

Contents lists available at ScienceDirect

Chemical Physics Letters



journal homepage: www.elsevier.com/locate/cplett

PAH335 – A diverse database of highly accurate CCSD(T) isomerization energies of 335 polycyclic aromatic hydrocarbons



Amir Karton^{a,*}, Bun Chan^b

^a School of Science and Technology, University of New England, Armidale, NSW 2351, Australia
^b Graduate School of Engineering, Nagasaki University, Bunkyo 1-14, Nagasaki 852-8521, Japan

ARTICLE INFO	A B S T R A C T		
Keywords: CCSD(T) G4(MP2) DFT Polycyclic aromatic hydrocarbons Isomerization energies	We obtain CCSD(T) isomerization energies for 335 polycyclic aromatic hydrocarbons by means of the G4(MP2) thermochemical protocol. We use this large and diverse dataset of isomerization energies for identifying <i>a-priori</i> indicators for energetic instability and benchmarking DFT and semiempirical methods. We find that dipole moments above 1.0 D, HOMO-LUMO gaps below 2.0 eV, and $q_{neg,avg}$ values below $-0.06 e$ indicate energetic instability. Surprisingly, the local density approximation SVWN5 method attains a mean absolute deviation (MAD) of merely 3.8 kJ/mol and outperforms many functionals from higher rungs of Jacob's Ladder. PBE0-D4 is the best performer with a MAD of 2.8 kJ/mol.		

1. Introduction

Due to their intrinsic energetic stability via π -electron delocalization and resonance stabilization, polycyclic aromatic hydrocarbons (PAHs) are one of the most ubiquitous classes of naturally occurring and synthetic compounds. PAHs have found numerous applications in organic and materials chemistry [1-6]. Many PAHs are carcinogenic and genotoxic, and therefore computational chemistry plays a critical role in exploring their reactivity and potential chemical applications. Two areas in which computational chemistry is playing an increasingly important role in investigating the reactivity and chemical properties of PAHs are materials and combustion chemistry [3-6]. However, these computational investigations normally involve large PAHs or multiple potential reaction pathways involving many reaction intermediates. It is, therefore, important to identify reliable quantum chemical procedures such as density functional theory (DFT) and semiempirical molecular orbital (SMO) methods that are applicable to large PAHs. To this end, we construct an extensive and highly diverse database of 335 PAHs and calculate their relative isomerization energies at the CCSD(T) level (i.e., coupled cluster with singles, doubles, and quasiperturbative triple excitations) via the high-level G4(MP2) composite ab initio method [7,8]. We use these benchmark isomerization energies to identify chemical indicators for predicting highly energetic PAHs and to identify computationally economical DFT and SMO methods that are able to produce chemically accurate isomerization energies for this extensive dataset.

2. Computational details

High-level isomerization energies were obtained using the G4(MP2) thermochemical protocol [7,8]. The G4(MP2) composite ab initio protocol is a computationally efficient composite procedure for approximating the CCSD(T) energy in conjunction with a large triple- ζ -quality basis set [7,8]. G4(MP2) theory uses the following MP2-based additivity scheme:

$$E[G4(MP2)] = E[CCSD(T)/6 - 31G(d)] + E(MP2) + E(HF)$$
(1)

where $\Delta E(MP2)$ and $\Delta E(HF)$ are basis set correction terms calculated using triple- ζ and quadruple- ζ quality basis sets, respectively [7]. We note that for the largest PAHs considered in the present work, namely the C₃₆H₁₈ isomers with C_s symmetry, each of the CCSD(T)/6-31G(d) calculations ran for over 10 days on 6 cores of a dual Intel Xeon machine with 192 GB of RAM. The G4(MP2) protocol is widely used for the calculation of thermochemical properties such as bond dissociation and isomerization energies of main-group compounds (for a review of the Gaussian-*n* methods, see reference [8]). G4(MP2) theory has been found to produce thermochemical properties of hydrocarbons with mean absolute deviations (MADs) from accurate benchmark data below the threshold of chemical accuracy (arbitrarily defined as 4.2 kJ mol⁻¹) [7–13]. The geometries and harmonic vibrational frequencies of all structures have been obtained at the B3LYP/6-31G(2df,p) level of theory [14,15] as prescribed in the G4(MP2) protocol [7,8]. We note in passing

* Corresponding author. *E-mail address:* amir.karton@une.edu.au (A. Karton).

https://doi.org/10.1016/j.cplett.2023.140544

Received 6 February 2023; Received in revised form 2 April 2023; Accepted 22 April 2023 Available online 24 April 2023 0009-2614/© 2023 Elsevier B.V. All rights reserved. that the B3LYP functional has been found to provide excellent performance for calculating equilibrium structures relative to a wide and diverse set of 122 organic and inorganic molecules [16].

All equilibrium structures were verified to have all real harmonic frequencies. We use the benchmark isomerization energies to evaluate the performance of a variety of DFT methods, namely SVWN5 [17], BLYP [14,18], PBE [19], TPSS [20], M06-L [21], MN15-L [22], B3LYP [14,15], PBE0 [23], BH&HLYP [24], CAM-B3LYP [25], @B97X [26], ωB97X-D [27], TPSSh [28], PW6B95 [29], M06 [30], M06-2X [30], and MN15 [22], The DFT calculations were performed in conjunction with the Def2-TZVPP basis set [31]. Where relevant the atomic-charge dependent D4 empirical dispersion correction is included [32]. We note that our G4(MP2) reference values are not deemed sufficiently accurate for evaluating the performance of double-hybrid DFT (DHDFT) methods. In addition, due to their steep computational cost and basis set dependency, DHDFT methods are computationally too demanding for studying relatively large PAHs. Therefore, DHDFT methods are not considered here. However, we do consider the performance of SMO methods which are computationally more economical than DFT methods. In particular, we consider the performance of the following SMO methods AM1 [33], PM6 [34], PM7 [35], and XTB [36].

The CCSD(T) calculations involved in the G4(MP2) procedure were performed with the Molpro 2016 program [37]. The XTB calculations were performed with Grimme's XTB code (version 6.5.1). All other ab initio, DFT, and SMO calculations were carried out using the Gaussian 16 program suite [38].

3. Results and discussion

The isomerization energy is a fundamental thermochemical quantity that is needed for comparing the thermodynamic stability of species that have the same atomic composition. In the present work, we examine the isomerization energies of 335 PAHs, where the isomerization energy (ΔE_{iso}) is defined as the energy difference between a PAH with the general molecular formula C_nH_m and the energetically most stable structure within that isomer space. All the initial structures in the PAH335 database were taken from the NIST PAH Database [39]. Of the 660 PAHs in that database, only C_nH_m isomer sets with at least ten isomers are considered in the present work. The resulting PAH335 database includes the following isomer sets (the number of isomers in each subset is given in parentheses): C17H12 (11), C19H12 (12), C20H12 (21), C₂₁H₁₄ (27), C₂₂H₁₂ (15), C₂₂H₁₄ (12), C₂₃H₁₄ (18), C₂₄H₁₄ (85), C25H16 (10), C26H14 (16), C26H16 (37), C28H14 (10), C28H16 (48), and C₃₆H₁₆ (13). Fig. 1a shows the distribution of the G4(MP2) isomerization energies in the entire PAH335 database. The G4(MP2) isomerization energies are given in Table S1 of the Supporting Information. The isomerization energies vary smoothly and continuously up to 291.1 kJ mol⁻¹. Three isomers have significantly higher relative energies of 351.6-365.8 kJ mol⁻¹ (Fig. 1a). It can also be seen that the energy distribution has two fairly linear regimes with different slopes. The first region varies between 0 and 150 kJ mol⁻¹ with a squared correlation coefficient of $R^2 = 0.95$, and the second region varies between 150 and 300 kJ mol⁻¹ with a squared correlation coefficient of $R^2 = 0.96$. In the first region (0–150 kJ mol⁻¹), the energy difference between consecutive isomers is small and amounts to about 0.5 kJ mol^{-1} on average. In the second region (150–300 kJ mol^{-1}), the energy difference between consecutive isomers is much larger and amounts to about 5 kJ mol⁻¹ on average. Fig. 1b shows the distribution of the isomerization energies in each of the subsets. As mentioned above, the PAH335 database includes 14 different isomer subsets, where each subset contains between 10 and 85 isomers. Table 1 gives an overview of the number of isomers and their energy distribution for the subsets in the PAH335 database. The distribution of the electronic isomerization energies (ΔE_{iso}) can vary substantially between the different subsets. The average isomerization energy across each of the sets varies between 14.8 (C₂₅H₁₆) and 121.5 $(C_{24}H_{14})$ kJ mol⁻¹. Similarly, the largest isomerization energies in each



Fig. 1. Overview of the G4(MP2) CCSD(T) isomerization energies in (a) the PAH335 database and (b) in each of the C_nH_m subsets, the number of isomers in each subset is given in parentheses (energies are given in kJ mol⁻¹).

Table 1

Overview of the CCSD(T) benchmark electronic isomerization energies (ΔE_{iso}) obtained from G4(MP2) theory (in kJ mol⁻¹). All isomerization energies are calculated relative to the most stable isomer in each subset.

Subset	Number of isomers	Average isomerization energy	Smallest isomerization energy	Largest isomerization energy
$C_{17}H_{12}$	11	33.5	1.9	76.2
C19H12	12	54.6	24.5	92.6
$C_{20}H_{12}$	21	111.1	12.0	365.8
$C_{21}H_{14}$	27	27.7	7.6	65.0
$C_{22}H_{12}$	15	111.9	27.0	204.3
$C_{22}H_{14}$	12	23.0	2.3	77.7
$C_{23}H_{14}$	18	45.1	2.7	86.0
$C_{24}H_{14}$	85	121.5	7.0	355.4
C25H16	10	14.8	1.6	28.5
C26H14	16	56.4	5.5	135.2
C26H16	37	32.8	0.1	108.7
C28H14	10	76.8	12.7	288.6
C28H16	48	47.6	8.2	126.6
C36H16	13	27.7	3.2	85.4
All	335	68.3	0.1	365.8

of the sets vary between 28.5 ($C_{25}H_{16}$) and 365.8 ($C_{20}H_{12}$) kJ mol⁻¹. These results demonstrate that the PAH335 database is structurally highly diverse and provides excellent means for identifying structural and energetic trends and for evaluating the performance of computationally economical quantum chemical procedures.

It is of interest to explore the electronic properties of the PAHs in the PAH335 database that may correlate with the relative energetic stability of the isomers. Two such properties are the molecular dipole moments and the HOMO–LUMO gaps – i.e., the energy difference between the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO). We note in passing that HOMO–LUMO gaps have been found to correlate with optical band gaps of a wide range of PAHs [40,41]. Fig. 2a plots the G4(MP2) isomerization energies and the



Fig. 2. Overview of the (a) dipole moments (μ , in Debye) and (b) HOMO–LUMO gaps (ΔE_{HL} , in eV) along with the CCSD(T) G4(MP2) isomerization energies (ΔE_{iso} , in kJ mol⁻¹) for the PAHs in the PAH335 database.

dipole moments (calculated at the B3LYP/6-31G(2df,p) level of theory) for all the PAHs in the PAH335 database. It is evident that there is a poor statistical correlation between the two sets of values. Nevertheless, a closer look at this plot reveals that a dipole moment larger than 1.0 D can provide a useful indicator for the relative energy. Nearly 10% of the isomers in the PAH335 database are associated with dipole moments larger than or equal to 1.0 D. Table S2 of the Supporting Information lists the 26 PAHs with dipole moments ranging between 1.0 and 3.8 D, along with their isomerization energies. Excluding one PAH, the isomerization energies for these PAHs are relatively high or very high, ranging between 55.4 and 365.8 kJ mol⁻¹. Moreover, for nearly all the isomers in this subset, the isomerization energy exceeds 100 kJ mol^{-1} . We also note that the five isomers with $\mu > 2.0 \mbox{ D}$ are associated with very high isomerization energies ranging between 172.1 and 365.8 kJ mol⁻¹. Fig. 2b plots the G4(MP2) isomerization energies and the HOMO–LUMO gaps (ΔE_{HL}) calculated at the B3LYP/6-31G(2df,p) level of theory for all the PAHs in the PAH335 database. Again, we find a poor statistical correlation between the ΔE_{iso} and ΔE_{HL} values. However, a closer inspection of Fig. 2b reveals that all PAHs with $\Delta E_{HL} < 2.0$ eV are energetically unstable, with ΔE_{iso} values above 75 kJ mol⁻¹. Thus, we show here that either dipole moments larger than 1.0 D or HOMO--LUMO gaps below 2.0 eV indicate energetically unstable PAHs. However, as clearly illustrated in Fig. 2, it is important to emphasize that a dipole moment below 1.0 D or a HOMO-LUMO gap larger than 2.0 eV does not indicate energetic stability.

All the PAH isomers in the in the PAH335 database are unsubstituted and do not contain heteroatoms. Therefore, the finding that PAHs with dipole moments larger than 1.0 D are associated with high isomerization energies is perhaps not surprising. However, most PAHs (and indeed 92% of the PAHs in the PAH335 database) have dipole moments below 1.0 D. Thus, it may be more useful to examine the local negative charges within the PAH, since any PAH, even those with a nil or low dipole moment, may have a significant degree of atomic charge separation within the carbon skeleton of the PAH. For this purpose, we will define two useful quantities (i) the sum of all the atomic charges on the negatively charged carbons in a PAH (denoted by $q_{neg,tot}$) and (ii) the average atomic charge on the negatively charged carbons in a PAH (denoted by $q_{neg,avg}$). These quantities are defined by the following equations:

$$q_{neg,tot} = \sum_{allC} q_{neg,C} \tag{2}$$

$$q_{neg,avg} = \frac{q_{neg,tot}}{n_{neg,C}} \tag{3}$$

where $q_{neg,C}$ is the atomic charge of a negatively charged carbon, and $n_{neg,C}$ is the number of negatively charged carbons in the PAH. Here, we consider the atomic polar tensor (APT) charges [42], since they have been shown to give reliable molecular charge distributions [43,44]. We also note that the PAH335 database contains only neutral systems; therefore, the negative atomic charges are perfectly balanced by positive atomic charges. Fig. 3a plots the total atomic charge on all the negatively charged carbons $(q_{neg,tot})$ vs. the G4(MP2) isomerization energies. It is evident that PAHs for which this sum exceeds -1.0 e are highly energetic, with relative energies ranging between 70 and 370 kJ mol⁻¹ (Fig. 3a). We note that 52 PAHs in the PAH335 database are associated with $q_{neg,tot} < -1.0 e$, namely, about one out of seven PAHs. Therefore, $q_{neg,tot}$ may not serve as a generally applicable diagnostic for energetic instability. An additional drawback of $q_{neg,tot}$ is that it increases with the size of the PAH. This is illustrated by examining nanographenes of increasing size. For example, we obtain the following $q_{neg,tot}$ values -0.11 (benzene, C₆H₆), -0.22 (coronene, C₂₄H₁₂), -0.68 (circumcoronene, C₅₄H₁₈), -1.27 (circumcircumcoronene, C₉₆H₂₄), and -2.16 (C₁₅₀H₃₀) [45]. This trend in the total partial negative charges does not mean that the larger PAHs in this series become energetically unstable. Therefore, it may be more instructive to look at the average atomic charge per



Fig. 3. Overview of the (a) sum of all the atomic charges on the negatively charged carbons ($q_{neg,tot}$) and (b) average atomic charge on the negatively charged carbons ($q_{neg,avg}$) along with the CCSD(T) G4(MP2) isomerization energies (ΔE_{iso} , in kJ mol⁻¹) for the PAHs in the PAH335 database.

negatively charged carbon in a PAH, i.e., $q_{neg,avg}$ defined in eq. (3). Fig. 3b plots the $q_{neg,avg}$ values vs. the G4(MP2) isomerization energies. Fig. 3b shows that $q_{neg,avg} < -0.06 \ e$ indicates energetic instability. Practically all of the PAHs with $q_{neg,avg} < -0.06 \ e$ are associated with isomerization energies above 66 kJ mol⁻¹ (namely, 92 out of 95 PAHs). In addition, this diagnostic is able to capture nearly all PAHs with isomerization energies above 150 kJ mol⁻¹. That is, with the exception of three isomers, all the isomers associated with isomerization energies above 150 kJ mol⁻¹ are associated with $q_{neg,avg}$ values below $-0.06 \ e$ (Fig. 3b).

Having established that the average partial negative charge within a PAH ($q_{neg,avg}$) is a useful diagnostic for identifying highly energetic PAHs, we now turn to the direct calculation of the isomerization energies via computationally economical quantum chemical methods. We consider a range of DFT methods from rungs 1–4 of Jacob's Ladder [46], as well as SMO methods. Table 2 gives an overview of the performance of the DFT methods. Let us begin with one of our key findings, that the local density approximation (LDA), specifically the SVWN5 exchange–correlation functional, results in an excellent performance with a MAD of merely 3.8 kJ mol⁻¹ over the entire PAH335 database. We also note that the near-zero mean signed deviation of -0.4 kJ mol⁻¹ indicates that SVWN5 is not biased toward underestimation or overestimation of the isomerization energies.

With no exceptions, SVWN5 outperforms all the functionals from rungs 2 and 3 of Jacob's Ladder. Remarkably, SVWN5 outperforms most of the functionals from rungs 3.5 and 4 of Jacob's Ladder as well, in particular, the functionals without a dispersion correction. It is well-established that the local density approximation provides poor performance for challenging thermochemical properties such as atomization and bond dissociation energies [47]. However, it has been previously found that SVWN5 significantly outperforms many DFT methods from all rungs of Jacob's Ladder in predicting the linear alkane \rightarrow branched

Table 2

Performance of a representative set of DFT and SMO methods for the isomerization energies in the PAH335 database relative to CCSD(T) isomerization energies obtained from G4(MP2) theory (in kJ mol⁻¹).^{a.}

	MAD	MSD	RMSD
SVWN5	3.8	-0.4	6.3
BLYP	11.7	-8.0	16.0
BLYP-D4	5.1	-3.6	8.4
PBE	9.3	-7.1	13.0
PBE-D4	5.8	-5.1	9.1
TPSS	9.8	-7.1	13.5
TPSS-D4	5.1	-4.3	8.1
M06-L	4.9	-0.8	7.5
MN15-L	8.1	-5.5	10.8
B3LYP	6.5	-1.1	8.8
B3LYP-D4	3.3	2.1	4.7
PBE0	4.3	0.4	5.8
PBE0-D4	2.8	2.2	3.8
BH&HLYP	7.9	7.0	10.9
BH&HLYP-D4	9.4	9.4	12.3
CAM-B3LYP	7.1	6.4	9.8
CAM-B3LYP-D4	8.3	8.3	11.1
ωB97X	10.4	10.2	14.5
ωB97X-D	7.2	6.8	10.2
TPSSh	7.5	-4.0	10.0
TPSSh-D4	3.0	-1.5	4.8
PW6B95	4.4	-0.3	6.1
PW6B95-D4	3.0	1.3	4.0
M06	3.3	1.7	4.9
M06-2X	7.2	7.2	9.5
MN15	3.7	1.8	4.7
AM1	23.3	21.1	32.2
PM6	11.4	6.6	15.3
PM7	12.3	8.1	16.2
XTB	9.6	-9.3	13.8

 a MAD = mean absolute deviation, MSD = mean signed deviation, RMSD = root mean square deviation.

alkane isomerization energy [13]. Reference [13] shows that SVWN5 is one of the few DFT functionals that are able to predict that the branched alkane is more stable than the linear alkane across a set of alkane isomers. The current results confirm that LDA provides excellent performance for isomerization reactions across a much larger and more diverse dataset of 335 PAHs.

Let us move to the performance of the generalized gradient approximation (GGA) and *meta*-GGA methods. Both BLYP and PBE result in poor performance with MADs of 11.7 and 9.3 kJ mol⁻¹. The *meta*-GGAs TPSS and MN15-L do not improve on this performance; however, M06-L results in a significantly better performance with a MAD of 4.9 kJ mol⁻¹. With the exception of M06-L, all the GGA and *meta*-GGA methods tend to systematically underestimate the CCSD(T) benchmark isomerization energies, as demonstrated by MAD \approx –MSD. Therefore, adding a dispersion correction is expected to improve the performance. Indeed, the MADs are reduced by very significant amounts of 38–56% upon including the D4 dispersion correction. Overall, BLYP-D4 and TPSS-D4 provide similar performance to M06-L with MADs of 5.1 kJ mol⁻¹ (Table 2).

Of the global hybrid functionals, B3LYP and PBE0 emerge as the best performers with MADs of 6.5 and 4.3 kJ mol⁻¹, respectively. Similar to the GGA and *meta*-GGA methods, including the D4 dispersion correction significantly improves the performance. Namely, B3LYP-D4 and PBE0-D4 attain MADs of 3.3 and 2.8 kJ mol⁻¹, respectively. In fact, the remarkably low MAD for PBE0-D4 is not surpassed by any of the hybrid*meta*-GGA methods (*vide infra*). The range-separated hybrids CAM-B3LYP and ω B97X (and their dispersion-corrected counterparts) provide relatively poor performance with MADs ranging between 7.1 and 10.4 kJ mol⁻¹.

As expected, a number of hybrid-*meta*-GGA methods provide excellent performance. In particular, (MADs given in parentheses) TPSSh-D4 (3.0), PW6B95-D4 (3.0), M06 (3.3), and MN15 (3.7 kJ mol⁻¹). Whilst PW6B95 and PW6B95-D4, which involve 28% of exact Hartree–Fock (HF) exchange, perform well with MADs of 4.4 and 3.0 kJ mol⁻¹, respectively. The increase in the exact exchange from M06 (27% HF exchange) to M06-2X (54% HF exchange) results in a significant deterioration in performance from a MAD of 3.3 to 7.2 kJ mol⁻¹, respectively. Overall, of all the considered functionals, PBE0-D4 emerges as the best performer.

It is interesting to note that some of the MADs obtained here are very similar to those obtained for a much smaller and less diverse database of PAH isomerization energies, namely the PAH5 database, which involves only five isomerization energies [48]. In particular, the MADs for the PAH5 database are within 1.0 kJ mol⁻¹ from those obtained for the PAH335 database for the BLYP, PBE, B3LYP, and PBE0 functionals. For another two functionals, M06-L and M06, the MADs obtained for the two databases are within 2.0 kJ mol⁻¹ from each other. Table S3 of the Supporting Information compares the MADs obtained for the two databases. We note, however, that for some of the functionals that were considered in both benchmark studies (e.g., BH&HLYP, CAM-B3LYP, ωB97X, ωB97X-D, and M06-2X), the differences in MAD can vary by over 1 kcal mol⁻¹, namely by 4.3–5.6 kJ mol⁻¹. In addition, Table S3 of the Supporting Information demonstrates that the MADs obtained for the PAH335 database are, with no exceptions, larger than those obtained for the PAH5 database. These results indicate that although the PAH5 database can capture the performance obtained against a much larger and highly diverse database of PAHs for some functionals, a small database cannot reproduce the performance across a larger suite of functionals. In this context, it should be mentioned that small databases have to be carefully designed and benchmarked in order to capture the performance of larger databases [49,50].

Let us move on to the performance of the SMO methods, which are computationally much more economical than DFT methods. The parametric methods PM6 and PM7 provide similar performance with MADs of 11.4 and 12.3 kJ mol⁻¹, respectively. This is a significant improvement over the AM1 method, which attains a MAD of 23.3 kJ mol⁻¹. The

more recently developed tight binding XTB method of Grimme and coworkers attains a MAD of 9.6 kJ mol⁻¹ and emerges as the bestperforming SMO method. We note that XTB provides better performance than several DFT methods (namely, TPSS, ω B97X, and BLYP) at a significantly reduced computational cost. We, therefore, recommend that in cases where large PAHs have to be examined, initial screening with XTB followed by refined energy calculations with PBE0-D4 (or even SVWN5) is a highly effective approach.

4. Conclusions

We obtain accurate CCSD(T) isomerization energies for a large and diverse set of 335 PAHs with up to 36 carbon atoms by means of the G4 (MP2) thermochemical protocol. The database includes the following PAHs (number of isomers given in parentheses): C₁₇H₁₂ (11), C₁₉H₁₂ (12), C₂₀H₁₂ (21), C₂₁H₁₄ (27), C₂₂H₁₂ (15), C₂₂H₁₄ (12), C₂₃H₁₄ (18), C24H14 (85), C25H16 (10), C26H14 (16), C26H16 (37), C28H14 (10), C28H16 (48), and C₃₆H₁₆ (13). The PAHs in this database include 5- and 6membered rings, both sp² and sp³ carbons, and planar and nonplanar structures. The smallest PAHs in the database (C17H12) include four condensed rings whereas the largest PAHs (C₃₆H₁₆) include eleven condensed rings. As such, this database is structurally highly diverse and provides an excellent means for identifying structural and energetic trends and for benchmarking computationally economical quantum chemical procedures. The isomerization energies in the PAH335 database vary over a wide range of 0.1–365.8 kJ mol⁻¹. Similarly, the dipole moments for the PAHs spread over a wide range of 0.00-3.75 D, and the HOMO-LUMO gaps vary between 0.84 and 4.77 eV. These properties demonstrate the structural and electronic diversity in the PAH335 database. We use this dataset of accurate CCSD(T) isomerization energies for two purposes (a) identifying indicators for energetic instability and (b) identifying DFT and semiempirical methods that are able to accurately reproduce the G4(MP2) isomerization energies.

We find that a dipole moment above 1.0 D or a HOMO–LUMO gap below 2.0 eV, are general indications of energetic instability. However, a better indicator for energetic instability is the average atomic charge of the negatively charged carbons in the PAH (i.e., $q_{neg,avg}$ defined in eq. (3). We find that $q_{neg,avg}$ values below –0.06 *e* indicate with no exceptions energetic instability. Practically all the isomers associated with isomerization energies above 150 kJ mol⁻¹ are associated with $q_{neg,avg}$ values below –0.06 *e* (Fig. 3b).

A number of DFT methods are able to reproduce the CCSD(T) isomerization energies in the PAH335 database with mean-absolute deviations below the threshold of chemical accuracy (i.e., MADs < $4.2 \text{ kJ} \text{ mol}^{-1}$). Perhaps the most surprising result is that the LDA method SVWN5 attains a MAD of merely 3.8 kJ mol⁻¹ and outperforms all the considered functionals from rungs two (GGA) and three (*meta*-GGA) of Jacob's Ladder. SVWN5 also outperforms most of the functionals from rung four of Jacob's Ladder as well, in particular, the functionals without a dispersion correction. Overall, the best-performing functionals from each rung of Jacob's Ladder are:

- LDA, SVWN5, MAD = 3.8 kJ mol^{-1}
- GGA, BLYP-D4, MAD = 5.1 kJ mol^{-1}
- Meta-GGA, M06-L, MAD = 4.9 kJ mol^{-1}
- Hybrid-GGA, PBE0-D4, MAD = 2.8 kJ mol^{-1}
- Hybrid-meta-GGA, TPSSh-D4 and PW6B95-D4, MAD = 3.0 kJ mol⁻¹

Of the considered semiempirical methods, XTB emerges as the bestperforming method with a MAD of 9.6 kJ mol⁻¹. This performance is followed by PM6 and PM7 with MADs of 11.4 and 12.3 kJ mol⁻¹, respectively. Therefore, in cases where large PAHs (or a large number of PAHs) have to be examined, initial screening with XTB followed by refined energy calculations with PBE0-D4 (or even SVWN5) is recommended as a highly cost-effective approach.

CRediT authorship contribution statement

Amir Karton: Conceptualization, Methodology, Software, Formal analysis, Investigation, Resources, Writing – original draft, Writing – review & editing, Project administration. **Bun Chan:** Methodology, Formal analysis, Investigation, Writing – review & editing.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Amir Karton reports equipment, drugs, or supplies was provided by National Computational Infrastructure. Amir Karton reports a relationship with Elsevier B.V. that includes: board membership. Amir Karton is an Editor of Chemical Physics Letters but was blinded from the peerreview process for this paper.

Data availability

Data will be made available on request.

Acknowledgments

We gratefully acknowledge the generous allocation of computing time from the National Computational Infrastructure (NCI) National Facility, and system administration support provided by the Faculty of Science Agriculture Business and Law to the Linux cluster of the Karton group.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cplett.2023.140544.

References

- [1] R.G. Harvey, Polycyclic aromatic hydrocarbons, Wiley-VCH, New York, 1997.
- [2] J.E. Anthony, Functionalized Acenes and Heteroacenes for Organic Electronics, Chem. Rev. 106 (2006) 5028.
- [3] K. Kohse-Höinghaus, Combustion in the future: The importance of chemistry, Proc. Combust. Inst. 38 (1) (2021) 1–56.
- [4] A. Wahab, L. Pfuderer, E. Paenurk, R. Gershoni-Poranne, The COMPAS Project: A Computational Database of Polycyclic Aromatic Systems. Phase 1: cata-Condensed Polybenzenoid Hydrocarbons, J. Chem. Inf. Model. 62 (2022) 3704–3713.
- [5] D.C. Langreth, B.I. Lundqvist, S.D. Chakarova-Käck, V.R. Cooper, M. Dion, P. Hyldgaard, A. Kelkkanen, J. Kleis, L. Kong, S. Li, P.G. Moses, E. Murray, A. Puzder, H. Rydberg, E. Schröder, T. Thonhauser, A density functional for sparse matter, J. Phys.: Condens. Matter 21 (8) (2009) 084203.
- [6] L. Zhang, Y. Čao, N.S. Colella, Y. Liang, J.-L. Brédas, K.N. Houk, A.L. Briseno, Unconventional, Chemically Stable, and Soluble Two-Dimensional Angular Polycyclic Aromatic Hydrocarbons: From Molecular Design to Device Applications, Acc. Chem. Res. 48 (2015) 500.
- [7] L.A. Curtiss, P.C. Redfern, K. Raghavachari, Gaussian-4 theory using reduced order perturbation theory, J. Chem. Phys. 127 (2007), 124105.
- [8] L.A. Curtiss, P.C. Redfern, K. Raghavachari, Gn theory, WIREs Comput. Mol. Sci. 1 (5) (2011) 810–825.
- [9] A. Karton, N. Sylvetsky, J. M. L. Martin, L. W4-17: A diverse and high-confidence dataset of atomization energies for benchmarking high-level electronic structure methods, J. Comput. Chem. 38 (2017) 2063.
- [10] A. Karton, P.R. Schreiner, J.M.L. Martin, Heats of formation of platonic hydrocarbon cages by means of high-level thermochemical procedures, J. Comput. Chem. 37 (2016) 49.
- [11] L.-J. Yu, A. Karton, Assessment of theoretical procedures for a diverse set of isomerization reactions involving double-bond migration in conjugated dienes, Chem. Phys. 441 (2014) 166.
- [12] A. Karton, J.M.L. Martin, Explicitly correlated benchmark calculations on C₈H₈ isomer energy separations: How accurate are DFT, double-hybrid and composite ab initio procedures? Mol. Phys. 110 (2012) 2477.
- [13] A. Karton, D. Gruzman, J.M.L. Martin, Benchmark thermochemistry of the C_nH_{2n+2} alkane isomers (n = 2–8) and performance of DFT and composite ab initio methods for dispersion-driven isomeric equilibria, J. Phys. Chem. A 113 (2009) 8434.
- [14] C. Lee, W. Yang, R.G. Parr, Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density, Phys. Rev. B 37 (1988) 785.
- [15] A. D. Becke, Density-functional thermochemistry. III. The role of exact exchange, J. Chem. Phys. 98 (1993) 5648.

A. Karton and B. Chan

- [16] A. Karton, P.R. Spackman, Evaluation of Density Functional Theory for A Large and Diverse Set of Organic and Inorganic Equilibrium Structures, J. Comput. Chem. 42 (2021) 1590.
- [17] S.H. Vosko, L. Wilk, M. Nusair, Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis, Can. J. Phys. 58 (1980) 1200.
- [18] A.D. Becke, Density-Functional Exchange-Energy Approximation with Correct Asymptotic Behavior, Phys. Rev. A 38 (1988) 3098.
- [19] J.P. Perdew, K. Burke, M. Ernzerhof, Generalized Gradient Approximation Made Simple, Phys. Rev. Lett. 77 (1996) 3865.
- [20] J.M. Tao, J.P. Perdew, V.N. Staroverov, G.E. Scuseria, Phys. Rev. Lett. 91 (2003), 146401.
- [21] Y. Zhao, D.G. Truhlar, Climbing the Density Functional Ladder: Nonempirical Meta-Generalized Gradient Approximation Designed for Molecules and Solids, J. Chem. Phys. 125 (2006), 194101.
- [22] H.S. Yu, X. He, S.L. Li, D.G. Truhlar, MN15: a Kohn-Sham global-hybrid exchangecorrelation density functional with broad accuracy for multi-reference and singlereference systems and non-covalent interactions, Chem. Sci. 7 (2016) 5032.
- [23] C. Adamo, V. Barone, Toward reliable density functional methods without adjustable parameters: The PBE0 model, J. Chem. Phys. 110 (13) (1999) 6158–6170.
- [24] A.D. Becke, A new mixing of Hartree-Fock and local density functional theories, J. Chem. Phys. 98 (1993) 1372.
- [25] T. Yanai, D. Tew, N.C. Handy, A new hybrid exchange correlation functional using the Coulomb-attenuating method (CAM-B3LYP), Chem. Phys. Lett. 393 (2004) 51.
- [26] J.-D. Chai, M. Head-Gordon, Systematic optimization of long-range corrected hybrid density functionals, J. Chem. Phys. 128 (2008), 084106.
- [27] J.-D. Chai, M. Head-Gordon, Long-range corrected hybrid density functionals with damped atom-atom dispersion corrections, Phys. Chem. Chem. Phys. 10 (2008) 6615.
- [28] V.N. Staroverov, G.E. Scuseria, J. Tao, J.P. Perdew, Comparative Assessment of a New Nonempirical Density Functional: Molecules and Hydrogen-Bonded Complexes, J. Chem. Phys. 119 (2003) 12129.
- [29] Y. Zhao, D.G. Truhlar, Design of Density Functionals That Are Broadly Accurate for Thermochemistry, Thermochemical Kinetics, and Nonbonded Interactions, J. Phys. Chem. A 109 (2005) 5656.
- [30] Y. Zhao, D.G. Truhlar, The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals, Theor. Chem. Acc. 120 (2008) 215.
- [31] F. Weigend, R. Ahlrichs, Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy, Phys. Chem. Chem. Phys. 7 (18) (2005) 3297.
- [32] E. Caldeweyher, S. Ehlert, A. Hansen, H. Neugebauer, S. Spicher, C. Bannwarth, S. Grimme, A generally applicable atomic-charge dependent London dispersion correction, J. Chem. Phys. 150 (2019), 154112.
- [33] M.J.S. Dewar, E.G. Zoebisch, E.F. Healy, J.J.P. Stewart, Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model, J. Am. Chem. Soc. 107 (1985) 3902.
- [34] J.J.P. Stewart, Optimization of parameters for semiempirical methods V: Modification of NDDO approximations and application to 70 elements, J. Mol. Model. 13 (2007) 1173.

- [35] J.J.P. Stewart, Optimization of parameters for semiempirical methods VI: more modifications to the NDDO approximations and re-optimization of parameters, J. Mol. Model. 19 (2013) 1.
- [36] C. Bannwarth, E. Caldeweyher, S. Ehlert, A. Hansen, P. Pracht, J. Seibert, S. Spicher, S. Grimme, Extended tight-binding quantum chemistry methods, WIREs Comput. Mol. Sci. 11 (2020) e01493.
- [37] H.-J. Werner, P.J. Knowles, G. Knizia, F.R. Manby, M. Schütz, Molpro: a generalpurpose quantum chemistry program package, WIREs Comput. Mol. Sci. 2 (2012) 242.
- [38] Gaussian 16, Revision C.01, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D. J. Fox,, See also, Gaussian Inc, Wallingford CT, 2016.
- [39] The NIST PAH Database is available here: https://pah.nist.gov/text/about.html.
- [40] A. Menon, J.A.H. Dreyer, J.W. Martin, J. Akroyd, J. Robertsonc, M. Kraft, Optical band gap of cross-linked, curved, and radical polyaromatic hydrocarbons, Phys. Chem. Chem. Phys. 21 (2019) 16240.
- [41] Y. Xu, Q. Chu, D. Chen, A. Fuentes, HOMO–LUMO Gaps and Molecular Structures of Polycyclic Aromatic Hydrocarbons in Soot Formation, Front. Mech. Eng. 7 (2021), 744001.
- [42] J. Ciosłowski, A new population analysis based on atomic polar tensors, J. Am. Chem. Soc. 111 (1989) 8333.
- [43] F. De Proft, J.M.L. Martin, P. Geerlings, On the performance of density functional methods for describing atomic populations, dipole moments and infrared intensities, Chem. Phys. Lett. 250 (1996) 393.
- [44] M. Cho, N. Sylvetsky, S. Eshafi, G. Santra, I. Efremenko, J.M.L. Martin, The Atomic Partial Charges Arboretum: Trying to See the Forest for the Trees, ChemPhysChem 21 (2020) 688.
- [45] A. Karton, π–π interactions between benzene and graphene by means of large-scale DFT-D4 calculations, Chem. Phys. 561 (2022), 111606.
- [46] J.P. Perdew, K. Schmidt, Jacob's ladder of density functional approximations for the exchange-correlation energy, AIP Conf. Proc. 577 (2000) 1.
- [47] A. Karton, S. Daon, J.M.L. Martin, W4–11: A high-confidence dataset for computational thermochemistry derived from W4 ab initio data, Chem. Phys. Lett. 510 (2011) 165.
- [48] A. Karton, How reliable is DFT in predicting the relative energies of polycyclic aromatic hydrocarbon isomers? Comparison of functionals from different rungs of Jacob's Ladder, J. Comput. Chem. 38 (2017) 370.
- [49] T. Gould, 'Diet GMTKN55' offers accelerated benchmarking through a representative subset approach, Phys. Chem. Chem. Phys. 20 (2018) 27735.
- [50] B.J. Lynch, D.G. Truhlar, Small representative benchmarks for thermochemical calculations, J. Phys. Chem. A 107 (2003) 8996.