CSIRO PUBLISHING

BECKWITH REVIEW
https://doi.org/10.1071/CH24130

AUSTRALIAN JOURNAL OF CHEMISTRY

Three decades of quantum science: how quantum chemistry transformed thermochemical database generation for benchmarking DFT and machine learning

Amir Karton[A,*]

For full list of author affiliations and declarations see end of paper

***Correspondence to:**
Amir Karton
School of Science and Technology, University of New England, Armidale, NSW 2351, Australia
Email: amir.karton@une.edu.au

Handling Editor:
Curt Wentrup

ABSTRACT

In celebration of the United Nations' declaration of 2025 as the International Year of Quantum Science and Technology, marking 100 years since the development of quantum mechanics, this review highlights how accurate quantum mechanical calculations have transformed gas-phase thermochemistry. In particular, the developments of high-level composite *ab initio* methods over the past 30 years enable the calculations of thermochemical properties with confident chemical accuracy (i.e. with 95% confidence intervals ≤1 kcal mol^{-1}) for molecules with up to 12 non-hydrogen atoms. Lower-level composite *ab initio* methods can be applied to molecules containing up to ~50 non-hydrogen atoms; however, they cannot achieve confident chemical accuracy in terms of 95% confidence intervals. Over the past three decades, hundreds of composite *ab initio* methods have been developed, covering different theoretical frameworks, levels of accuracy and computational costs. To guide users in selecting an appropriate composite *ab initio* method for a given system size and level of accuracy, we present a general approach for categorising the accuracy of these methods. This approach places composite *ab initio* methods on four rungs of Jacob's Ladder. Lower rungs offer less accuracy but are applicable to larger systems, and higher rungs offer greater accuracy but are applicable to smaller systems. Each consecutive rung of this ladder represents an improvement in the treatment of the one-particle space, *n*-particle space, or both, leading toward the exact solution of the relativistic Schrödinger equation. The Jacob's Ladder of composite *ab initio* methods can be considered as an extension to the Jacob's Ladder of density functional theory (DFT), which leads from 'Hartree Hell' to the 'Heaven' of double-hybrid DFT methods.

Keywords: CCSD(T), CCSDTQ, chemical databases, composite *ab initio* methods, density functional theory, machine-learning, quantum chemistry.

Received: 12 September 2024
Accepted: 10 February 2025
Published: 17 March 2025

Cite this: Karton A (2025) Three decades of quantum science: how quantum chemistry transformed thermochemical database generation for benchmarking DFT and machine learning. *Australian Journal of Chemistry* **78**, CH24130. doi:10.1071/CH24130

Experimental thermochemical databases

Experimental thermochemistry is a foundational branch of chemistry that is critical to understanding the heat and energy changes involved in chemical transformations. This field generates valuable thermochemical properties through various experimental techniques such as calorimetric, spectrophotometric and mass-spectrometric measurements. These measurements assist in determining reaction mechanisms and predicting reaction outcomes and have broad applications in material science, biochemistry, environmental science and pharmaceuticals. Experimental thermochemistry has been a highly active field of chemical research throughout the 20th Century. Towards the last quarter of the century, the critical mass of thermochemical determinations was compiled into several experimental thermochemical databases. These databases included thermochemical properties, such as heats of formation, bond dissociation energies, electron affinities and ionisation potentials. Popular experimental databases include the National Institute of Standards and Technology (NIST),[1] NIST-JANAF Thermochemical Tables,[2,3,A] Gas-Phase Ion and Neutral Thermochemistry (GIANT),[4]

[A]Data are also accessible online at http://webbook.nist.gov/chemistry and http://srdata.nist.gov/cccbdb.

Thermodynamics Research Center (TRC),[5] CODATA,[6] Gurvich *et al.* [7–9] and Pedley *et al.*[10] There data underpin the prediction of reaction feasibility and elucidation of reaction mechanisms across diverse chemical disciplines. However, with the advancement of correlated wavefunction methods and computer hardware in the 1980s and 1990s, the availability of reliable thermochemical data served another critical role – namely, as reference data in the development of quantum chemical methods capable of accurate thermochemical predictions – the so-called composite *ab initio* methods (also known as composite wavefunction methods). The development of the Active Thermochemical Tables (ATcT) by Ruscic and co-workers in 2004 represented a significant breakthrough in experimental thermochemistry.[11,12] In contrast to traditional thermochemical tabulations, which compile thermochemical experimental determinations, ATcT combines thermochemical data from many experiments and high-level theoretical calculations forming a network where multiple pathways lead to the heat of formation of the same molecule in the network. Since ATcT considers a vast quantity of thermochemical data in one interconnected network, it can identify inconsistencies and potential errors in the data and find the most consistent set of enthalpies for the molecules involved in the network. Overall, this approach leads to thermochemical determinations that are more robust and reliable than the traditional thermochemical databases. Nevertheless, although highly accurate, this approach is still limited to the molecules in the thermochemical network. High-accuracy computational thermochemistry offers an attractive alternative since it is applicable to a much wider range of systems, offering a broader perspective on the world of thermochemistry. Although they do not replace experiments, computational tools have become indispensable, allowing chemists to explore and understand thermodynamic phenomena with unprecedented speed and efficiency. Composite *ab initio* methods are also starting to play a critical role in obtaining reliable thermochemical properties for machine learning databases with hundreds of thousands of thermochemical determinations.[13]

Bird's eye view of composite *ab initio* methods

In the broadest sense, composite *ab initio* methods are multi-step theoretical procedures designed to obtain energetic or spectroscopic chemical properties directly comparable to experimental data. The quest for the development of these methods began in the late 1980s with the pioneering work of Pople and colleagues, who introduced the so-called Gaussian-1 (G1) theory.[14–17] G1 theory established a framework for combining a series of wavefunction electron correlation methods with secondary energetic corrections to obtain thermochemical properties for small molecules. G1 theory was soon followed by Gausian-2 (G2) theory,[18] which provided a precise recipe for obtaining thermochemical data. In short, G2 theory calculates the electronic energy using quadratic configuration interaction theory (QCISD(T)) with the 6-311G(d,p) basis set, whereas basis set corrections to account for the effect of diffuse and higher polarisation functions are calculated using fourth-order Møller–Plesset (MP4) perturbation theory. The core-valence correction is calculated using second-order Møller–Plesset (MP2) perturbation theory and the zero-point vibrational energy at the Hartree–Fock level. G2 theory also includes an empirical higher-level correction (HLC) term to account for some of the remaining deficiencies in the theoretical model. The empirical parameter involved in the HLC term was parameterised against the 55 experimental total atomisation energies (TAEs) in the G2 test set. For the entire G2 test set, which includes 55 TAEs, 38 ionisation energies, 25 electron affinities and 7 proton affinities, G2 theory attains a mean absolute deviation (MAD) of 1.24 kcal mol^{-1}. This pioneering work demonstrated the critical role that reliable experimental thermochemical data play in the development and evaluation of composite *ab initio* methods. Incidentally, 1 kcal mol^{-1} has become the yardstick for 'chemical accuracy' in the computational determination of thermochemical properties (Karton[19] provides a detailed overview of the term 'chemical accuracy'). G2 theory was followed by G3 theory[20] in 1998 – the year John Pople was awarded the Nobel Prize in Chemistry for the 'development of computational methods in quantum chemistry'. Undoubtedly, the Nobel Prize was also awarded for the development of the Gaussian-*n* composite *ab initio* methods, which were noted in Pople's Nobel Lecture (8 December 1998). Here, we must note the highly influential computational and theoretical chemists who had major contributions to these remarkable developments, namely Larry Curtiss, Paul Redfern and Krishnan Raghavachari.[21] For a comprehensive review of the Gaussian-*n* methods by these authors, see Curtiss *et al.*[22] Overall, dozens of Gaussian-*n* type methods have been developed over the years, including the reduced order perturbation theory G*n*(MP2) methods,[23–26] which are computationally more economical and are applicable to systems as large as buckminsterfullerene (C_{60}) and its isomers.[27,28] The development of G*n*-type methods by other internationally renowned groups, such as the group of Leo Radom, should also be noted here.[29–35]

The development of the early variants of the Gaussian-*n* methods has sparked extensive theoretical developments in the area of computational thermochemistry, leading to the development of many more types and variants of composite *ab initio* methods. These include the complete basis set (CBS) model chemistries,[36–42] focal-point analysis (FPA),[43–47] Weizmann-*n* (W*n*),[48–53] W*n*X,[54–56] multi-coefficient correlation methods (MCCMs),[57–62] high-accuracy extrapolated *ab initio* thermochemistry (HEAT),[63–67] correlation consistent composite approach (ccCA),[68–77] Feller–Peterson–Dixon (FPD),[78–84] *ab initio* thermochemistry using optimal-balance models with isodesmic corrections (ATOMIC),[85–87] interference-corrected explicitly correlated second-order perturbation theory (INT-MP2-F12)[88] and the so-called cheap composite scheme (ChS)[89,90] procedures. For in-depth reviews of composite *ab initio* methods, see Curtiss *et al.*,[18] Raghavachari,[21] Karton,[91,92] Patel *et al.*,[93]

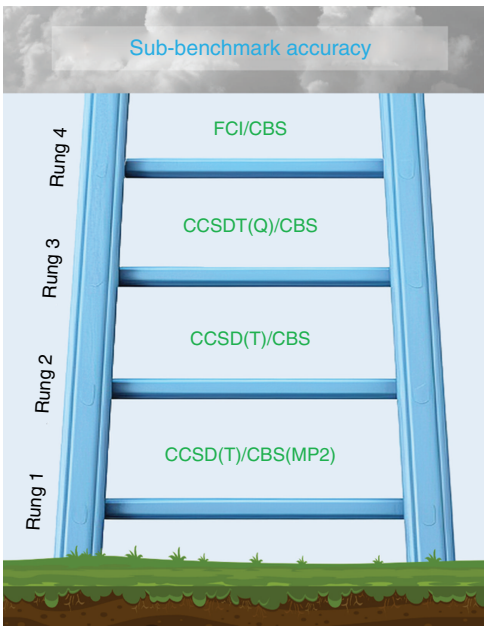

Fig. 1. Jacob's Ladder of composite *ab initio* methods illustrates the four categories of these methods. Each successive rung represents a more rigorous treatment of the one-particle space, the *n*-particle space, or both. Methods on the first rung (e.g. G4(MP2) theory) are suitable for large systems such as C_{60} isomers, whereas those on the fourth rung (e.g. W4-F12 theory) are limited to much smaller molecules like pentane (C_5H_{12}).

Chan,[94] C. Peterson *et al.*,[95] Feller *et al.* [79,96], Dixon *et al.*,[78] K. A. Peterson *et al.*,[97] Jiang and Wilson,[98] Klopper *et al.*,[99] DeYonker *et al.*,[100] Helgaker *et al.*,[101,102] and Martin.[103,104] Here, we will give a bird's eye view of the different types of composite *ab initio* methods and their expected accuracies. Broadly speaking, composite *ab initio* methods can be classified based on the highest level of coupled-cluster excitation considered (e.g. CCSD(T), CCSDT(Q), CCSDTQ, CCSDTQ5 and CCSDTQ56) and the level of basis set completeness it approximates. With numerous composite *ab initio* methods, spanning dozens of variants across multiple families (as outlined above), it is essential to establish clear guidelines for assessing their expected accuracy. Here, we propose a Jacob's Ladder framework, where the methods are placed on rungs based on their expected accuracy. Each successive rung of Jacob's Ladder represents a more rigorous treatment of the one-particle space, *n*-particle space, or both. Fig. 1 depicts the Jacob's Ladder of composite *ab initio* methods (for further details, see also Karton[19]).

The hierarchical structure of composite *ab initio* methods

The first rung of Jacob's Ladder represents methods that approximate the CCSD(T)/CBS energy using MP2-based basis-set additivity schemes. This rung includes computationally economical composite *ab initio* methods such as G*n*, CBS and ccCA. These methods are commonly denoted as CCSD(T)/CBS(MP2)[105–107] and use the following simple basis set additivity scheme:

$$\begin{aligned}\mathrm{CCSD(T)/Large} &= \mathrm{CCSD(T)/Small} \\ &+ \mathrm{MP2/Large} - \mathrm{MP2/Small} \\ &+ [\text{additional corrections}] \end{aligned} \quad (1)$$

In Eqn 1, Small and Large denote different basis set sizes, typically corresponding to at least double-ζ and triple-ζ quality, respectively. For example, in one of the most computationally efficient CCSD(T)/CBS(MP2) methods (i.e. G4(MP2)),[108] 'Small' refers to the 6-31G(d) basis set and 'Large' refers to the G3MP2LargeXP basis set (i.e. a modified version of the 6-311 + G(3df,2p) basis set). The computationally most demanding step in the G4(MP2) method is typically the CCSD(T)/6-31G(d) calculation. Thus, the G4(MP2) method can be routinely applied to large systems with over 40 carbon atoms, e.g. the entire set of C_{40} isomers[109,110] and even to a subset of C_{60} isomers.[26,27] It is also noteworthy that the G4(MP2) method has been applied to the 133,000 molecules with up to nine non-hydrogen first-row atoms,[13,111] thus providing a valuable database for benchmarking density functional theory (DFT) methods.[111] For comparison, the ccCA-PS3 method is an example of a computationally more demanding CCSD(T)/CBS(MP2) method.[71] In this method, 'Small' refers to the cc-pV(T + d)Z basis set and 'Large' refers to a basis set extrapolation from basis sets of up to aug-cc-pV(Q + d)Z quality.[112–114] Thus, the ccCA-PS3 method has been applied to smaller systems than G4(MP2). Large systems to which the ccCA-PS3 method was applied to include up to ~20 non-hydrogen atoms, e.g. highly energetic molecules such as HMX ($C_4H_8N_8O_8$) and PETN ($C_5H_8N_4O_{12}$),[115] melatonin conformers ($C_{13}H_{16}N_2O_2$)[116] and ditetrazinetetroxide ($C_2N_8O_4$).[117] Depending on the specific CCSD(T)/CBS(MP2) composite *ab initio* at hand, additional corrections to the electronic energy in Eqn 1 may include basis set correction terms for the effects of diffuse and higher polarisation functions, a complete basis set extrapolation correction for the Hartree–Fock energy and a core-valence correction.

The second rung of Jacob's Ladder involves CCSD(T)/CBS composite *ab initio* methods that do not involve Moller–Plesset perturbation theory (MP*n*) based corrections terms. That is, methods that extrapolate the components of the CCSD(T) energy to the complete basis set limit by the following general expression:

$$\begin{aligned}\mathrm{CCSD(T)/CBS} &= \mathrm{HF/CBS} + \mathrm{CCSD^{corr}/CBS} + \mathrm{(T)/CBS} \\ &+ [\text{additional corrections}] \end{aligned} \quad (2)$$

Here $\mathrm{CCSD^{corr}}$ and (T) are the CCSD and (T) correlation energies respectively. Methods on rung two of Jacob's Ladder involve a more rigorous treatment of the one-particle space when compared to methods from the first rung. Examples of such methods include the lower members of the Weizmann-*n* family of composite *ab initio* methods (e.g. W1, W2, W1-F12 and

W2-F12).[48,52,104,118] W1-F12 theory is an example of a computationally economical method of this category. In this method, the HF energy and CCSD-F12 correlation energy are extrapolated separately to the CBS limit from the cc-pVDZ-F12 and cc-pVTZ-F12 basis sets.[119] The (T) correlation energy is extrapolated from the regular jul-cc-pV{D,T}Z basis set pair.[120] The computationally most demanding steps in W1-F12 theory are the CCSD-F12/cc-pVTZ-F12 and CCSD(T)/jul-cc-pV(T + d)Z calculations. Therefore, methods such as W1 and W1-F12 theories are still applicable to reasonably large systems. For example, molecules with up to 21 non-hydrogen atoms such as sumanene ($C_{21}H_{12}$),[27,121] dodecahedrane ($(CH)_{20}$),[122] corannulene ($C_{20}H_{10}$),[27,121] terphenyl ($C_{18}H_{14}$)[123] and chrysene ($C_{18}H_{12}$).[123] W2-F12 theory is an example of a higher-end CCSD(T)/CBS method, which attains results closer to the true basis set limit compared to W1-F12 theory. The computationally most demanding steps in W2-F12 theory are the CCSD-F12/cc-pVQZ-F12 and CCSD(T)/cc-pVTZ-F12 calculations. The largest systems W2-F12 theory was applied to include molecules with ~10 non-hydrogen atoms such as methionine ($C_5H_{11}NO_2S$),[124] adenine ($C_5H_5N_5$)[52] and cubane ($(CH)_8$).[122]

Rung three of Jacob's Ladder additionally incorporates post-CCSD(T) contributions up to the CCSDT(Q) level into Eqn 2:

$$\begin{aligned}\mathrm{CCSDT(Q)/CBS} = {} & \mathrm{HF/CBS} + \mathrm{CCSD^{corr}/CBS} + \mathrm{(T)} \\ & \mathrm{/CBS} + \mathrm{T\text{–}(T)/CBS} + \mathrm{(Q)/CBS} \\ & + \mathrm{[additional\ corrections]} \end{aligned} \quad (3)$$

Thus, moving from rung two to rung three represents a more rigorous treatment of the *n*-particle space. Methods that belong to this rung typically build on the top-end methods of rung two by adding higher-order triples excitations CCSDT–CCSD(T) (T–(T)) and quasiperturbative quadruple excitations ((Q)). For example, W3-F12 theory, which approximates the CCSDT(Q)/CBS energy, uses W2-F12 theory as a baseline and adds a T–(T) contribution extrapolated from the cc-pV{D,T}Z basis set pair and a (Q) contribution calculated with the cc-pVDZ basis set. It is important to mention here that successively higher coupled cluster expansion terms (ΔCCSD, Δ(T), ΔT–(T), Δ(Q), ΔQ–(Q), Δ(5), etc.) converge increasingly faster with the basis set size.[19,50,51] Indeed, this is the reason that composite *ab initio* methods from rungs three and four of Jacob's Ladder can be performed at a realistic computational cost. The computationally most demanding step in W3-F12 theory is typically the CCSDT/cc-pVTZ calculation.[125] Therefore, rung three methods are not generally applicable to systems with over 10 non-hydrogen atoms. The largest systems W3-F12 (or W3) theory have been applied to include the benzene dimer ($C_{12}H_{12}$),[107] bullvalene ($(CH)_{10}$),[126] octasulfur ring (S_8),[127] phosphorus sulfide cages (P_4S_3 and P_4S_4)[128] and hexahaloethanes (C_2X_6).[129,130]

Composite *ab initio* methods on the fourth rung of Jacob's Ladder represent the most accurate methods in contemporary quantum chemistry that are capable of confident sub-benchmark accuracy even for pathologically multireference systems such as O_3 and $C_2(\Sigma_g^+)$.[50] Moving from rung three to rung four typically involves a more rigorous treatment of both the one-particle and *n*-particle spaces in order to achieve confident sub-benchmark accuracy. To illustrate this, it is instructive to compare the CCSD(T) and post-CCSD(T) components in W3 and W4 theories. In W3 theory, which sits on rung three of Jacob's Ladder, the CCSD(T) energy is obtained by CCSD(T)/CBS = HF-AV{Q,5}Z + CCSD/AV{Q,5}Z + (T)/AV{T,Q}Z. Whereas in W4 theory (rung four), all the basis sets are upgraded by one angular momentum, namely CCSD(T)/CBS = HF/AV{5,6}Z + CCSD/AV{5,6}Z + (T)/AV{Q,5}Z. Likewise, the basis set used for calculating the parenthetical connected quadruples ((Q)) is cc-pVDZ in W3 and cc-pVTZ in W4. Moving to the *n*-particle space, W3 does not include correlation contributions beyond CCSDT(Q), whereas W4 theory calculates the CCSDTQ–CCSDT(Q) component with a cc-pVDZ basis set and the CCSDTQ5–CCSDTQ component with a truncated version of the cc-pVDZ basis set.[19,50,92] The significantly improved treatment of the one-particle and *n*-particle spaces in rung four composite *ab initio* methods results in a prohibitive computational cost. For example, the largest systems W4 theory has been applied to include up to five non-hydrogen atoms (e.g. tetrafluorosilane, tetrachloromethane, acetic acid, tetrahedrane and *n*-butane).[53]

So far, we have categorised the composite *ab initio* methods on the rungs of Jacob's Ladder based on the levels of accuracy used for calculating the non-relativistic electronic energy. In order to obtain thermochemical and kinetic data that are directly comparable to experimental measurements, composite *ab initio* methods include secondary energetic corrections such as the core-valence, scalar relativistic, spin-orbit, Born–Oppenheimer and zero-point vibrational energy (ZPVE) corrections. Excluding the ZPVE correction, these calculations are typically computationally less demanding than those required for the non-relativistic electronic energy. For a comprehensive overview of secondary energetic corrections in composite *ab initio* methods, the reader is referred to several recent reviews.[19,91,92]

The performance of composite *ab initio* methods has been extensively benchmarked relative to accurate and reliable experimental and high-level theoretical data. It is well established that the performance of electronic structure methods depends on both the chemical properties being considered (e.g. heats of formation, reaction barrier heights, conformational energies or non-covalent interactions) and the specific composition of the evaluation dataset (e.g. elemental composition and multireference character of the species involved). The atomisation energy is the energy required to break a molecule into its constituent atoms in the gas phase. As such, TAEs are one of the most challenging

thermochemical properties and, therefore, are commonly used for the evaluation of composite *ab initio* methods.[131] In this context, it should be pointed out that TAEs can be converted to molecular heats of formation at 0 K from the atomic heats of formation at 0 K (which can be taken from experiments) and the zero-point vibrational energy (which can be obtained from CCSD(T) harmonic frequencies and DFT quartic force fields).[50,92,132,133]

The performance of composite *ab initio* methods from rungs one and two of Jacob's can be evaluated relative to a wide and diverse set of TAEs obtained from methods on the fourth rung of Jacob's Ladder. The W4-17 database of 200 TAEs represents such a dataset.[129] It comprises first- and second-row molecules with up to eight non-hydrogen atoms, which cover a broad spectrum of bonding situations, electronic states and multireference characters. Since methods from rungs one and two approximate the CCSD(T) energy, which is not suitable for highly multireference systems, pathologically multireference species (e.g. halogen oxides and atomic clusters)[19,53,91,92,129,134–137] are excluded from the evaluation dataset. Two recent reviews[19,91] give a comprehensive overview of the performance of popular composite *ab initio* methods for TAEs. As customary in experimental (and high-level computational) thermochemistry, we will use 95% confidence intervals (CIs) rather than root-mean-square deviations (RMSDs) or MADs for robust uncertainty quantification of composite *ab initio* methods.[138,139] We note that the 95% CI is approximately equal to twice the RMSD for a normal distribution. As previously noted,[19] methods from rung one attain 95% CIs ranging between 2 and 5 kcal mol^{-1} and methods from the second rung attain 95% CIs ranging between 1 and 2 kcal mol^{-1}. Thus, in terms of 95% CIs, methods from rung one are not capable of chemical accuracy for TAEs (arbitrarily defined as deviations of ~1.0 kcal mol^{-1}), and methods from rung two are capable of near chemical accuracy for TAEs. Methods from the third and fourth rungs can only be evaluated against highly accurate and reliable experimental TAEs, which are typically taken from the ATcT network of Ruscic and co-workers.[11,12] Relative to a subset of ATcT TAEs, methods from rung three of Jacob's Ladder attain 95% CI of near or sub-benchmark accuracy for TAEs (arbitrarily defined as deviations of ~1.0 kJ mol^{-1}), and methods from rung four are capable of confident sub-benchmark accuracy.[19] These results show a progressive improvement in performance as we ascend the rungs of Jacob's Ladder, which validates the more rigorous theoretical framework. It should be emphasised that the above 95% CIs are obtained for one of the most challenging thermochemical properties and that the performance of the composite *ab initio* methods can improve significantly for less challenging thermochemical properties, in particular, thermochemical properties associated with reactions that conserve large molecular fragments on both sides of the reaction.[122,130,140–148] Finally, it is important to note that composite *ab initio* methods are extensively used not just for obtaining accurate energy-based thermochemical properties, but also spectroscopic properties (e.g. equilibrium geometries, vibrational frequencies and rotational constants)[89,90,96,149–158] and electrical properties (e.g. dipole moments, polarisabilities and hyperpolarisabilities).[84,154,159]

Theorretical thermochemical databases

As mentioned in the previous section, experimental thermochemical databases such as G2,[18] G2/97,[160] G3/99,[161] Database/3[59] and G3/05,[162] played a foundational role in the development of composite *ab initio* methods. Since the development of the ATcT network in 2004, ATcT has become the dominant thermochemical database because it provides highly accurate, reliable and internally consistent thermochemical values. In addition, ATcT is a dynamic database that is regularly updated, for example, the current version of ATcT includes ~3000 heats of formation determined both at 0 and 298 K (ATcT Network, Argonne National Laboratory, see https://atct.anl.gov/). For comparison, the largest of the test sets listed above,G3/05, includes 454 thermochemical determinations, namely 270 enthalpies of formation, 105 ionisation energies, 63 electron affinities, 10 proton affinities and 6 hydrogen-bonded complexes.

When it comes to benchmarking electronic structure methods, one limitation of the above experimental databases is that they focus on a small number of thermochemical properties (e.g. ΔH_f, TAEs, IPs, EAs, PAs), most of which are obtained for small species with 1–9 non-hydrogen atoms. Theoretical databases generated using composite *ab initio* methods overcome these limitations since they are accessible to a much wider range of thermochemical and kinetic properties, and composite *ab initio* methods from rungs one and two of Jacob's Ladder are applicable to much larger systems with dozens of non-hydrogen atoms. Contemporary composite *ab initio* methods span from methods capable of near chemical accuracy and applicable to systems with ~20 non-hydrogen atoms, to methods capable of sub-benchmark accuracy and applicable to systems with 6 non-hydrogen atoms.[19,91,92] These theoretical developments, along with advances in high-performance computer technology, have led to a dramatic increase in the scale and diversity of theoretical thermochemical databases since the mid-2000s.[163] The databases generated by composite *ab initio* methods are chemically more diverse and cover a wider range of chemical properties than the experimental ones. For example, they paved the way for the generation of databases focusing on chemical properties that are typically not included in experimental databases such as isomerisation energies,[130,164,165] conformational energies,[166–170] various reaction energies,[130,171,172] (including hypothetical species),[173] reaction barrier heights[174–181] and other chemical properties (e.g. self-interaction errors[182] and radical stabilisation energies).[183–185]

There are many dozens of examples of high-quality theoretical thermochemical databases. For an overview, the reader is referred to Goerigk and Grimme,[182,186] Goerigk *et al.*,[187] Mardirossian and Head-Gordon,[188] and Goerigk.[189] Here, we will only describe two extreme cases the W4-11 database,[130] which includes close to 1000 reaction energies calculated by W4 theory from rung four of Jacob's Ladder, and the GDB-9 database, which includes over 133,000 TAEs calculated by G4(MP2) theory from rung one of Jacob's Ladder.[13]

Let us begin with the W4-11 database, which includes the following sets of reaction energies: 140 TAEs, 99 bond dissociation energies, 707 heavy-atom transfer reaction energies, 20 isomerisation energies and 13 nucleophilic substitution reaction energies, totalling 979 reaction energies. All the reaction energies are all-electron, relativistic, ZPVE-inclusive and DBOC-inclusive CCSDTQ5/CBS energies. The largest systems represented in the W4-11 database are molecules such as CF_4, FOOF, acetic acid, SiF_4, SO_3, P_4 and S_4. In the original work, this extensive set of 797 highly accurate reaction energies was used to benchmark the performance of composite *ab initio* methods from rungs one and two of Jacob's Ladder as well as a range of contemporary DFT and DHDFT methods.[130] In 2017, this set of 140 TAEs was extended to include 27 additional TAEs from W4 theory (rung four) and 33 TAEs from W4lite (rung three of Jacob's Ladder) in the W4-17 database.[129] Although the W4-11 and W4-17 databases achieve the most accurate thermochemical properties by composite *ab initio* methods, the high computational cost associated with rung four methods, and in particular the CCSDTQ/cc-pVDZ and CCSDTQ5/DZ calculations, limits their applicability to relatively small systems with up to 5–6 non-hydrogen atoms. It should be pointed out that the atomisation energies in the W4-11 and W4-17 databases are not more accurate than the most accurate TAEs in the ATcT database. However, unlike experiments, W4 theory is applicable to any arbitrary system with up to 5–6 first- and second-row atoms, whether it is poisonous, explosive, short-lived or hypothetical.

On the other extreme, methods such as G4(MP2) are applicable to systems with dozens of non-hydrogen atoms. Therefore they have been applied for the calculation of the isomerisation energies of 40 C_{40} isomers[109,110] and 8 C_{60} isomers.[26,27] A particularly impressive application of the G4(MP2) method has been the calculation of the atomisation energies of over 133,000 molecules with 9 non-hydrogen atoms in the GDB-9 database.[13,190] This makes this database of atomisation energies highly diverse, albeit at the price of reduced accuracy relative to the W4-11 and W4-17 databases.[129,130] (For a comprehensive discussion of the composition of the molecules in the GDB-9 database, see Narayanan *et al.*,[13] Ramakrishnan *et al.*,[190] and Huang and von Lilienfeld[191].) The largest systems W1 and W1-F12 theories have been applied to include medium-sized hydrocarbons such as corannulene ($C_{20}H_{10}$),[121] sumanene ($C_{21}H_{12}$),[121] dodecahedrane ($(CH)_{20}$)[122] and the C_{20} and C_{24} carbon clusters.[192]

Finally, it should be emphasised that computational acceleration techniques such as the resolution-of-the-identity (RI)[193–196] and explicitly correlated approximations[52,197–201] and localised-orbital approaches[202–208] have enabled the application of composite *ab initio* methods to large molecular systems by reducing their computational cost. Examples of composite *ab initio* methods that use these acceleration techniques include Wn-F12,[52,53] WnX,[54–56] L-W1X,[209] ccCA-F12,[76] G4(MP2)-XK,[26] G4(MP2)-XK-D[35] and cc-G4-n.[210]

A significant advantage of using theoretical rather than experimental data for benchmarking electronic structure methods is that the theoretical benchmark data are directly comparable with the more approximate data obtained from DFT and low-level electronic structure calculations. Namely, electronic structure methods calculate gas-phase, non-relativistic electronic energies at the bottom of the well. By contrast, thermochemical properties obtained from experiments include additional components such as relativistic, zero-point vibrational, thermo-statistical, entropic and solvation corrections. In order to compare the electronic structure and experimental data, these components have to be added to the theoretical calculations or backtracked from the experimental values.[211] Alternatively, reference values that are calculated by non-relativistic high-level wavefunction methods (e.g. CCSD(T) or CCSDT(Q)) are directly comparable to those obtained from DFT (e.g. B3LYP, PBE, M06-2X, ωB97X-D) or low-level *ab initio* (e.g. HF, MP2, SCS-MP2) calculations. Therefore, data obtained from high-level wavefunction methods can be readily used for the evaluation and development of lower-level electronic structure methods.

It should be emphasised that backtracking the above secondary energetic contributions from experimental determinations (or adding them to non-relativistic, bottom-of-the-well theoretical determinations) is not a trivial task. This process is likely to increase the uncertainty of the experimental values when the secondary energetic contributions are not calculated with sufficient accuracy (e.g. when they are obtained from DFT calculations). Indeed, much of the research into high-level composite *ab initio* methods has been dedicated to obtaining secondary energetic contributions (such as scalar relativistic, spin-orbit, zero-point vibrational and Born–Oppenheimer corrections) with sufficient levels of accuracy (see Karton[19,91,92] for an overview). As an illustrative example, let us consider the zero-point vibrational energy (ZPVE) component of small-to-medium-sized molecules. The ZPVE for simple organic molecules ethane, pentane, arginine and dodecahedrane is 46.39, 98.88, 138.18 and 222.97 kcal mol^{-1} respectively.[122,124,164,212] Thus, even a 1% error in the calculated ZPVE due to the neglect of explicit anharmonicity can translate to errors on the order of 1–2 kcal mol^{-1} in the ZPVE for small molecules.

Electronic structure calculations are generally easier to perform than experiments and can be applied to a wide

range of systems that may be difficult or impossible to study experimentally, including poisonous, explosive, short-lived or hypothetical species. Electronic structure calculations can be applied to a broad spectrum of thermochemical and kinetic properties, including those that are difficult to measure accurately, such as some reaction barrier heights and arbitrary bond dissociation energies. This allows the construction of large, diverse and systematic datasets. However, a significant limitation of low-level electronic structure methods, such as DFT, is their inability to consistently and confidently achieve high accuracy across a wide range of chemical properties and systems. This underscores the need to regularly benchmark DFT methods to ensure reliability and precision across diverse applications.[186–189] With the exception of multireference systems, composite *ab initio* methods from rungs one and two of Jacob's Ladder largely overcome this limitation. Whereas composite *ab initio* methods from rungs three and four of Jacob's Ladder can be safely applied to moderately and highly multireference systems respectively.[19,91,92]

An advantage of theoretical databases over experimental ones is that high-level thermochemical data can be obtained reasonably easily at a reasonable computational cost if the rung of Jacob's Ladder is chosen according to the system size. For instance, rung four methods are applicable for systems with up to ~5 non-hydrogen atoms, rung three for systems with up to ~10 non-hydrogen atoms, rung two for systems with up to ~16 non-hydrogen atoms and rung one for systems with up to ~32 non-hydrogen atoms. This ease of data generation offers significant flexibility in designing diverse and targeted databases for specific chemical systems and properties. This allows the generation of systematic databases targeted at certain elements, compounds and substituents, as well as tackling specific chemical properties.

Synergy of accurate computational thermochemistry databases and machine learning

Machine learning (ML) is transforming the field of chemistry by enabling novel approaches to predict and analyse the chemical properties of molecules, proteins and materials. A recent testament to the impact of ML in chemistry was the 2024 Nobel Prize in Chemistry awarded to David Baker, a computational biochemist from the University of Washington, and to Demis Hassabis and John Jumper from Google, for their groundbreaking work in computational protein design and protein structure prediction. ML applications in chemistry span from designing functional molecules, proteins and materials with tailored properties to optimising reaction conditions and synthetic routes. A key approach to achieving these advances is through the prediction of critical properties, such as binding energies, chemical stabilities and catalytic activities. However, for reliable and accurate predictions, ML models require vast and diverse datasets of high-quality chemical data.[191,213–224]

Over the past two decades, composite *ab initio* methods have been extensively used for generating highly accurate thermochemical, kinetic and non-covalent interaction databases for (i) training empirical DFT methods and (ii) benchmarking the performance of empirical and non-empirical DFT methods.[91,163,189] These databases are typically calculated using composite *ab initio* methods from rungs one to four of Jacob's Ladder of composite *ab initio* methods (Fig. 1). Over the past decade, there have been several compilations of many small databases into large general-purpose ones that cover a wide range of chemical systems and properties. Examples include the GMTKN55,[188] MGCDB84[189] and 2015B[225] databases (for a broader overview, see Karton and de Oliveira[163] and Goerigk[189]). Together, these general-purpose databases include well over 5000 accurate thermochemical, kinetic and non-covalent interaction determinations. These compilations have proved to be highly valuable for training and benchmarking DFT methods over the past decade.[163,189] Nevertheless, in terms of their size, these compilations are still not comparable to databases used for training ML models, which typically require tens to hundreds of thousands of data points to capture complex patterns and dependencies across diverse chemical systems.

DFT has become the workhorse of quantum chemistry due to its attractive accuracy-to-computational cost ratio. DFT methods enable the efficient generation of extensive databases encompassing hundreds of thousands (or more) of molecular systems that are invaluable for training and testing ML models. Consequently, databases for training ML algorithms have been largely generated using DFT methods.[190,191,226–232] However, this scalability comes with a trade-off in accuracy compared to composite *ab initio* methods.[111,187,188] Needless to say, this limits the accuracy of ML models trained on these data. The accuracy of any ML algorithm is constrained by the quality of the data it is trained on. Therefore, care must be taken with the accuracy of the reference data in databases used for training ML models for applications demanding high precision. A clear example is the prediction of chemical kinetics, where even an ~1 kcal mol^{-1} variation in the reaction barrier height leads to an order-of-magnitude change in reaction rate at room temperature. This sensitivity to the accuracy of the reaction barrier height becomes even more critical for reactions at lower temperatures. Another example is the development of ML-based DFT methods designed to outperform conventional DFT methods. Achieving such enhanced performance is a primary driving force behind the development of ML-DFT methods, aiming to surpass the limitations of conventional DFT in both accuracy and computational efficiency.

Currently, the availability of thermochemical, kinetic and non-covalent interaction datasets with hundreds of thousands of chemical properties calculated by composite *ab initio* methods is an underdeveloped area. To the best of

our knowledge, the main such database is the recalculation of the 133,000 total atomisation energies in the GDB-9 database, originally calculated at the B3LYP/6-31G(2df,p) level of theory,[190] with the rung one composite *ab initio* method G4(MP2).[13] This important work by Curtiss and co-workers, demonstrates that with contemporary supercomputer hardware, it is possible to generate large thermochemical databases containing hundreds of thousands of systems with up to nine non-hydrogen atoms, with composite *ab initio* methods from rung one of Jacob's Ladder of composite *ab initio* methods (Fig. 1). Excluding multireference systems, these methods provide reliable thermochemical data, superior to that generated by DFT and DHDFT methods.[111] Continuation of the development of such databases will improve the accuracy of machine learning-based classical and quantum mechanical methods, such as ML-based force fields[233–236] and ML-DFT methods.[237–240] Nevertheless, an inherent limitation of highly accurate databases obtained by composite *ab initio* methods is that they will always be biased towards smaller (or fewer) systems compared to databases obtained from DFT calculations. Thus, it might be beneficial to augment such databases with thermochemical data obtained from DHDFT calculations that are applicable to larger systems and capable of achieving an intermediate accuracy between composite *ab initio* and conventional DFT methods.

When dealing with large and diverse chemical databases with tens or even hundreds of thousands of molecules generated by either composite *ab initio* or DFT methods, manual analysis of the structural and energetic data becomes impractical. Machine learning algorithms can efficiently analyse this vast data, for example, by identifying structure-function relationships and uncovering hidden patterns within the thermochemical data. This synergy between machine learning and large-scale thermochemical databases will ultimately lead to advances across various fields of chemistry, such as the design and optimisation of more selective and efficient catalysts and drugs with enhanced efficacy and fewer side effects. Nevertheless, we stress that the accuracy of the thermochemical data is critical to making reliable predictions and designing effective molecules and materials.

Conclusions

This review examines the synergy between experimental and computational thermochemistry. It highlights the historical role of experimental thermochemistry in developing highly accurate multilevel wavefunction theories (also known as composite *ab initio* methods) capable of near-kilocalories per mole to sub-kilojoules per mole thermochemical predictions. It continues to introduce the framework of Jacob's Ladder of composite *ab initio* methods, which categorises the large number of composite *ab initio* methods developed over the past three decades based on the treatment of the one-particle and *n*-particle spaces. Each rung of the ladder represents progressively more accurate and computationally demanding methods:

- **Rung 1** includes CCSD(T)/CBS(MP2) methods, which involve CCSD(T) calculations with a reasonably small basis set and MP2-based basis-set corrections calculated with larger basis sets. This rung includes methods such as G4(MP2), which are applicable to systems as large as C_{60}.
- **Rung 2** includes CCSD(T)/CBS methods that offer a more rigorous treatment of the one-particle space and extrapolate the HF, CCSD and (T) components individually to the complete basis set limit using reasonably large basis sets. This rung includes methods such as W1-F12, which are applicable to systems as large as dodecahedra $(CH)_{20}$.
- **Rung 3** offers a more rigorous treatment of the *n*-particle space by introducing higher-order triples and quasiperturbative quadruples excitations (i.e. methods that approximate the CCSDT(Q)/CBS energy). This rung includes methods such as W3-F12, which are applicable to systems as large as bullvalene $(CH)_{10}$.
- **Rung 4** represents the most accurate composite *ab initio* methods that approximate the FCI/CBS energy and are capable of confident sub-benchmark accuracy. These methods (e.g. W4-F12 theory) treat the one-particle and *n*-particle spaces with exceptional rigor but are limited to small molecules such as tetrahedrane $(CH)_4$.

This Jacob's Ladder framework facilitates better understanding and usage of the wide range of composite *ab initio* methods, which will assist researchers in selecting appropriate methods for different system sizes and levels of accuracy. We continue to explore how the landscape of thermochemical predictions has evolved due to the development of the above composite *ab initio* methods, which enable the creation of large theoretical thermochemical databases critical for DFT development and benchmarking. Finally, the growing synergy between quantum chemistry and machine learning is explored, with large chemical databases generated by composite *ab initio* methods representing the pinnacle of this approach. Machine learning models trained on large databases of highly accurate thermochemical data could revolutionise predictions for larger or more complex systems that are otherwise computationally prohibitive. These developments are set to accelerate advancements across various fields of chemistry, such as the design of novel materials, catalysts and bioactive molecules.

References

1 Afeefy HY, Liebman JF, Stein SE. *NIST Chemistry WebBook, SRD 69*. Linstrom PJ, Mallard WG, editors. Gaithersburg, MD, USA: National Institute of Standards and Technology. Available at http://webbook.nist.gov
2 Chase MW, Davies CA, Downey JR, Frurip DJ, McDonald RA, Syverud AN. JANAF thermochemical tables. *J Phys Chem Ref Data* 1985; 14(Suppl. 1).

3 Chase MW. NIST-JANAF thermochemical tables. *J Phys Chem Ref Data* 1998; 9.
4 Lias SG, Bartmess JE, Liebman JF, Holmes JL, Levin RD, Mallard WG. Gas-phase ion and neutral thermochemistry. *J Phys Chem Ref Data* 1988; 17(Suppl. 1): 861.
5 Rossini FD, Wagman DD, Evans WH, Levine S, Jaffe I, *Selected values of chemical thermodynamic properties*. Circular of the Bureau of Standards Number 500. Washington, DC, USA: US Government Printing Office; 1952.
6 Cox JD, Wagman DD, Medvedev VA. *CODATA Key Values for Thermodynamics*. New York, NY, USA: Hemisphere Publishing Corp.; 1989. Available at http://www.codata.org/resources/databases/key1.html
7 Gurvich LV, Veyts IV, Alcock CB. *Thermodynamic Properties of Individual Substances. Vol. 1, O, H(D, T), F, Cl, Br, I, He, Ne, Ar, Kr, Xe, Rn, S, N, P, and Their Compounds*, 4th edn. New York, NY, USA: Hemisphere Publishing Corp.; 1989.
8 Gurvich LV, Veyts IV, Alcock CB. *Thermodynamic Properties of Individual Substances. Vol. 2, C, Si, Ge, Sn, Pb, and Their Compounds*, 4th edn. New York, NY, USA: Hemisphere Publishing Corp.; 1989.
9 Gurvich LV, Veyts IV, Alcock CB. *Thermodynamic Properties of Individual Substances. Vol. 3, B, Al, Ga, In, Tl, Be, Mg, Ca, Sr, and Ba and Their Compounds*, 4th edn. New York, NY, USA: Hemisphere Publishing Corp.; 1989.
10 Pedley JB, Naylor RD, Kirby SP. *Thermochemical Data or Organic Compounds*. London, UK: Chapman and Hall; 1986.
11 Ruscic B, Pinzon RE, Morton ML, von Laszewski G, Bittner SJ, Nijsure SG, Amin KA, Minkoff M, Wagner AF. Introduction to active thermochemical tables: several "key" enthalpies of formation revisited. *J Phys Chem A* 2004; 108: 9979–9997. doi:10.1021/jp047912y
12 Ruscic B, Pinzon RE, von Laszewski G, Kodeboyina D, Burcat A, Leahy D, Montoya D, Wagner AF. Active thermochemical tables: thermochemistry for the 21st Century. *J Phys Conf Ser* 2005; 16: 561. doi:10.1088/1742-6596/16/1/078
13 Narayanan B, Redfern PC, Assary RS, Curtiss LA. Accurate quantum chemical energies for 133 000 organic molecules. *Chem Sci* 2019; 10: 7449–7455. doi:10.1039/C9SC02834J
14 Pople JA, Luke BT, Frisch MJ, Binkley JS. Theoretical thermochemistry. 1. Heats of formation of neutral AH_n molecules (A = Li to Cl). *J Phys Chem* 1985; 89: 2198–2203. doi:10.1021/j100257a013
15 Pople JA, Curtiss LA. Theoretical thermochemistry. 2. Ionization energies and proton affinities of AHn species (A = C to F and Si to Cl)—heats of formation of their cations. *J Phys Chem* 1987; 91: 155–162. doi:10.1021/j100285a035
16 Pople JA, Head-Gordon M, Fox DJ, Raghavachari K, Curtiss LA. Gaussian-1 theory: a general procedure for prediction of molecular energies. *J Chem Phys* 1989; 90: 5622–5629. doi:10.1063/1.456415
17 Curtiss LA, Jones C, Trucks GW, Raghavachari K, Pople JA. Gaussian-1 theory of molecular-energies for 2nd-row compounds. *J Chem Phys* 1990; 93: 2537–2545. doi:10.1063/1.458892
18 Curtiss LA, Raghavachari K, Trucks GW, Pople JA. Gaussian-2 theory for molecular-energies of 1st-row and 2nd-row compounds. *J Chem Phys* 1991; 94: 7221–7230. doi:10.1063/1.460205
19 Karton A. Quantum mechanical thermochemical predictions 100 years after the Schrödinger equation. In: Dixon DA, editor. *Annual Reports in Computational Chemistry*. Vol. 18. Elsevier; 2022. pp. 123–166.
20 Curtiss LA, Raghavachari K, Redfern PC, Rassolov V, Pople JA. Gaussian-3 (G3) theory for molecules containing first and second-row atoms. *J Chem Phys* 1998; 109: 7764–7776. doi:10.1063/1.477422
21 Raghavachari K. Autobiography of Krishnan Raghavachari. *J Phys Chem A* 2024; 128: 2526–2533. doi:10.1021/acs.jpca.4c01319
22 Curtiss LA, Redfern PC, Raghavachari K. G*n* theory. *WIREs Comput Mol Sci* 2011; 1: 810–825. doi:10.1002/wcms.59
23 Curtiss LA, Raghavachari K, Redfern PC, Pople JA. Investigation of the use of B3LYP zero-point energies and geometries in the calculation of enthalpies of formation. *Chem Phys Lett* 1997; 270: 419–426. doi:10.1016/S0009-2614(97)00399-0
24 Curtiss LA, Raghavachari K, Redfern PC, Baboul AG, Pople JA. Gaussian-3 theory using coupled cluster energies. *Chem Phys Lett* 1999; 314: 101–107. doi:10.1016/S0009-2614(99)01126-4
25 Curtiss LA, Redfern PC, Raghavachari K. Gaussian-4 theory. *J Chem Phys* 2007; 126: 084108. doi:10.1063/1.2436888
26 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: 4478–4484. doi:10.1021/acs.jctc.9b00449
27 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: 34–38. doi:10.1016/j.cplett.2015.11.009
28 Karton A, Waite SL, Page AJ. Performance of DFT for C_{60} isomerization energies: a noticeable exception to Jacob's Ladder. *J Phys Chem A* 2019; 123: 257–266. doi:10.1021/acs.jpca.8b10240
29 Henry DJ, Sullivan MB, Radom L. G3-RAD and G3X-RAD: modified Gaussian-3 (G3) and Gaussian-3X (G3X) procedures for radical thermochemistry. *J Chem Phys* 2003; 118: 4849–4860. doi:10.1063/1.1544731
30 Chan B, Coote ML, 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: 2647–2653. doi:10.1021/ct100266u
31 Chan B, Deng J, Radom L. G4(MP2)-6X: a cost-effective improvement to G4(MP2). *J Chem Theory Comput* 2011; 7: 112–120. doi:10.1021/ct100542x
32 Karton A, O'Reilly RJ, 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: 3128–3136. doi:10.1021/ct3004723
33 da Silva G. G3X-K theory: a composite theoretical method for thermochemical kinetics. *Chem Phys Lett* 2013; 558: 109–113. doi:10.1016/j.cplett.2012.12.045
34 Chan B, Karton A, Raghavachari K, Radom L. Restricted open-shell G4(MP2)-type procedures. *J Phys Chem A* 2016; 120: 9299–9304. doi:10.1021/acs.jpca.6b09361
35 Semidalas E, Martin JML. Canonical and DLPNO-based G4(MP2) XK-inspired composite wave function methods parametrized against large and chemically diverse training sets: are they more accurate and/or robust than double-hybrid DFT? *J Chem Theory Comput* 2020; 16: 4238–4255. doi:10.1021/acs.jctc.0c00189
36 Petersson GA, Bennett A, Tensfeldt TG, Al-Laham MA, Shirley WA, Mantzaris J. A complete basis set model chemistry. I. The total energies of closed-shell atoms and hydrides of the first-row elements. *J Chem Phys* 1988; 89: 2193–2218. doi:10.1063/1.455064
37 Petersson GA, Al-Laham MA. A complete basis set model chemistry. II. Open-shell systems and the total energies of the first-row atoms. *J Chem Phys* 1991; 94: 6081–6090. doi:10.1063/1.460447
38 Petersson GA, Bennett A, Tensfeldt TG. A complete basis set model chemistry. III. The complete basis set-quadratic configuration interaction family of methods. *J Chem Phys* 1991; 94: 6091–6101. doi:10.1063/1.460448
39 Montgomery JA, Ochterski JW, Petersson GA. A complete basis set model chemistry. IV. An improved atomic pair natural orbital method. *J Chem Phys* 1994; 101: 5900–5909. doi:10.1063/1.467306
40 Ochterski JW, Petersson GA, Montgomery JA. A complete basis set model chemistry. V. Extensions to six or more heavy atoms. *J Chem Phys* 1996; 104: 2598–2619. doi:10.1063/1.470985
41 Montgomery JA, Frisch MJ, Ochterski JW, Petersson GA. A complete basis set model chemistry. VI. Use of density functional geometries and frequencies. *J Chem Phys* 1999; 110: 2822–2827. doi:10.1063/1.477924
42 Wood GPF, Radom L, Petersson GA, Barnes EC, Frisch MJ, Montgomery JA. A restricted-open-shell complete-basis-set model chemistry. *J Chem Phys* 2006; 125: 094106. doi:10.1063/1.2335438
43 Allen WD, East ALL, Császár AG. *Ab initio* anharmonic vibrational analyses of non-rigid molecules. In: Laane J, Dakkouri M, van der Veken B, Oberhammer H, editors. *Structures and Conformations of Non-Rigid Molecules*. Dordrecht, Netherlands: Kluwer; 1993. pp. 343–373.

44 East ALL, Allen WD. The heat of formation of NCO. *J Chem Phys* 1993; 99: 4638–4650. doi:10.1063/1.466062
45 Klippenstein SJ, East ALL, Allen WD. A high level *ab initio* map and direct statistical treatment of the fragmentation of singlet ketene. *J Chem Phys* 1996; 105: 118–140. doi:10.1063/1.471858
46 Császár AG, Allen WD, Schaefer HF. In pursuit of the *ab initio* limit for conformational energy prototypes. *J Chem Phys* 1998; 108: 9751–9764. doi:10.1063/1.476449
47 Schuurman MS, Muir SR, Allen WD, Schaefer HF. 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: 11586–11599. doi:10.1063/1.1707013
48 Martin JML, de Oliveira G. Towards standard methods for benchmark quality *ab initio* thermochemistry—W1 and W2 theory. *J Chem Phys* 1999; 111: 1843–1856. doi:10.1063/1.479454
49 Boese AD, Oren M, Atasoylu O, Martin JML, Kallay M, Gauss J. W3 theory: robust computational thermochemistry in the kJ/mol accuracy range. *J Chem Phys* 2004; 120: 4129–4141. doi:10.1063/1.1638736
50 Karton A, Rabinovich E, Martin JML, Ruscic B. W4 theory for computational thermochemistry: in pursuit of confident sub-kJ/mol predictions. *J Chem Phys* 2006; 125: 144108. doi:10.1063/1.2348881
51 Karton A, Taylor PR, Martin JML. Basis set convergence of post-CCSD contributions to molecular atomization energies. *J Chem Phys* 2007; 127: 064104. doi:10.1063/1.2755751
52 Karton A, Martin JML. Explicitly correlated Wn theory: W1-F12 and W2-F12. *J Chem Phys* 2012; 136: 124114. doi:10.1063/1.3697678
53 Sylvetsky N, Peterson KA, Karton A, Martin JML. Toward a W4-F12 approach: Can explicitly correlated and orbital-based *ab initio* CCSD(T) limits be reconciled? *J Chem Phys* 2016; 144: 214101. doi:10.1063/1.4952410
54 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: 4259–4269. doi:10.1021/ct300632p
55 Chan B, Radom L. W3X: a cost-effective post-CCSD (T) composite procedure. *J Chem Theory Comput* 2013; 9: 4769–4778. doi:10.1021/ct4005323
56 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: 2109–2119. doi:10.1021/acs.jctc.5b00135
57 Fast PL, Sanchez ML, Truhlar DG. Multi-coefficient Gaussian-3 method for calculating potential energy surfaces. *Chem Phys Lett* 1999; 306: 407–410. doi:10.1016/S0009-2614(99)00493-5
58 Fast PL, Corchado JC, Sanchez ML, Truhlar DG. Multi-coefficient correlation method for quantum chemistry. *J Phys Chem* 1999; 103: 5129–5136. doi:10.1021/jp9903460
59 Fast PL, Truhlar DG. MC-QCISD: multi-coefficient correlation method based on quadratic configuration interaction with single and double excitations. *J Phys Chem A* 2000; 104: 6111–6116. doi:10.1021/jp000408i
60 Lynch BJ, Truhlar DG. Robust and affordable multicoefficient methods for thermochemistry and thermochemical kinetics: the MCCM/3 suite and SAC/3. *J Phys Chem A* 2003; 107: 3898–3906. doi:10.1021/jp0221993
61 Lynch BJ, Zhao Y, Truhlar DG. The 6-31B(d) basis set and the BMC-QCISD and BMC-CCSD multicoefficient correlation methods. *J Phys Chem A* 2005; 109: 1643–1649. doi:10.1021/jp045847m
62 Zhang W, Kong X, Liu S, Zhao Y. Multi-coefficients correlation methods. *WIREs Comput Mol Sci* 2020; 10: e1474. doi:10.1002/wcms.1474
63 Tajti A, Szalay PG, Császár AG, Kállay M, Gauss J, Valeev EF, Flowers BA, Vázquez J, Stanton JF. HEAT: high accuracy extrapolated *ab initio* thermochemistry. *J Chem Phys* 2004; 121: 11599–11613. doi:10.1063/1.1811608
64 Bomble YJ, Vázquez J, Kállay M, Michauk C, Szalay PG, Császár AG, Gauss J, Stanton JF. High-accuracy extrapolated *ab initio* thermochemistry. II. Minor improvements to the protocol and a vital simplification. *J Chem Phys* 2006; 125: 064108. doi:10.1063/1.2206789
65 Harding ME, Vázquez J, Ruscic B, Wilson AK, Gauss J, Stanton JF. High-accuracy extrapolated *ab initio* thermochemistry. III. Additional improvements and overview. *J Chem Phys* 2008; 128: 114111. doi:10.1063/1.2835612
66 Thorpe JH, Lopez CA, Nguyen TL, Baraban JH, Bross DH, Ruscic B, Stanton JF. High-accuracy extrapolated *ab initio* thermochemistry. IV. A modified recipe for computational efficiency. *J Chem Phys* 2019; 150: 224102. doi:10.1063/1.5095937
67 Thorpe JH, Kilburn JL, Feller D, Changala PB, Bross DH, Ruscic B, Stanton JF. Elaborated thermochemical treatment of HF, CO, N_2, and H_2O: Insight into HEAT and its extensions. *J Chem Phys* 2021; 155: 184109. doi:10.1063/5.0069322
68 DeYonker NJ, Cundari TR, Wilson AK. The correlation consistent composite approach (ccCA): an alternative to the Gaussian-n methods. *J Chem Phys* 2006; 124: 114104. doi:10.1063/1.2173988
69 DeYonker NJ, Peterson KA, Steyl G, Wilson AK, Cundari TR. Quantitative computational thermochemistry of transition metal species. *J Phys Chem A* 2007; 111: 11269–11277. doi:10.1021/jp0715023
70 DeYonker NJ, Williams TG, Imel AE, Cundari TR, Wilson AK. Accurate thermochemistry for transition metal complexes from first-principles calculations. *J Chem Phys* 2009; 131: 024106. doi:10.1063/1.3160667
71 Mintz B, Williams TG, Howard L, Wilson AK. Computation of potential energy surfaces with the multireference correlation consistent composite approach. *J Chem Phys* 2009; 130: 234104. doi:10.1063/1.3149387
72 DeYonker NJ, Wilson BR, Pierpont AW, Cundari TR, Wilson AK. Towards the intrinsic error of the correlation consistent composite approach (ccCA). *Mol Phys* 2009; 107: 1107–1121. doi:10.1080/00268970902744359
73 Prascher BP, Lai JD, Wilson AK. The resolution of the identity approximation applied to the correlation consistent composite approach. *J Chem Phys* 2009; 131: 044130. doi:10.1063/1.3176507
74 Laury ML, DeYonker NJ, Jiang W, Wilson AK. 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. doi:10.1063/1.3662415
75 Laury ML, Wilson AK. Examining the heavy p-block with a pseudopotential-based composite method: atomic and molecular applications of rp-ccCA. *J Chem Phys* 2012; 137: 214111. doi:10.1063/1.4768420
76 Mahler A, Wilson AK. Explicitly correlated methods within the ccCA methodology. *J Chem Theory Comput* 2013; 9: 1402–1407. doi:10.1021/ct300956e
77 Welch BK, Almeida NMS, Wilson AK. Super ccCA (s-ccCA): an approach for accurate transition metal thermochemistry. *Mol Phys* 2021; 119: e1963001. doi:10.1080/00268976.2021.1963001
78 Dixon DA, Feller D, Peterson KA. A practical guide to reliable first principles computational thermochemistry predictions across the periodic table. In: Wheeler RA, editor. *Annual Reports in Computational Chemistry*. Vol. 8. Elsevier; 2012. pp. 1–28.
79 Feller D, Peterson KA, Dixon D. The impact of larger basis sets and explicitly correlated coupled cluster theory on the Feller–Peterson–Dixon Composite Method. In: Dixon DA, editor. *Annual Reports in Computational Chemistry*. Vol. 12. Elsevier; 2016. pp. 47–78.
80 Feller D, Peterson KA, Dixon DA. Refined theoretical estimates of the atomization energies and molecular structures of selected small oxygen fluorides. *J Phys Chem A* 2010; 114: 613–623. doi:10.1021/jp908128g
81 Feller D, Peterson KA, Dixon DA. *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: 1440–1451. doi:10.1021/jp111644h
82 Feller D, Peterson KA, Dixon DA. Further benchmarks of a composite, convergent, statistically calibrated coupled cluster-based approach for thermochemical and spectroscopic studies. *Mol Phys* 2012; 110: 2381–2399. doi:10.1080/00268976.2012.684897

83 Feller D, Peterson KA, Ruscic B. Improved accuracy benchmarks of small molecules using correlation consistent basis sets. *Theor Chem Acc* 2014; 133: 1407. doi:10.1007/s00214-013-1407-z

84 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. doi:10.1063/1.4993625

85 Bakowies D. *Ab initio* thermochemistry using optimal-balance models with isodesmic corrections: the ATOMIC protocol. *J Chem Phys* 2009; 130: 144113. doi:10.1063/1.3089241

86 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: 5230–5251. doi:10.1021/acs.jctc.9b00343

87 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: 399–426. doi:10.1021/acs.jctc.9b00974

88 Vogiatzis KD, Haunschild R, Klopper W. Accurate atomization energies from combining coupled-cluster computations with interference-corrected explicitly correlated second-order perturbation theory. *Theor Chem Acc* 2014; 133: 1446. doi:10.1007/s00214-014-1446-0

89 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: 988–1006. doi:10.1021/acs.jctc.9b01037

90 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: 6974–6992. doi:10.1021/acs.jctc.1c00869

91 Karton A. Benchmark accuracy in thermochemistry, kinetics, and noncovalent interactions. In: Boyd RJ, Yanez M, editors. *Comprehensive Computational Chemistry*. Vol. 1, 1st edn. Elsevier; 2023. pp. 47–68.

92 Karton A. A computational chemist's guide to accurate thermochemistry for organic molecules. *WIREs Comput Mol Sci* 2016; 6: 292–310. doi:10.1002/wcms.1249

93 Patel P, Melin TRL, North SC, Wilson AK. *Ab initio* composite methodologies: their significance for the chemistry community. In: Dixon DA, editor. *Annual Reports in Computational Chemistry*. Vol. 17. Elsevier; 2021. pp. 113–161.

94 Chan B. How to computationally calculate thermochemical properties objectively, accurately, and as economically as possible. *Pure Appl Chem* 2017; 89: 699–713. doi:10.1515/pac-2016-1116

95 Peterson C, Penchoff DA, Wilson AK. Prediction of thermochemical properties across the periodic table: a review of the correlation consistent composite approach (ccCA) strategies and applications. In: Dixon DA, editor. *Annual Reports in Computational Chemistry*. Vol. 12. Elsevier; 2016. pp. 3–45.

96 Feller D, Peterson KA, Dixon DA. A survey of factors contributing to accurate theoretical predictions of atomization energies and molecular structures. *J Chem Phys* 2008; 129: 204105. doi:10.1063/1.3008061

97 Peterson KA, Feller D, Dixon DA. Chemical accuracy in *ab initio* thermochemistry and spectroscopy: current strategies and future challenges. *Theor Chem Acc* 2012; 131: 1079. doi:10.1007/s00214-011-1079-5

98 Jiang W, Wilson AK. *Ab initio* composite approaches: potential energy surfaces and excited electronic states. In: Wheeler RA, editor. *Annual Reports in Computational Chemistry*. Vol. 8. Elsevier; 2012. pp. 29–51.

99 Klopper W, Bachorz RA, Hättig C, Tew DP. Accurate computational thermochemistry from explicitly correlated coupled-cluster theory. *Theor Chem Acc* 2010; 126: 289–304. doi:10.1007/s00214-010-0733-7

100 DeYonker N, Cundari TR, Wilson AK. The correlation consistent composite approach (ccCA): efficient and pan-periodic kinetics and thermodynamics. In: Piecuch P, Maruani J, Delgado-Barrio G, Wilson S, editors. *Advances in the Theory of Atomic and Molecular Systems (Progress in Theoretical Chemistry and Physics)*. Vol. 19. Dordrecht, Netherlands: Springer; 2009. pp. 197–224.

101 Helgaker T, Klopper W, Tew DP. Quantitative quantum chemistry. *Mol Phys* 2008; 106: 2107. doi:10.1080/00268970802258591

102 Helgaker T, Klopper W, Bak KL, Halkier A, Jørgensen P, Olsen J. Highly accurate *ab initio* computation of thermochemical data. In: Cioslowski J, editor. *Understanding Chemical Reactivity, Vol. 22: Quantum–Mechanical Prediction of Thermochemical Data*. Dordrecht, Netherlands: Kluwer; 2001. pp. 1–30.

103 Martin JML. Computational thermochemistry: a brief overview of quantum mechanical approaches. In: *Annual Reports in Computational Chemistry*. Vol. 1. Elsevier; 2005. pp. 31–43.

104 Martin JML, Parthiban S. W1 and W2 theory and their variants: thermochemistry in the kJ/mol accuracy range. In: Cioslowski J, editor. *Understanding Chemical Reactivity, Vol. 22: Quantum–Mechanical Prediction of Thermochemical Data*. Dordrecht, Netherlands: Kluwer; 2001. pp. 31–65.

105 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: 622–632. doi:10.1002/jcc.23837

106 Yu LJ, Sarrami F, O'Reilly RJ, Karton A. Reaction barrier heights for cycloreversion of heterocyclic rings: an Achilles' heel for DFT and standard *ab initio* procedures. *Chem Phys* 2015; 458: 1–8. doi:10.1016/j.chemphys.2015.07.005

107 Karton A, Martin JML. Prototypical π–π dimers re-examined by means of high-level CCSDT(Q) composite *ab initio* methods. *J Chem Phys* 2021; 154: 124117. doi:10.1063/5.0043046

108 Curtiss LA, Redfern PC, Raghavachari K. Gaussian-4 theory using reduced order perturbation theory. *J Chem Phys* 2007; 127: 124105. doi:10.1063/1.2770701

109 Karton A. Fullerenes pose a strain on hybrid density functional theory. *J Phys Chem A* 2022; 126: 4709–4720. doi:10.1021/acs.jpca.2c02246

110 Karton A, Chan B. Performance of local G4(MP2) composite *ab initio* procedures for fullerene isomerization energies. *Comput Theor Chem* 2022; 1217: 113874. doi:10.1016/j.comptc.2022.113874

111 Karton A. Big data benchmarking: how do DFT methods across the rungs of Jacob's Ladder perform for a dataset of 122k CCSD(T) total atomization energies? *Phys Chem Chem Phys* 2024; 26: 14594–14606. doi:10.1039/D4CP00387J

112 Dunning TH. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J Chem Phys* 1989; 90: 1007–1023. doi:10.1063/1.456153

113 Kendall RA, Dunning TH, Harrison RJ. Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions. *J Chem Phys* 1992; 96: 6796–6806. doi:10.1063/1.462569

114 Dunning TH, Peterson KA, Wilson AK. Gaussian basis sets for use in correlated molecular calculations. X. The atoms aluminum through argon revisited. *J Chem Phys* 2001; 114: 9244–9253. doi:10.1063/1.1367373

115 Manaa MR, Fried LE, Kuo I-FW. Determination of enthalpies of formation of energetic molecules with composite quantum chemical methods. *Chem Phys Lett* 2016; 648: 31–35. doi:10.1016/j.cplett.2016.01.071

116 Fogueri UR, Kozuch S, Karton A, Martin JML. 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: 2269–2277. doi:10.1021/jp312644t

117 Jorgensen KR, Oyedepo GA, Wilson AK. Highly energetic nitrogen species: reliable energetics via the correlation consistent composite approach (ccCA). *J Hazard Mater* 2011; 186: 583–589. doi:10.1016/j.jhazmat.2010.11.035

118 Barnes EC, Petersson GA, Montgomery JAJr, Frisch MJ, Martin JML. Unrestricted coupled cluster and brueckner doubles variations of W1 theory. *J Chem Theory Comput* 2009; 5: 2687–2693. doi:10.1021/ct900260g

119 Peterson KA, Adler TB, Werner H-J. Systematically convergent basis sets for explicitly correlated wavefunctions: the atoms H, He, B–Ne, and Al–Ar. *J Chem Phys* 2008; 128: 084102. doi:10.1063/1.2831537

120 Papajak E, Truhlar DG. Convergent partially augmented basis sets for post-Hartree–Fock calculations of molecular properties

and reaction barrier heights. *J Chem Theory Comput* 2011; 7: 10–18. doi:10.1021/ct1005533
121 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* 2013; 117: 1834–1842. doi:10.1021/jp312585r
122 Karton A, Schreiner PR, Martin JML. Heats of formation of platonic hydrocarbon cages by means of high-level thermochemical procedures. *J Comput Chem* 2016; 37: 49–58. doi:10.1002/jcc.23963
123 Karton A, Chan B. Accurate heats of formation for polycyclic aromatic hydrocarbons: a high-level *ab initio* perspective. *J Chem Eng Data* 2021; 66: 3453–3462. doi:10.1021/acs.jced.1c00256
124 Karton A, Yu L-J, Kesharwani MK, Martin JML. Heats of formation of the amino acids re-examined by means of W1-F12 and W2-F12 theories. *Theor Chem Acc* 2014; 133: 1483. doi:10.1007/s00214-014-1483-8
125 Karton A, Kaminker I, Martin JML. Economical post-CCSD(T) computational thermochemistry protocol and applications to some aromatic compounds. *J Phys Chem A* 2009; 113: 7610–7620. doi:10.1021/jp900056w
126 Karton A. Cope rearrangements in shapeshifting molecules re-examined by means of high-level CCSDT(Q) composite *ab initio* methods. *Chem Phys Lett* 2020; 759: 138018. doi:10.1016/j.cplett.2020.138018
127 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. doi:10.1016/j.chphi.2021.100047
128 Kroeger AA, Karton A. Thermochemistry of phosphorus sulfide cages: an extreme challenge for high-level *ab initio* methods. *Struct Chem* 2019; 30: 1665–1675. doi:10.1007/s11224-019-01352-7
129 Karton A, Sylvetsky N, Martin JML. W4-17: a diverse and high-confidence dataset of atomization energies for benchmarking high-level electronic structure methods. *J Comput Chem* 2017; 38: 2063–2075. doi:10.1002/jcc.24854
130 Karton A, Daon S, Martin JML. W4-11: a high-confidence benchmark dataset for computational thermochemistry derived from first-principles W4 data. *Chem Phys Lett* 2011; 510: 165–178. doi:10.1016/j.cplett.2011.05.007
131 Martin JML. Electron correlation: nature's weird and wonderful chemical glue. *Isr J Chem* 2022; 62: e202100111. doi:10.1002/ijch.202100111
132 Boese AD, Klopper W, Martin JML. Anharmonic force fields and thermodynamic functions using density functional theory. *Mol Phys* 2005; 103: 863–876. doi:10.1080/00268970512331339369
133 Boese AD, Martin JML. Vibrational spectra of the azabenzenes revisited: anharmonic force fields. *J Phys Chem A* 2004; 108: 3085–3096. doi:10.1021/jp0369589
134 Karton A, Martin JML. The lowest singlet-triplet excitation energy of BN: a converged coupled cluster perspective. *J Chem Phys* 2006; 125: 144313. doi:10.1063/1.2352752
135 Karton A, Tarnopolsky A, Martin JML. Atomization energies of the carbon clusters Cn (n = 2–10) revisited by means of W4 theory as well as density functional, Gn, and CBS methods. *Mol Phys* 2009; 107: 977–990. doi:10.1080/00268970802708959
136 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. doi:10.1016/j.cplett.2019.136810
137 Karton A. Post-CCSD(T) contributions to total atomization energies in multireference systems. *J Chem Phys* 2018; 149: 034102. doi:10.1063/1.5036795
138 Ruscic B. Uncertainty quantification in thermochemistry, benchmarking electronic structure computations, and active thermochemical tables. *Int J Quantum Chem* 2014; 114: 1097–1101. doi:10.1002/qua.24605
139 Ruscic B, Bross DH. Thermochemistry. In: Faravelli T, Manenti F, Ranzi E, editors. *Computer Aided Chemical Engineering*. Vol. 45. Elsevier; 2019. pp. 3–114.
140 Wheeler SE. Homodesmotic reactions for thermochemistry. *WIREs Comput Mol Sci* 2012; 2: 204–220. doi:10.1002/wcms.72
141 Chan B, Collins E, Raghavachari K. Applications of isodesmic-type reactions for computational thermochemistry. *WIREs Comput Mol Sci* 2021; 11: e1501. doi:10.1002/wcms.1501
142 Wheeler SE, Houk KN, Schleyer PvR, Allen WD. A hierarchy of homodesmotic reactions for thermochemistry. *J Am Chem Soc* 2009; 131: 2547–2560. doi:10.1021/ja805843n
143 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: 166–177. doi:10.1016/j.chemphys.2014.07.015
144 Karton A. How reliable is DFT in predicting relative energies of polycyclic aromatic hydrocarbon isomers? Comparison of functionals from different rungs of Jacob's Ladder. *J Comput Chem* 2017; 38: 370–382. doi:10.1002/jcc.24669
145 Karton A, Chan B. PAH335 – a diverse database of highly accurate CCSD(T) isomerization energies of 335 polycyclic aromatic hydrocarbons. *Chem Phys Lett* 2023; 824: 140544. doi:10.1016/j.cplett.2023.140544
146 Ramabhadran RO, Raghavachari K. Theoretical thermochemistry for organic molecules: development of the generalized connectivity-based hierarchy. *J Chem Theory Comput* 2011; 7: 2094–2103. doi:10.1021/ct200279q
147 Ramabhadran RO, Raghavachari K. Connectivity-based hierarchy for theoretical thermochemistry: assessment using wave function-based methods. *J Phys Chem A* 2012; 116: 7531–7537. doi:10.1021/jp301421a
148 Ramabhadran RO, Raghavachari K. The successful merger of theoretical thermochemistry with fragment-based methods in quantum chemistry. *Acc Chem Res* 2014; 47: 3596–3604. doi:10.1021/ar500294s
149 Boese AD, Martin JML. Vibrational spectra of the azabenzenes revisited: anharmonic force fields. *J Phys Chem A* 2004; 108: 3085–3096. doi:10.1021/jp0369589
150 Heckert M, Kállay M, Tew DP, Klopper W, Gauss J. Basis-set extrapolation techniques for the accurate calculation of molecular equilibrium geometries using coupled-cluster theory. *J Chem Phys* 2006; 125: 44108. doi:10.1063/1.2217732
151 Tew DP, Klopper W, Heckert M, Gauss J. Basis set limit CCSD(T) harmonic vibrational frequencies. *J Phys Chem A* 2007; 111: 11242–11248. doi:10.1021/jp070851u
152 Puzzarini C, Heckert M, Gauss J. The accuracy of rotational constants predicted by high-level quantum-chemical calculations. I. Molecules containing first-row atoms. *J Chem Phys* 2008; 128: 194108. doi:10.1063/1.2912941
153 Puzzarini C. Extrapolation to the complete basis set limit of structural parameters: comparison of different approaches. *J Phys Chem A* 2009; 113: 14530–14535. doi:10.1021/jp9040327
154 Karton A, Martin JML. Performance of W4 theory for spectroscopic constants and electrical properties of small molecules. *J Chem Phys* 2010; 133: 144102. doi:10.1063/1.3489113
155 Puzzarini C, Barone V. Extending the molecular size in accurate quantum-chemical calculations: the equilibrium structure and spectroscopic properties of uracil. *Phys Chem Chem Phys* 2011; 13: 7189–7197. doi:10.1039/c0cp02636k
156 Puzzarini C, Stanton JF. Connections between the accuracy of rotational constants and equilibrium molecular structures. *Phys Chem Chem Phys* 2023; 25: 1421–1429. doi:10.1039/D2CP04706C
157 Franke PR, Stanton JF. Rotamers of methanediol: composite *ab initio* predictions of structures, frequencies, and rovibrational constants. *J Phys Chem A* 2023; 127: 924–937. doi:10.1021/acs.jpca.2c06686
158 Spiegel M, Semidalas E, Martin JML, Bentley MR, Stanton JF. Post-CCSD(T) corrections to bond distances and vibrational frequencies: the power of Λ. *Mol Phys* 2023; 122: e2252114. doi:10.1080/00268976.2023.2252114
159 Christiansen O, Coriani S, Gauss J, Hattig C, Jorgensen P, Pawlowski F, Rizzo A. Accurate nonlinear optical properties for small molecules. In: Papadopoulos MG, Sadlej AJ, Leszczynski J, editors. *Non-linear optical properties of matter: from molecules to condensed phases*. Dordrecht, Netherlands: Springer; 2006. pp. 51–99. doi:10.1007/1-4020-4850-5_2
160 Curtiss LA, Raghavachari K, Redfern PC, Pople JA. Assessment of Gaussian-2 and density functional theories for the computation

of enthalpies of formation. *J Chem Phys* 1997; 106: 1063–1079. doi:10.1063/1.473182
161 Curtiss LA, Raghavachari K, Redfern PC, Pople JA. Assessment of Gaussian-3 and density functional theories for a larger experimental test set. *J Chem Phys* 2000; 112: 7374–7383. doi:10.1063/1.481336
162 Curtiss LA, Redfern PC, Raghavachari K. Assessment of Gaussian-3 and density-functional theories on the G3/05 test set of experimental energies. *J Chem Phys* 2005; 123: 124107. doi:10.1063/1.2039080
163 Karton A, de Oliveira MT. Good practices in database generation for benchmarking DFT. *WIREs Comput Mol Sci* 2024; 15: e1737. doi:10.1002/wcms.1737
164 Karton A, Gruzman D, Martin JML. Benchmark thermochemistry of the CnH2n + 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: 8434–8447. doi:10.1021/jp904369h
165 Karton A, Martin JML. Explicitly correlated benchmark calculations on C_8H_8 isomer energy separations: how accurate are DFT, double-hybrid and composite *ab initio* procedures? *Mol Phys* 2012; 110: 2477–2491. doi:10.1080/00268976.2012.698316
166 Gruzman D, Karton A, Martin JML. Performance of *ab initio* and density functional methods for conformational equilibria of C_nH_{2n+2} alkane isomers (n = 4–8). *J Phys Chem A* 2009; 113: 11974–11983. doi:10.1021/jp903640h
167 Fogueri UR, Kozuch S, Karton A, Martin JML. The melatonin conformer space: benchmark and assessment of wave function and DFT methods for a paradigmatic biological and pharmacological molecule. *J Phys Chem A* 2013; 117: 2269–2277. doi:10.1021/jp312644t
168 Reha D, Valdés H, Vondrásek J, Hobza P, Abu-Riziq A, Crews B, de Vries MS. Structure and IR spectrum of phenylalanyl-glycyl-glycine tripetide in the gas-phase: IR/UV experiments, *ab initio* quantum chemical calculations, and molecular dynamic simulations. *Chem Eur J* 2005; 11: 6803–6817. doi:10.1002/chem.200500465
169 Csonka GI, French AD, Johnson GP, Stortz CA. Evaluation of density functionals and basis sets for carbohydrates. *J Chem Theory Comput* 2009; 5: 679–692. doi:10.1021/ct8004479
170 Wilke JJ, Lind MC, Schaefer HF, Császár AG, Allen WD. Conformers of gaseous cysteine. *J Chem Theory Comput* 2009; 5: 1511–1523. doi:10.1021/ct900005c
171 Johnson ER, Mori-Sánchez P, Cohen AJ, Yang W. Delocalization errors in density functionals and implications for main-group thermochemistry. *J Chem Phys* 2008; 129: 204112. doi:10.1063/1.3021474
172 Krieg H, Grimme S. Thermochemical benchmarking of hydrocarbon bond separation reaction energies: Jacob's Ladder is not reversed! *Mol Phys* 2010; 108: 2655–2666. doi:10.1080/00268976.2010.519729
173 Korth M, Grimme S. "Mindless" DFT benchmarking. *J Chem Theory Comput* 2009; 5: 993–1003. doi:10.1021/ct800511q
174 Karton A. Highly accurate CCSDT(Q)/CBS reaction barrier heights for a diverse set of transition structures: basis set convergence and cost-effective approaches for estimating post-CCSD (T) contributions. *J Phys Chem A* 2019; 123: 6720–6732. doi:10.1021/acs.jpca.9b04611
175 Grimme S, Ehrlich S, Goerigk L. Effect of the damping function in dispersion corrected density functional theory. *J Comput Chem* 2011; 32: 1456–1465. doi:10.1002/jcc.21759
176 Guner V, Khuong KS, Leach AG, Lee PS, Bartberger MD, Houk KN. A standard set of pericyclic reactions of hydrocarbons for the benchmarking of computational methods: the performance of *ab initio*, density functional, CASSCF, CASPT2, and CBS-QB3 methods for the prediction of activation barriers, reaction energetics, and transition state geometries. *J Phys Chem A* 2003; 107: 11445–11459. doi:10.1021/jp035501w
177 Ess DH, Houk KN. Activation energies of pericyclic reactions: performance of DFT, MP2, and CBS-QB3 methods for the prediction of activation barriers and reaction energetics of 1,3-dipolar cycloadditions, and revised activation enthalpies for a standard set of hydrocarbon pericyclic reactions. *J Phys Chem A* 2005; 109: 9542–9553. doi:10.1021/jp052504v
178 Grimme S, Mück-Lichtenfeld C, Würthwein E-U, Ehlers AW, Goumans TPM, Lammertsma K. Consistent theoretical description of 1,3-dipolar cycloaddition reactions. *J Phys Chem A* 2006; 110: 2583–2586. doi:10.1021/jp057329x
179 Dinadayalane TC, Vijaya R, Smitha A, Sastry GN. Diels–Alder reactivity of butadiene and cyclic five-membered dienes ($(CH)_4X$, X = CH_2, SiH_2, O, NH, PH, and S) with ethylene: a benchmark study. *J Phys Chem A* 2002; 106: 1627–1633. doi:10.1021/jp013910r
180 Zhao Y, Lynch BJ, Truhlar DG. Development and assessment of a new hybrid density functional model for thermochemical kinetics. *J Phys Chem A* 2004; 108: 2715–2719. doi:10.1021/jp049908s
181 Zhao Y, González-García N, Truhlar DG. Benchmark database of barrier heights for heavy atom transfer, nucleophilic substitution, association, and unimolecular reactions and its use to test theoretical methods. *J Phys Chem A* 2005; 109: 2012–2018. doi:10.1021/jp045141s
182 Goerigk L, Grimme S. A general database for main group thermochemistry, kinetics, and noncovalent interactions – assessment of common and reparameterized (meta-)GGA density functionals. *J Chem Theory Comput* 2010; 6: 107–126. doi:10.1021/ct900489g
183 Neese F, Schwabe T, Kossmann S, Schirmer B, Grimme S. Assessment of orbital-optimized, spin-component scaled second-order many-body perturbation theory for thermochemistry and kinetics. *J Chem Theory Comput* 2009; 5: 3060–3073. doi:10.1021/ct9003299
184 Izgorodina EI, Coote ML, Radom L. Trends in R–X bond dissociation energies (R = Me, Et, *i*-Pr, *t*-Bu; X = H, CH_3, OCH_3, OH, F): a surprising shortcoming of density functional. *J Phys Chem A* 2005; 109: 7558–7566. doi:10.1021/jp052021r
185 Coote ML, Lin CY, Beckwith ALJ, Zavitsas AA. A comparison of methods for measuring relative radical stabilities of carbon-centred radicals. *Phys Chem Chem Phys* 2010; 12: 9597–9610. doi:10.1039/c003880f
186 Goerigk L, Grimme S. A thorough benchmark of density functional methods for general main group thermochemistry, kinetics, and noncovalent interactions. *Phys Chem Chem Phys* 2011; 13: 6670–6688. doi:10.1039/c0cp02984j
187 Goerigk L, Hansen A, Bauer C, Ehrlich S, Najibi A, Grimme S. A look at the density functional theory zoo with the advanced GMTKN55 database for general main group thermochemistry, kinetics and noncovalent interactions. *Phys Chem Chem Phys* 2017; 19: 32184–32215. doi:10.1039/C7CP04913G
188 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: 2315–2372. doi:10.1080/00268976.2017.1333644
189 Goerigk L. Benchmarking modern density functionals for broad applications in chemistry. In: Boyd RJ, Yanez M, editors. *Comprehensive Computational Chemistry*. Vol. 1, 1st edn. Elsevier; 2023. pp. 78–93.
190 Ramakrishnan R, Dral PO, Rupp M, von Lilienfeld OA. Quantum chemistry structures and properties of 134 kilo molecules. *Sci Data* 2014; 1: 140022. doi:10.1038/sdata.2014.22
191 Huang B, von Lilienfeld OA. *Ab initio* machine learning in chemical compound space. *Chem Rev* 2021; 121: 10001–10036. doi:10.1021/acs.chemrev.0c01303
192 Manna D, Martin JML. What are the ground state structures of C_{20} and C_{24}? An explicitly correlated *ab initio* approach. *J Phys Chem A* 2016; 120: 153–160. doi:10.1021/acs.jpca.5b10266
193 Feyereisen MW, Fitzgerald G, Kormornicki A. Use of approximate integrals in *ab initio* theory. An application in MP2 energy calculations. *Chem Phys Lett* 1993; 208: 359–363. doi:10.1016/0009-2614(93)87156-W
194 Vahtras O, Almlof J, Feyereisen MW. Integral approximations for LCAO-SCF calculations. *Chem Phys Lett* 1993; 213: 514–518. doi:10.1016/0009-2614(93)89151-7
195 Kendall R, Fruchtl HA. The impact of the resolution of the identity approximate integral method on modern *ab initio* algorithm development. *Theor Chim Acta* 1997; 97: 158–163. doi:10.1007/s002140050249

196 Weigend F, Haser M, Patzelt H, Ahlrichs R. RI-MP2: optimized auxiliary basis sets and demonstration of efficiency. *Chem Phys Lett* 1998; 294: 143–152. doi:10.1016/S0009-2614(98)00862-8
197 Klopper W. Highly accurate coupled-cluster singlet and triplet pair energies from explicitly correlated calculations in comparison with extrapolation techniques. *Mol Phys* 2001; 99: 481–507. doi:10.1080/00268970010017315
198 Ten-no S. Initiation of explicitly correlated Slater-type geminal theory. *Chem Phys Lett* 2004; 398: 56–61. doi:10.1016/j.cplett.2004.09.041
199 Klopper W, Manby FR, Ten-no S, Valeev EF. R12 methods in explicitly correlated molecular electronic structure theory. *Int Rev Phys Chem* 2006; 25: 427–468. doi:10.1080/01442350600799921
200 Werner H-J, Adler TB, Manby FR. General orbital invariant MP2-F12 theory. *J Chem Phys* 2007; 126: 164102. doi:10.1063/1.2712434
201 Ten-no S, Noga J. Explicitly correlated electronic structure theory from R12/F12 ansätze. *WIREs Comput Mol Sci* 2012; 2: 114–125. doi:10.1002/wcms.68
202 Ma Q, Werner H-J. Explicitly correlated local coupled-cluster methods using pair natural orbitals. *WIREs Comput Mol Sci* 2018; 8: e1371. doi:10.1002/wcms.1371
203 Riplinger C, Neese F. An efficient and near linear scaling pair natural orbital based local coupled cluster method. *J Chem Phys* 2013; 138: 034106. doi:10.1063/1.4773581
204 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. doi:10.1063/1.4821834
205 Nagy PR, Kállay M. Optimization of the linear-scaling local natural orbital CCSD(T) method: redundancy-free triples correction using Laplace transform. *J Chem Phys* 2017; 146: 214106. doi:10.1063/1.4984322
206 Nagy PR, Samu G, Kállay M. Optimization of the linear-scaling local natural orbital CCSD(T) method: improved algorithm and benchmark applications. *J Chem Theory Comput* 2018; 14: 4193–4215. doi:10.1021/acs.jctc.8b00442
207 Nagy PR, Kállay M. Approaching the basis set limit of CCSD(T) energies for large molecules with local natural orbital coupled-cluster methods. *J Chem Theory Comput* 2019; 15: 5275–5298. doi:10.1021/acs.jctc.9b00511
208 Liakos DG, Sparta M, Kesharwani MK, Martin JML, Neese F. Exploring the accuracy limits of local pair natural orbital coupled cluster theory. *J Chem Theory Comput* 2015; 11: 1525–1539. doi:10.1021/ct501129s
209 Chan B, Karton A. Assessment of DLPNO-CCSD(T)-F12 and its use for the formulation of the low-cost and reliable L-W1X composite method. *J Comput Chem* 2022; 43: 1394–1402. doi:10.1002/jcc.26892
210 Semidalas E, Martin JML. Canonical and DLPNO-based composite wavefunction methods parametrized against large and chemically diverse training sets. 2: Correlation-consistent basis sets, core − valence correlation, and F12 alternatives. *J Chem Theory Comput* 2020; 16: 7507–7524. doi:10.1021/acs.jctc.0c01106
211 Bursch M, Mewes J-M, Hansen A, Grimme S. Best-practice DFT protocols for basic molecular computational chemistry. *Angew Chem Int Ed* 2022; 61: e202205735. doi:10.1002/anie.202205735
212 Karton A, Ruscic B, Martin JML. Benchmark atomization energy of ethane: importance of accurate zero-point vibrational energies and diagonal Born–Oppenheimer corrections for a 'simple' organic molecule. *J Mol Struct Theochem* 2007; 811: 345–353. doi:10.1016/j.theochem.2007.01.013
213 Jiang J, Ke L, Chen L, Dou B, Zhu Y, Liu J, Zhang B, Zhou T, Wei G-W. Transformer technology in molecular science. *WIREs Comput Mol Sci* 2024; 14: e1725. doi:10.1002/wcms.1725
214 Abraham BM, Jyothirmai MV, Sinha P, Viñes F, Singh JK, Illas F. Catalysis in the digital age: unlocking the power of data with machine learning. *WIREs Comput Mol Sci* 2024; 14: e1730. doi:10.1002/wcms.1730
215 Xue H, Cheng G, Yin W-J. Computational design of energy-related materials: from first-principles calculations to machine learning. *WIREs Comput Mol Sci* 2024; 14: e1732. doi:10.1002/wcms.1732
216 Dalmau D, Alegre-Requena JV. ROBERT: bridging the gap between machine learning and chemistry. *WIREs Comput Mol Sci* 2024; 14: e1733. doi:10.1002/wcms.1733
217 Back S, Aspuru-Guzik A, Ceriotti M, Grynova G, Grzybowski B, Gu GH, Hein J, Hippalgaonkar K, Hormázabal R, Jung Y, Kim S, Kim WY, Moosavi SM, Noh J, Park C, Schrier J, Schwaller P, Tsuda K, Vegge T, von Lilienfeld OA, Walsh A. Accelerated chemical science with AI. *Digit Discov* 2024; 3: 23–33. doi:10.1039/D3DD00213F
218 Bender A, Schneider N, Segler M, Patrick Walters W, Engkvist O, Rodrigues T. Evaluation guidelines for machine learning tools in the chemical sciences. *Nat Rev Chem* 2022; 6: 428–442. doi:10.1038/s41570-022-00391-9
219 Meuwly M. Machine learning for chemical reactions. *Chem Rev* 2021; 121: 10218–10239. doi:10.1021/acs.chemrev.1c00033
220 Keith JA, Vassilev-Galindo V, Cheng B, Chmiela S, Gastegger M, Müller K-R, Tkatchenko A. Combining machine learning and computational chemistry for predictive insights into chemical systems. *Chem Rev* 2021; 121: 9816–9872. doi:10.1021/acs.chemrev.1c00107
221 Westermayr J, Gastegger M, Schütt KT, Maurer RJ. Perspective on integrating machine learning into computational chemistry and materials science. *J Chem Phys* 2021; 154: 230903. doi:10.1063/5.0047760
222 Tkatchenko A. Machine learning for chemical discovery. *Nat Commun* 2020; 11: 4125. doi:10.1038/s41467-020-17844-8
223 von Lilienfeld OA, Müller K-R, Tkatchenko A. Exploring chemical compound space with quantum-based machine learning. *Nat Rev Chem* 2020; 4: 347–358. doi:10.1038/s41570-020-0189-9
224 Strieth-Kalthoff F, Sandfort F, Segler MHS, Glorius F. Machine learning the ropes: principles, applications and directions in synthetic chemistry. *Chem Soc Rev* 2020; 49: 6154–6168. doi:10.1039/C9CS00786E
225 Yu HS, He X, Li SL, Truhlar DG. MN15: a Kohn–Sham global-hybrid exchange-correlation density functional with broad accuracy for multi-reference and single-reference systems and noncovalent interactions. *Chem Sci* 2016; 7: 5032–5051. doi:10.1039/C6SC00705H
226 Saal JE, Kirklin S, Aykol M, Meredig B, Wolverton C. Materials design and discovery with high-throughput density functional theory: the Open Quantum Materials Database (OQMD). *JOM* 2013; 65: 1501–1509. doi:10.1007/s11837-013-0755-4
227 Kirklin S, Saal EJ, Meredig B, Thompson A, Doak JW, Aykol M, Rühl S, Wolverton C. The Open Quantum Materials Database (OQMD): assessing the accuracy of DFT formation energies. *Npj Comput Mater* 2015; 1: 15010. doi:10.1038/npjcompumats.2015.10
228 Gubernatis JE, Lookman T. Machine learning in materials design and discovery: examples from the present and suggestions for the future. *Phys Rev Mater* 2018; 2: 120301. doi:10.1103/PhysRevMaterials.2.120301
229 Bhattacharjee H, Vlachos DG. Thermochemical data fusion using graph representation learning. *J Chem Inf Model* 2020; 60: 4673–4683. doi:10.1021/acs.jcim.0c00699
230 Li Q, Wittreich G, Wang Y, Bhattacharjee H, Gupta U, Vlachos DG. Accurate thermochemistry of complex lignin structures via density functional theory, group additivity, and machine learning. *ACS Sustain Chem Eng* 2021; 9: 3043–3049. doi:10.1021/acssuschemeng.0c08856
231 Bhattacharjee H, Anesiadis N, Vlachos DG. Regularized machine learning on molecular graph model explains systematic error in DFT enthalpies. *Sci Rep* 2021; 11: 14372. doi:10.1038/s41598-021-93854-w
232 Formalik FS, Kaihang J, Faramarz W, Xijun S, Randall Q. Exploring the structural, dynamic, and functional properties of metal–organic frameworks through molecular modeling. *Adv Funct Mater* 2024; 34: 2308130. doi:10.1002/adfm.202308130
233 Rowe P, Deringer VL, Gasparotto P, Csányi G, Michaelides A. An accurate and transferable machine learning potential for carbon. *J Chem Phys* 2020; 153: 034702. doi:10.1063/5.0005084
234 Deringer VL, Caro MA, Csányi G. A general-purpose machine-learning force field for bulk and nanostructured phosphorus. *Nat Commun* 2020; 11: 5461. doi:10.1038/s41467-020-19168-z
235 Milardovich D, Waldhoer D, Jech M, El-Sayed A-MB, Grasser T. Building robust machine learning force fields by composite Gaussian approximation potentials. *Solid-State Electron* 2023; 200: 108529. doi:10.1016/j.sse.2022.108529

236 Klawohn S, Darby JP, Kermode JR, Csányi G, Caro MA, Bartók AP. Gaussian approximation potentials: theory, software implementation and application examples. *J Chem Phys* 2023; 159: 174108. doi:10.1063/5.0160898
237 Nagai R, Akashi R, Sugino O. Completing density functional theory by machine learning hidden messages from molecules. *npj Comput Mater* 2020; 6: 43. doi:10.1038/s41524-020-0310-0
238 Kirkpatrick J, McMorrow B, Turban DHP, Gaunt AL, Spencer JS, Matthews AGDG, Obika A, Thiry L, Fortunato M, Pfau D, Castellanos LR, Petersen S, Nelson AWR, Kohli P, Mori-Sánchez P, Hassabis D, Cohen AJ. Pushing the frontiers of density functionals by solving the fractional electron problem. *Science* 2021; 374: 1385–1389. doi:10.1126/science.abj6511
239 del Rio BG, Phan B, Ramprasad R. A deep learning framework to emulate density functional theory. *npj Comput Mater* 2023; 9: 158. doi:10.1038/s41524-023-01115-3
240 Xiao J, Chen Y, Zhang L, Wang H, Zhu T. A machine learning-based high-precision density functional method for drug-like molecules. *Artificial Intelligence Chemistry* 2024; 2(1): 100037. doi:10.1016/j.aichem.2023.100037

Data availability. This review is based on previously published research.

Conflicts of interest. Amir Karton is an Associate Editor of the *Australian Journal of Chemistry* but did not at any stage have editor-level access to this manuscript while in peer review, as is the standard practice when handling manuscripts submitted by an editor to this journal. *Australian Journal of Chemistry* encourages its editors to publish in the journal and they are kept totally separate from the decision-making processes for their manuscripts. The author has no further conflicts of interest to declare.

Declaration of funding. This research did not receive any specific funding.

Author affiliation
[A]School of Science and Technology, University of New England, Armidale, NSW 2351, Australia.