phys.* **2012**, *136*, 197101.

109. Oyedepo, G. A.; Wilson, A. K. Multireference Correlation Consistent Composite Approach [MR-ccCA]: Toward Accurate Prediction of the Energetics of Excited and Transition State Chemistry. *J. Phys. Chem. A* **2010**, *114*, 8806.
110. Oyeyemi, V. B.; Pavone, M.; Carter, E. A. Accurate Bond Energies of Hydrocarbons from Complete Basis Set Extrapolated Multi-Reference Singles and Doubles Configuration Interaction. *ChemPhysChem* **2011**, *12*, 3354.
111. Oyeyemi, V. B.; Krisiloff, D. B.; Keith, J. A.; Libisch, F.; Pavone, M.; Carter, E. A. Size-Extensivity-Corrected Multireference Configuration Interaction Schemes to Accurately Predict Bond Dissociation Energies of Oxygenated Hydrocarbons. *J. Chem. Phys.* **2014**, *140*, 044317.
112. Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. Gaussian-4 Theory Using Reduced Order Perturbation Theory. *J. Chem. Phys.* **2007**, *127*, 124105.
113. Yu, L.-J.; Karton, A. Assessment of Theoretical Procedures for a Diverse set of Isomerization Reactions Involving Double-Bond Migration in Conjugated Dienes. *Chem. Phys.* **2014**, *441*, 166.
114. Friedrich, J. Efficient Calculation of Accurate Reaction Energies—Assessment of Different Models in Electronic Structure Theory. *J. Chem. Theory Comput.* **2015**, *11*, 3596.
115. Yu, L.-J.; Sarrami, F.; O'Reilly, R. J.; Karton, A. Can DFT and Ab Initio Methods Describe All Aspects of the Potential Energy Surface of Cycloreversion Reactions? *Mol. Phys.* **2016**, *114*, 21.
116. Karton, A.; Goerigk, L. Accurate Reaction Barrier Heights of Pericyclic Reactions: Surprisingly Large Deviations for the CBS-QB3 Composite Method and Their Consequences in DFT Benchmark Studies. *J. Comput. Chem.* **2015**, *36*, 622.
117. Yu, L.-J.; Sarrami, F.; O'Reilly, R. J.; Karton, A. Reaction Barrier Heights for Cycloreversion of Heterocyclic Rings: An Achilles' Heel for DFT and Standard Ab Initio Procedures. *Chem. Phys.* **2015**, *458*, 1.
118. Hohenstein, E. G.; Sherrill, C. D. Wavefunction Methods for Noncovalent Interactions. *WIREs Comput. Mol. Sci.* **2012**, *2*, 304.
119. Liakos, D. G.; Neese, F. Improved Correlation Energy Extrapolation Schemes Based on Local Pair Natural Orbital Methods. *J. Phys. Chem. A* **2012**, *116*, 4801.
120. Goerigk, L.; Karton, A.; Martin, J. M. L.; Radom, L. Accurate Quantum Chemical Energies for Tetrapeptide Conformations: Why MP2 Data with an Insufficient Basis Set Should Be Handled with Caution. *Phys. Chem. Chem. Phys.* **2013**, *15*, 7028.
121. Sedlak, R.; Janowski, T.; Pitonák, M.; Rezác, J.; Pulay, P.; Hobza, P. Accuracy of Quantum Chemical Methods for Large Noncovalent Complexes. *J. Chem. Theory Comput.* **2013**, *9*, 3364.
122. Fogueri, U. R.; Kozuch, S.; Karton, A.; Martin, J. M. L. The melatonin Conformer Space: Benchmark and Assessment of Wavefunction and DFT Methods for a Paradigmatic Biological and Pharmacological Molecule. *J. Phys. Chem. A* **2013**, *117*, 2269.
123. Kesharwani, M. K.; Karton, A.; Martin, J. M. L. Benchmark Ab Initio Conformational Energies for the Proteinogenic Amino Acids Through Explicitly Correlated Methods. Assessment of Density Functional Methods. *J. Chem. Theory Comput.* **2016**, *12*, 444.
124. Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. Assessment of Gaussian-3 and Density Functional Theories on the G3/05 Test Set of Experimental Geometries. *J. Chem. Phys.* **2005**, *123*, 124107.
125. Wan, W.; Karton, A. Heat of Formation for C_{60} by Means of the G4(MP2) Thermochemical Protocol Through Reactions in Which C_{60} Is Broken Down into Corannulene and Sumanene. *Chem. Phys. Lett.* **2016**, *643*, 34.
126. Karton, A.; Waite, S. L.; Page, A. J. Performance of DFT for C_{60} Isomerization Energies: A Noticeable Exception to Jacob's Ladder. *J. Phys. Chem. A* **2019**, *123*, 257.

127. Karton, A. Fullerenes Pose a Strain on Hybrid Density Functional Theory. *J. Phys. Chem. A* **2022**, *126*, 4709.
128. Karton, A.; Chan, B.; Raghavachari, K.; Radom, L. Evaluation of the Heats of Formation of Corannulene and C_{60} by Means of High-Level Theoretical Procedures. *J. Phys. Chem. A* **1834**, *2013*, 117.
129. Karton, A.; Schreiner, P. R.; Martin, J. M. L. Heats of Formation of Platonic Hydrocarbon Cages by Means of High-Level Thermochemical Procedures. *J. Comput. Chem.* **2016**, *37*, 49.
130. Manna, D.; Martin, J. M. L. What Are the Ground State Structures of C_{20} and C_{24}? An Explicitly Correlated Ab Initio Approach. *J. Phys. Chem. A* **2016**, *120*, 153.
131. Harding, M. E.; Vázquez, J.; Gauss, J.; Stanton, J. F.; Kállay, M. Towards Highly Accurate Ab Initio Thermochemistry of Larger Systems: Benzene. *J. Chem. Phys.* **2011**, *135*, 044513.
132. Karton, A.; Yu, L.-J.; Kesharwani, M. K.; Martin, J. M. L. Heats of Formation of the Amino Acids Re-Examined by Means of W1-F12 and W2-F12 Theories. *Theor. Chem. Accounts* **2014**, *133*, 1483.
133. Karton, A.; Chan, B. Accurate Heats of Formation for Polycyclic Aromatic Hydrocarbons: A High-Level Ab Initio Perspective. *J. Chem. Eng. Data* **2021**, *66*, 3453.
134. Karton, A. High-Level Thermochemistry for the Octasulfur Ring: A Converged Coupled Cluster Perspective for a Challenging Second-Row System. *Chem. Phys. Impact* **2021**, *3*, 100047.
135. Kroeger, A. A.; Karton, A. Thermochemistry of Phosphorus Sulfide Cages: An Extreme Challenge for High-Level Ab Initio Methods. *Struct. Chem.* **2019**, *30*, 1665.
136. Sylvetsky, N.; Martin, J. M. L. Probing the Basis Set Limit for Thermochemical Contributions of Inner-Shell Correlation: Balance of Core-Core and Core-Valence Contributions. *Mol. Phys.* **2019**, *117*, 1078.
137. Karton, A. Basis Set Convergence of High-Order Coupled Cluster Methods up to CCSDTQ567 for a Highly Multireference Molecule. *Chem. Phys. Lett.* **2019**, *737*, 136810.
138. Karton, A.; Martin, J. M. L. The Lowest Singlet-Triplet Excitation Energy of BN: A Converged Coupled Cluster Perspective. *J. Chem. Phys.* **2006**, *125*, 144313.
139. Karton, A. How Large Are Post-CCSD(T) Contributions to the Total Atomization Energies of Medium-Sized Alkanes? *Chem. Phys. Lett.* **2016**, *645*, 118.
140. Ramabhadran, R. O.; 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.
141. Kozuch, S.; Schleif, T.; Karton, A. Quantum Mechanical Tunnelling: The Missing Term to Achieve sub-kJ mol^{-1} Barrier Heights. *Phys. Chem. Chem. Phys.* **2021**, *23*, 10888.
142. Ranasinghe, D. S.; Frisch, M. J.; Petersson, G. A. A Density Functional for Core-Valence Correlation Energy. *J. Chem. Phys.* **2015**, *143*, 214111.
143. Pollack, L.; Windus, T. L.; de Jong, W. A.; Dixon, D. A. Thermodynamic Properties of the C5, C6, and C8 n-Alkanes from Ab Initio Electronic Structure Theory. *J. Phys. Chem. A* **2005**, *109*, 6934.
144. Douglas, M.; Kroll, N. M. Quantum electrodynamical corrections to the fine structure of helium. *Ann. Phys.* **1974**, *82*, 89.
145. Heß, B. A. Relativistic Electronic-Structure Calculations Employing a Two-Component No-Pair Formalism with External-Field Projection Operators. *Phys. Rev. A* **1986**, *33*, 3742.
146. Collins, C. L.; Dyall, K. G.; Schaefer, H. F. Relativistic and Correlation Effects in CuH, AgH, and AuH: Comparison of Various Relativistic Methods. *J. Chem. Phys.* **1995**, *102*, 2024.

147. Barysz, M.; Sadlej, A. J. Two-Component Methods of Relativistic Quantum Chemistry: From the Douglas–Kroll Approximation to the Exact Two-Component Formalism. *J. Mol. Struct. THEOCHEM* **2001**, *573*, 181.
148. Manby, F. R.; Stella, M.; Goodpaster, J. D.; Miller, T. F. A Simple, Exact Density-Functional Theory Embedding Scheme. *J. Chem. Theory Comput.* **2012**, *8*, 2564.
149. Karton, A.; Ruscic, B.; Martin, J. M. L. 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.
150. Tajti, A.; Szalay, P. G.; Gauss, J. Perturbative Treatment of the Electron-Correlation Contribution to the Diagonal Born–Oppenheimer Correction. *J. Chem. Phys.* **2007**, *127*, 014102.