Journal Name

ARTICLE TYPE

Cite this: DOI: 00.0000/xxxxxxxxxx

All hands on deck: Accelerating *ab initio* thermochemistry via wavefunction approximations †

Sambit Kumar Das,[a,‡] Salini Senthil,[a,‡], Sabyasachi Chakraborty,[a] and Raghunathan Ramakrishnan*[a]

Received Date
Accepted Date

DOI: 00.0000/xxxxxxxxxx

We accelerate the G4(MP2) composite model by fine-tuning the individual steps using resolution-of-identity and domain-based local pair-natural orbitals. The new variant, G4(MP2)-XP, has a low prediction error when tested on 1694 benchmark molecules. To showcase the method's relevance for large molecules, we determine and present a new reference value for the standard formation enthalpy of buckminsterfullerene. We expect G4(MP2)-XP to become the *de facto* method for rapid and accurate production of thermochemistry big data.

Modern composite wavefunction theories[1–4] can forecast small molecular atomization/dissociation energies to a degree of accuracy that rivals experimental precision. For routine applications to organic or combustion chemistry problems, G4(MP2)[5] is the chemist's workhorse. It predicts formation enthalpies, ΔH_f°, of 138 small hydrocarbon derivatives with a mean unsigned error (MUE) of 0.78 kcal/mol compared to experiment[5]. Curtiss *et al.*[6] showed this accuracy to be transferable to 459 small-to-medium sized hydrocarbons and their substituted analogs. When benchmarking on complex molecules of similar composition, from the probabilistically pruned dataset of 1694 entries (PPE1694), G4(MP2)'s MUE was found to be 1.67 kcal/mol[7].

Among density functional theory (DFT), only the range-separated hybrid method, ωB97M-D3BJ[8–10], has a similar accuracy, with an MUE of 1.73 kcal/mol[7]. Such high accuracy is made possible by using a large quadruple-zeta basis set along with dressed atom corrections to account residual systematic errors. Refitting an atom-type independent, higher-level correction (HLC) term in G4(MP2) to PPE1694 resulted in a negligible improvement, indicating the method's universality across datasets. Due to the favorable computational scaling and versatility, G4(MP2) qualified as the method of choice for thermochemical modeling[6] of 133,885 small organic molecules in the QM9 dataset[11].

A G4(MP2) calculation proceeds with geometry optimization and frequency calculations at the B3LYP/6-31G(2df,p) level. The harmonic wave numbers are scaled by 0.9854 to approximately account for anharmonicity. The zero-Kelvin internal energy, U_0, is the sum of electronic energy, determined through a series of single-point calculations, and the zero-point vibrational energy, ZPVE.

$$U_0 = E^{\mathrm{CCSD(T)}}_{6-31\mathrm{G}(d)} + \Delta E^{\mathrm{MP2}} + \Delta E^{\mathrm{HF}} + \mathrm{HLC} + \mathrm{SO} + \mathrm{ZPVE} \quad (1)$$

Electronic energy from CCSD(T)/6-31G(d) is used as a baseline to which basis set corrections determined at MP2 and HF are added: $\Delta E^{\mathrm{MP2}} = E^{\mathrm{MP2}}_{\mathrm{G3MP2LargeXP}} - E^{\mathrm{MP2}}_{6-31\mathrm{G}(d)}$ and $\Delta E^{\mathrm{HF}} = E^{\mathrm{HF}}_{\mathrm{CBS}} - E^{\mathrm{HF}}_{\mathrm{G3MP2LargeXP}}$. The small basis set terms, $E^{\mathrm{MP2}}_{6-31\mathrm{G}(d)}$ and $E^{\mathrm{HF}}_{\mathrm{G3MP2LargeXP}}$, are not determined separately but taken from CCSD(T) and MP2 calculations performed with the corresponding basis sets. The HF-limit, $E^{\mathrm{HF}}_{\mathrm{CBS}}$, is determined using a two-point extrapolation with two large basis sets[5]. All remaining basis set deficiencies are captured using a size-intensive (correction per electron) higher-level correction, HLC. Empirical spin-orbit coupling corrections, SO, are included for atoms, selected diatomics and one polyatomic molecule ($C_2H_2^+$) as per previous reports[5,7]. Thermal effects leading to finite-temperature internal energy (U_T) and enthalpy (H_T) are incorporated using the ideal gas, harmonic oscillator and rigid rotor partition functions.

G4(MP2)-6X[12] was introduced to quench the remaining errors in G4(MP2) by scaling the correlation energy components through six empirical factors. For 526 energies from experimental and theoretical sources, G4(MP2)-6X has an MUE of 0.87 kcal/mol, while G4(MP2)'s error was at 1.06 kcal/mol. The speed of G4(MP2)-6X was improved in G4(MP2)-XK by replacing the Pople-type basis sets with the Karlsruhe-type basis sets[13]. In the recently proposed G4(MP2)-XK-D and G4(MP2)-XK-T versions[14],

[a] *Tata Institute of Fundamental Research, Centre for Interdisciplinary Sciences, Hyderabad 500107, India*

† Electronic Supplementary Information (ESI) available: https://moldis-group.github.io/pople/, see Data Availability.

‡ These authors contributed equally to this work.

the HLC term was shown to have a negligible effect when the basis sets are augmented with diffuse functions. G4(MP2)-XK-T using large basis sets has a weighted MUE of 1.43 kcal/mol when tested on the GMTKN55 dataset [15]. Introducing the domain-based local pair natural orbital (DLPNO) approximation [16,17] at the CCSD(T)-level showed a better speed-to-accuracy tradeoff, incurring a smaller CPU time but the MUE rising to 1.66 kcal/mol. Approximating CCSD(T) with DLPNO in the correlation consistent composite approach (ccCA [2]) facilitated seamless calculations of intermolecular binding energies of large molecular dimers [18].

In the present study, we modify G4(MP2) by systematic introduction of wavefunction approximations. We show that achieving favorable error cancellation requires a judicious combination of resolution-of-identity [19,20] (RI) based HF and MP2, with DLPNO-based [16,17] CCSD(T) energies. When these approximations—with appropriate settings—are active at all levels, the fastest variation, G4(MP2)-XP, shows several fold speedup. In all correlated calculations, we used a small frozen-core approximation [5]. All calculations were done with the open-source thermochemistry toolkit Pople [21], developed as a part of the present study, that interfaces to the ORCA 4.2 quantum chemistry program package [22,23].

To comprehensively probe the effect of approximations, we calculated each contribution separately and merged them combinatorially. The improvement in efficiency at each level is depicted for selected alkanes in Fig. 1. The cost of geometry optimization and frequency calculations can often overtake that of SCF and post-SCF energies for large molecules, hence, we employ RI to accelerate the hybrid-DFT calculations. The Coulomb integrals are approximated using the general Weigend [24] def2/J auxiliary basis set while the exchange energy is calculated through a semi-numerical integration based on the 'chain-of-spheres' (COS) algorithm [25] (RIJCOSX) combined with Grid7/GridX9 settings. For icosane, RIJCOSX increases the speed of DFT-level geometry optimization by a factor of 3.38 (Fig. 1).

HF wavefunctions are approximated with RI not only for determining ΔE^{HF}, but also for calculating the one-electron wavefunctions for MP2 and CCSD(T) steps. At all levels, Coulomb (J) and exchange (K) integrals are modeled with the general Weigend def2/JK auxiliary basis set [24] leading to HF/RIJK, MP2/RIJK, and CCSD(T)/RIJK. The relative gain in speed due to RIJK is two-fold for the ΔE^{HF} step. Due to the higher-order computational scaling of CCSD(T), introduction of RIJK brings only a minor improvement, while it becomes effective in MP2. For the alkanes under consideration, evaluation of ΔE^{MP2} is made 9.00–9.36 times faster, when RI is introduced at the MP2 level through the auxiliary correlation fitting basis set def2-TZVP/C. The strongest response on the CPU time happens upon replacing CCSD(T) with DLPNO-CCSD(T). We found def2-SVP/C to be a suitable auxiliary correlation fitting basis set upon exploratory benchmarking. In all DLPNO-CCSD(T) calculations, we use a full-local MP2 guess and also request an RI-MP2 evaluation at this step that is used to determine ΔE^{MP2}. In the case of pentane, DLPNO does not bring any noticeable speedup as shown in the inset to Fig. 1. For large systems, DLPNO delivers improved

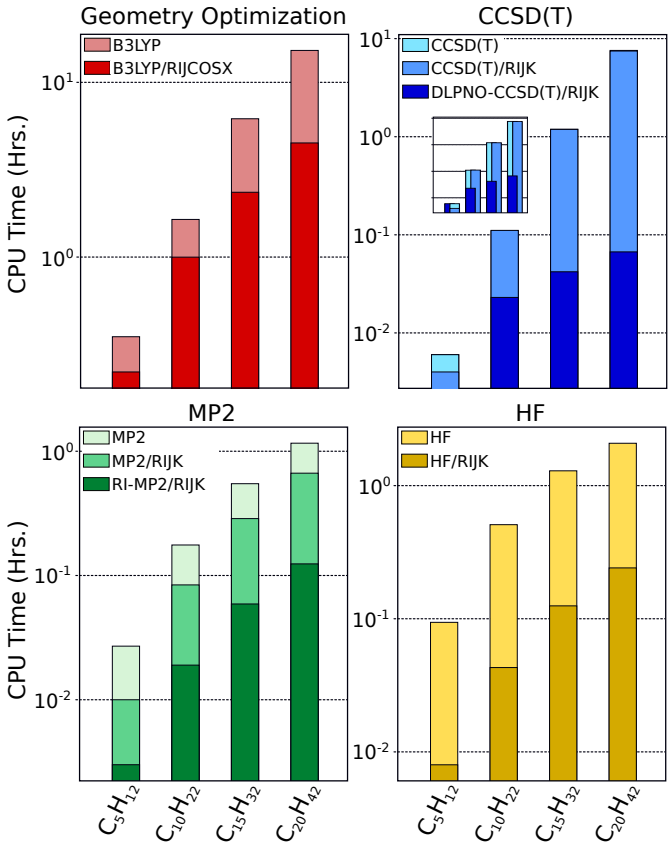

Fig. 1 Comparison of computational cost of individual steps in G4(MP2) with that of their RI/DLPNO approximated variants. CPU times (in log scale) are shown for pentane (C_5H_{12}), decane ($C_{10}H_{22}$), pentadecane ($C_{15}H_{32}$) and icosane ($C_{20}H_{42}$). All calculations were done on 8 cores of an Intel-E7-8867-v4 processor @ 2.40GHz.

performance reaching over orders of magnitude speedups for pentadecane (28.38) and icosane (113.18).

Individual contributions to H_{T} calculated at different levels with and without RI/DLPNO, culminate in 36 variations of G4(MP2) including the original formalism. For each variant, we optimized HLC parameters separately by minimizing the MUE compared to the experimental values for the G3/05 dataset. Some of the variants, neither result in consistent cancellation of terms nor lead to unique predictions. In Table. 1 we compare the accuracies of ten selected versions named in the increasing order of their speedups for calculating H_{T} for icosane. We benchmark on the G3/05 dataset and compare with G4(MP2) results [5], where we exclude ionization potentials for H_2S ($^2B_1$) and N_2 ($^2\Sigma_g$).

G4(MP2)-v1 is G4(MP2) with geometries and harmonic frequencies from RIJCOSX-B3LYP/6-31G(2df,p). For icosane, G4(MP2)-v1 is 1.70 faster than the parent method, while preserving the overall MUE at 1.05 kcal/mol. This improvement motivated us to focus on variants based on RI-DFT geometries (see Table. 1). Introducing HF/RIJK in G4(MP2)-v1 for $E_{\mathrm{CBS}}^{\mathrm{HF}}$ results in G4(MP2)-v3 with improved speed and accuracy compared to G4(MP2)-v2. In the latter, RI-MP2/RIJK is introduced at ΔE^{MP2}. An increase in MUE by 0.02 kcal/mol at G4(MP2)-v2 originates from an incomplete post-SCF basis

Table 1 Accuracy of the RI/DLPNO approximated G4(MP2) variants benchmarked for the G3/05 dataset: mean unsigned error (MUE) and root mean square error (RMSE) for standard formation enthalpy (ΔH_f° (298 K)), ionization potential (IP), electron affinity (EA), proton affinity (PA), and binding energy (BE) are given. Number of entries in parenthesis. ✓/✗ denotes presence/absence of an approximation. G4(MP2) results[7] are given for comparison. Relative speedup is given for each method using G4(MP2) computation time for icosane ($C_{20}H_{42}$) as the reference: 'Full' corresponds to a complete calculation while 'SP' denotes SCF and post-SCF single point calculations of the electronic energy.

Method	B3LYP	HF	MP2	CCSD(T)			Speedup		MUE (RMSE) in kcal/mol					
	RIJCOSX	RIJK	RIJK	RI-MP2	RIJK	DLPNO	Full	SP	ΔH_f° (270)	IP (103)	EA (63)	PA (10)	BE (6)	Total (452)
G4(MP2)	✗	✗	✗	✗	✗	✗	1.00	1.00	1.00 (1.40)	1.08 (1.64)	1.26 (1.66)	0.66 (0.87)	1.29 (1.91)	1.05 (1.50)
G4(MP2)-v1	✓	✗	✗	✗	✗	✗	1.70	1.00	1.00 (1.42)	1.08 (1.63)	1.27 (1.67)	0.67 (0.88)	1.28 (1.92)	1.05 (1.51)
G4(MP2)-v2	✓	✗	✓	✓	✗	✗	1.82	1.11	1.03 (1.45)	1.07 (1.63)	1.25 (1.62)	0.72 (0.91)	1.27 (1.91)	1.07 (1.52)
G4(MP2)-v3	✓	✓	✗	✗	✗	✗	1.93	1.20	1.00 (1.42)	1.07 (1.63)	1.26 (1.66)	0.69 (0.89)	1.29 (1.93)	1.05 (0.50)
G4(MP2)-v4	✓	✓	✓	✗	✓	✗	2.02	1.29	0.99 (1.38)	1.04 (1.56)	1.29 (1.65)	0.70 (0.89)	1.29 (1.93)	1.04 (1.46)
G4(MP2)-v5	✓	✓	✓	✓	✓	✗	2.11	1.38	1.01 (1.41)	1.03 (1.56)	1.28 (1.62)	0.75 (0.93)	1.28 (1.92)	1.05 (1.47)
G4(MP2)-v6	✓	✓	✓	✓	✗	✗	2.09	1.36	1.03 (1.45)	1.07 (1.63)	1.24 (1.61)	0.74 (0.92)	1.28 (1.92)	1.07 (1.51)
G4(MP2)-v7	✓	✗	✗	✗	✓	✓	3.34	3.27	0.96 (1.34)	1.06 (1.63)	1.29 (1.69)	0.71 (0.91)	1.35 (1.96)	1.03 (1.47)
G4(MP2)-v8	✓	✗	✓	✓	✓	✓	3.85	4.76	0.98 (1.40)	1.05 (1.62)	1.26 (1.61)	0.74 (0.93)	1.34 (1.95)	1.03 (1.48)
G4(MP2)-v9	✓	✓	✗	✗	✓	✓	4.37	7.37	0.96 (1.35)	1.06 (1.63)	1.29 (1.69)	0.72 (0.92)	1.36 (1.97)	1.03 (1.47)
G4(MP2)-XP	✓	✓	✓	✓	✓	✓	5.28	25.06	0.97 (1.37)	1.05 (1.62)	1.25 (1.61)	0.76 (0.94)	1.35 (1.96)	1.03 (1.47)

set correction, wherein ΔE^{MP2} is determined as $E^{RI-MP2/RIJK}_{G3MP2LargeXP} - E^{MP2}_{6-31G(d)}$. Similar accuracy is noted for G4(MP2)-v6, where HF/RIJK-limit and $\Delta E^{RI-MP2/RIJK}$ are incorporated in G4(MP2)-v1. G4(MP2)-v4 and G4(MP2)-v5, while showing somewhat inferior speed compared to G4(MP2)-v6, have better accuracies because of consistent cancellation of terms.

G4(MP2)-v7, where DLPNO-CCSD(T)/RIJK is introduced in G4(MP2)-v1, shows a stark performance upgradation with a 'Full' speedup of 3.34 and an improved MUE of 1.03 kcal/mol (Table. 1). Of all properties, maximum improvement in accuracy is noted for ΔH_f°, where G4(MP2)'s MUE of 1.00 kcal/mol is quenched to 0.96 kcal/mol. G4(MP2)-v7 combined with HF/RIJK-limit, G4(MP2)-v9, is faster than when combined with $\Delta E^{RI-MP2/RIJK}$, G4(MP2)-v8. The union of G4(MP2)-v8 and G4(MP2)-v9 gives rise to the 10th version, which upon optimized parameterization of the HLC term is henceforth denoted as G4(MP2)-XP. A calculation of H_T with G4(MP2)-XP is about 5.3 times faster than with the canonical G4(MP2). This acceleration must increase for molecules larger than the test case, icosane. Post geometry optimization, G4(MP2)-XP shows a 25-fold acceleration.

The favorable cost-vs-accuracy trade-off shown by G4(MP2)-XP prompted us to test its accuracies across larger and more diverse molecules. For this purpose, we selected the PPE1694 dataset[7]. In Table. 2, we tabulate the error metrics for G4(MP2)-XP across all subsets of PPE1694. G4(MP2)-XP with the DLPNO-CCSD(T)/RIJK component outperforms G4(MP2) consistently. For hydrocarbon derivatives, G4(MP2)-XP's accuracies are comparable to that of G4[26]. A previous work[18] has noted a similar trend, where DLPNO-CCSD(T) based ccCA resulted in lower errors for alkanes compared to ccCA and RI-ccCA. While both G4(MP2) variants have better accuracies for the inorganic hydrides, this comparison lies within the uncertainty threshold used for pruning the benchmark dataset[7]. Based on the accuracy for the PPE1694 dataset, G4(MP2)-XP is convincingly placed between G4 and G4(MP2), but its reduced cost extends its domain of applicability to larger datasets.

G4(MP2)-XP opens up the possibilities to determine the energetics of very large symmetric molecules in 'one-shot', in a single point fashion, provided the corresponding geometry and harmonic wavenumbers are available. We demonstrate this by presenting the ΔH_f° (298 K) of buckminsterfullerene, C_{60}, see Fig. 2. We exploited its icosahedral symmetry and scanned the potential energy surface along the radial mode to find a minimum at $r_{eq.} = 3.52$ Å. At the same geometry we determined the harmonic vibrational wavenumbers. The latest experimental reference for C_{60}'s ΔH_f° (298 K) is given by NIST[27] as 610 ± 30 kcal/mol, a value determined by averaging over 6 experimental estimations. The large experimental uncertainty compelled many to provide theoretical support for bracketing the NIST value[28–32]. However, ΔH_f° (298 K) of C_{60} has thus far been determined at the composite wavefunction level only by depending on group-additivity/isodesmic reaction schemes that may introduce non-negligible uncertainties.

Tasi *et al.*[33] pointed out how thermochemistry calculations of alkanes are prone to systematic errors accumulated through uncertainties in the zero-Kelvin formation enthalpies of carbon atom in gas phase, ΔH_f° (0 K; C_{gas}). Conventionally, this parameter is fixed at 169.98 kcal/mol—the cohesive energy of graphite. The quantity was revised to 170.11 kcal/mol by including an electronic spin splitting correction for C along with data from high-level theories[33]. The latter value was found to be appropriate for determining ΔH_f° (298 K) of alkanes at G4, G4(MP2), ccCA, and DFT methods[7]. A value of 170.027 ± 0.013 kcal/mol for the same quantity proposed by the Active Thermochemical Tables (ATcT)[34] has been used in theoretical

Table 2 Performance of G4(MP2)-XP for the PPE1694 dataset. Mean unsigned error (MUE) and root mean square error (RMSE) are given for ΔH_f° (298 K) across subsets of PPE1694. Error metrics for G4 and G4(MP2)[7] are given for comparison.

Subset	MUE (RMSE) in kcal/mol		
	G4	G4(MP2)	G4(MP2)-XP
Non Hydrogens (150)	1.95 (2.97)	2.21 (3.31)	2.17 (3.22)
Hydrocarbons (623)	1.51 (2.32)	1.81 (2.80)	1.52 (2.39)
Subst. Hydrocarbons (852)	1.40 (1.98)	1.57 (2.15)	1.44 (2.01)
Inorganic Hydrides (30)	0.98 (1.25)	0.85 (1.20)	0.82 (1.12)
Radicals (39)	1.13 (1.49)	1.37 (1.88)	1.34 (1.72)
Total (1694)	1.47 (2.18)	1.70 (2.50)	1.52 (2.27)

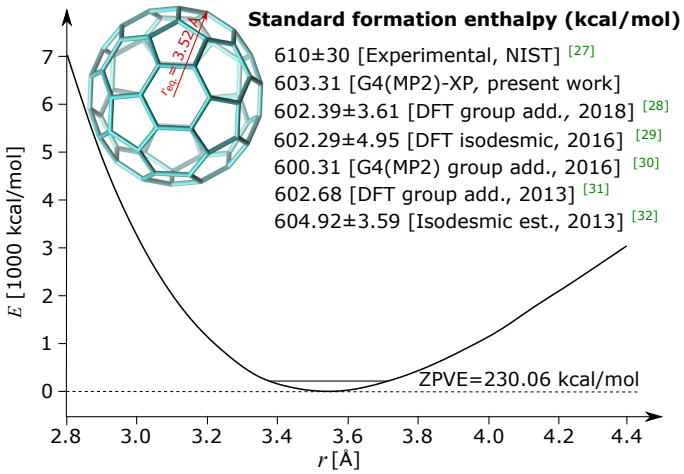

Fig. 2 ΔH_f° (298 K) of buckminsterfullerene, C_{60}, at the G4(MP2)-XP level presented with available experimental and calculated values.

investigations[35]. This empirical value accounts for the enthalpies of reactions such as $C_{gas} + O_{gas} \rightarrow CO_{gas}$. For three choices of ΔH_f° (0 K; C_{gas}): 169.98, 170.027, and 170.11 kcal/mol, we determine the G4MP2-XP value of ΔH_f° (298 K) of C_{60} as 600.49, 603.31, and 608.29 kcal/mol, respectively. Of these, the second value should be compared with more recent theoretical estimations that follow the ATcT conventions (see Fig. 2).

In summary, we explored ten variants of G4(MP2) designed by deploying RI and DLPNO at various steps, and reparameterizing the HLC term. The fastest variant, G4(MP2)-XP, is also the most accurate as vetted by benchmarks over G3/05 and PPE1694 datasets. Post geometry optimization and frequency steps, G4(MP2)-XP is 25-fold faster than G4(MP2) when clocked for icosane. The speedup is expected to increase for larger molecules. A direct calculation of C_{60}'s total energy at G4(MP2)-XP requires about 1 CPU day on 8-cores, which when using G4(MP2) may require over 1 CPU month. Upon evaluations of thermodynamic atomic parameters, we present 603.31 kcal/mol as an accurate benchmark value for ΔH_f° (298 K) of C_{60}.

The canonical G4(MP2) approach has been applied for thermochemical calculations of the QM9 dataset with 133,885 organic molecules[6]. The improved speed offered by G4(MP2)-XP should facilitate larger studies by sampling suitable subsets of the GDB17 molecular universe[36] comprising 166.4 billion closed-shell, organic molecules. When such grand high-throughput endeavors are undertaken with G4MP2-XP, measures[37] for preserving the intended Lewis structure connectivities during geometry optimization must be considered to establish the veracity of the data.

Acknowledgements

We acknowledge support of the Department of Atomic Energy, Government of India, under Project Identification No. RTI 4007. All calculations have been performed using the Helios computer cluster, which is an integral part of the MolDis Big Data facility, TIFR Hyderabad (http://moldis.tifrh.res.in).

Data Availability

The data that support the findings of this study are openly available at https://moldis-group.github.io/pople/. The site provides the source code, and manual for "Pople: A toolkit written in Python to perform ab initio thermochemistry calculations" along with example input/output files. G4(MP2) and G4(MP2)-XP results for G3/05 and PPE1694 datasets are available as benchmark suites importable in a python code. A stand-alone code to reproduce the results for C_{60} is also provided.

Conflicts of interest

"There are no conflicts to declare".

Notes and references

1 L. A. Curtiss, P. C. Redfern and K. Raghavachari, *WIREs Comput. Mol. Sci.*, 2011, **1**, 810–825.
2 C. Peterson, D. Penchoff and A. Wilson, *Annual Reports in Computational Chemistry*, Elsevier, 2016, vol. 12, pp. 3–45.
3 A. Tajti, P. G. Szalay, A. G. Császár, M. Kállay, J. Gauss, E. F. Valeev, B. A. Flowers, J. Vázquez and J. F. Stanton, *J. Chem. Phys.*, 2004, **121**, 11599–11613.
4 A. Karton, *WIREs Comput. Mol. Sci.*, 2016, **6**, 292–310.
5 L. A. Curtiss, P. C. Redfern and K. Raghavachari, *J. Chem. Phys.*, 2007, **127**, 124105.
6 B. Narayanan, P. C. Redfern, R. S. Assary and L. A. Curtiss, *Chem. Sci.*, 2019, **10**, 7449–7455.
7 S. K. Das, S. Chakraborty and R. Ramakrishnan, *J. Chem. Phys.*, 2021, **154**, 044113.
8 N. Mardirossian and M. Head-Gordon, *J. Chem. Phys.*, 2016, **144**, 214110.
9 S. Grimme, S. Ehrlich and L. Goerigk, *J. Comp. Chem.*, 2011, **32**, 1456–1465.
10 A. D. Becke and E. R. Johnson, *J. Chem. Phys.*, 2005, **123**, 154101.
11 R. Ramakrishnan, P. O. Dral, M. Rupp and O. A. von Lilienfeld, *Sci. Data*, 2014, **1**, 140022.
12 B. Chan, J. Deng and L. Radom, *J. Chem. Theory Comput.*, 2011, **7**, 112–120.
13 B. Chan, A. Karton and K. Raghavachari, *J. Chem. Theory Comput.*, 2019, **15**, 4478–4484.
14 E. Semidalas and J. M. Martin, *J. Chem. Theory Comput.*, 2020, **16**, 4238–4255.
15 L. Goerigk, A. Hansen, C. Bauer, S. Ehrlich, A. Najibi and S. Grimme, *Phys. Chem. Chem. Phys.*, 2017, **19**, 32184–32215.
16 C. Riplinger and F. Neese, *J. Chem. Phys.*, 2013, **138**, 034106.
17 C. Riplinger, B. Sandhoefer, A. Hansen and F. Neese, *J. Chem. Phys.*, 2013, **139**, 134101.
18 P. Patel and A. K. Wilson, *J. Comp. Chem.*, 2020, **41**, 800–813.
19 R. A. Kendall and H. A. Früchtl, *Theor. Chem. Acc.*, 1997, **97**, 158–163.
20 O. Vahtras, J. Almlöf and M. Feyereisen, *Chem. Phys. Lett.*, 1993, **213**, 514–518.
21 S. Das, S. Senthil, S. Chakraborty and R. Ramakrishnan, *Pople: A toolkit for ab initio thermochemistry*, 2021, https://moldis-group.github.io/pople/, accessed on 2021-04-29.
22 F. Neese, *WIREs Comput. Mol. Sci.*, 2012, **2**, 73–78.
23 F. Neese, *WIREs Comput. Mol. Sci.*, 2018, **8**, e1327.
24 F. Weigend, *Phys. Chem. Chem. Phys.*, 2006, **8**, 1057–1065.
25 R. Izsák and F. Neese, *J. Chem. Phys.*, 2011, **135**, 144105.
26 L. A. Curtiss, P. C. Redfern and K. Raghavachari, *J. Chem. Phys.*, 2007, **126**, 084108.
27 T. Allison, *NIST Polycyclic Aromatic Hydrocarbon Structure Index*, https://pah.nist.gov/, accessed on 2021-04-29.
28 J. A. Bumpus, *J. Phys. Chem. A*, 2018, **122**, 6615–6632.
29 B. Chan, Y. Kawashima, M. Katouda, T. Nakajima and K. Hirao, *J. Am. Chem. Soc.*, 2016, **138**, 1420–1429.
30 W. Wan and A. Karton, *Chem. Phys. Lett.*, 2016, **643**, 34–38.
31 A. Karton, B. Chan, K. Raghavachari and L. Radom, *J. Phys. Chem. A*, 2013, **117**, 1834–1842.
32 F. J. Dobek, D. S. Ranasinghe, K. Throssell and G. A. Petersson, *J. Phys. Chem. A*, 2013, **117**, 4726–4730.
33 G. Tasi, R. Izsák, G. Matisz, A. G. Császár, M. Kállay, B. Ruscic and J. F. Stanton, *ChemPhysChem*, 2006, **7**, 1664–1667.
34 *Active Thermochemical Tables*, https://atct.anl.gov/Thermochemical Data/version 201.122p/index.php, accessed on 2021-04-29.
35 A. Karton and J. M. Martin, *J. Chem. Phys.*, 2012, **136**, 124114.
36 L. Ruddigkeit, R. Van Deursen, L. C. Blum and J.-L. Reymond, *J. Chem. Inf. Model.*, 2012, **52**, 2864–2875.
37 S. Senthil, S. Chakraborty and R. Ramakrishnan, *Chem. Sci.*, 2021, **12**, 5566–5573.