'Diet GMTKN55' offers accelerated benchmarking through a representative subset approach[†]

Tim Gould,^a

Received Date
Accepted Date

DOI: 10.1039/xxxxxxxxxx

www.rsc.org/journalname

The GMTKN55 benchmarking protocol introduced by [Goerigk *et al.*, *Phys. Chem. Chem. Phys.*, 2017, 19, 32184] allows comprehensive analysis and ranking of density functional approximations with diverse chemical behaviours. But this comprehensiveness comes at a cost: GMTKN55's 1500 benchmarking values require energies for around 2500 systems to be calculated, making it a costly exercise. This manuscript introduces three subsets of GMTKN55, consisting of 30, 100 and 150 systems, as 'diet' substitutes for the full database. The subsets are chosen via a stochastic genetic approach, and consequently can reproduce key results of the full GMTKN55 database, including ranking of approximations.

Density functional theory (DFT) has established itself as a dominant approach for computational modelling of chemical systems. But this success has come at a cost – there is now a “zoo” of hundreds, if not thousands of approximations to choose from.^{1–4} Selecting a functional (here called a density functional approximation, DFA) has thus become a highly non-trivial task. This task is made harder by selection bias in scientific publishing: it is rare to find a new approach that doesn't claim to solve some important problem or other, whereas there is a delay until problems introduced by new approaches are publicised. Thus, there is significant trial and error in determining if a newly introduced (or indeed old) DFA is likely to offer systematic improvements over existing methods.

Benchmarking studies, which rank DFAs on their ability to reproduce key chemical physics calculated at a higher level of theory have thus become an integral part of method selection. The recently introduced “general main group thermochemistry, kinetics, and non-covalent interactions” (GMTKN55) benchmark database⁵ establishes an unsurpassed level of rigour in estab-

lishing the *overall* quality of DFAs, and must be considered the state-of-art of benchmarking for chemical systems. The reactions, atomisation energies, and other energy differences included in the database incorporate an extensive range of important chemical processes. The ranking systems introduced in that work critically evaluate DFAs on a comprehensive range of tests using a set of 1500 energy differences. Approaches which perform well on these tests should thus be expected to perform well in main group chemistry generally.

But to test a given method against GMTKN55 requires almost 2500 single point calculations be carried out. While this number of calculations is quite feasible in modern, well-developed quantum chemistry codes, it is more difficult to achieve using the planewave codes typically used in material science problems. Furthermore, methods that are highly innovative are often implemented inefficiently at first, with optimisation coming only after initial successes (e.g., the work of Román-Pérez and Soler⁶ for vdW-DFs⁷). Comprehensive testing across GMTKN55 may thus be infeasible or impossible at precisely the stage when it is most needed – when methods are being designed.

There is therefore a clear need for smaller databases that can be used to accurately assess methods for which comprehensive testing is infeasible. Ideally, such a database should not sacrifice rigour, and must capture the key features of existing state-of-art. This work establishes that a key ranking system of GMTKN55, called WTMAD-2, can be accurately reproduced using just 100 systems selected from the full database, with reproduction improving/degrading slightly when 150/30 systems are used.

The rest of paper is arranged as follows: first, the process used to find the selected systems is described. Then, some results are presented showing how well the smaller sets reproduce results from the full database. The list of 100 systems is then presented and discussed, with the lists for 30 and 150 systems included in the electronic supplementary information (ESI). Finally, the paper concludes with a secondary important feature of the reduced database – namely, its potential use in a more physically comprehensive benchmark database that includes systems in the solid

^a Qld Micro- and Nanotechnology Centre, Griffith University, Nathan, Qld 4111, Australia; E-mail: t.gould@griffith.edu.au

[†] Electronic Supplementary Information (ESI) available: [details of any supplementary information available should be included here]. See DOI: 10.1039/b000000x/

This work draws from the full GMTKN55 database provided on its official website⁸. This data was scraped, and then all errors stored as ΔE_I^d in a table on DFA d and system I indices. No new calculations were carried out. Interestingly, this process returned just 1499 systems, not the 1505 identified in the original paper. However, WTMAD-2 results calculated with the scraped data were within 2.1% of those reported on the GMTKN55 database, with most well inside 1%, making it more than fit for purpose. The one exception was results for BLYP (and its dispersion corrected counterparts), which were consequently discarded, giving us 213 DFAs, not 216.

$$\text{MAD}(d) = \frac{1}{N_{\text{GMTKN55}}} \sum_I W_I |\Delta E_I^d - \Delta E_I| \quad (1)$$
$$W_I = \frac{56.84 \text{ kcal mol}^{-1}}{N_{B(I)}^{-1} \sum_{J \in B(I)} |\Delta E_J|}, \quad (2)$$

This data then had to be condensed to a representative set of N_s systems, where N_s is the target subset size. This process involved two optimisation phases, both loosely based on evolution:

1. First, a primeval soup was introduced, in which “species” $S = \{I_1, I_2, \dots, I_{N_s}\}$ comprised of N_s systems are selected at random. The mean absolute deviation (MAD) was then calculated for the species, and an error

$$\text{Err}(S) = \frac{1}{N_{\text{DFAs}}} \sum_{d \in \text{DFAs}} |\text{MAD}_S(d) - \text{MAD}(d)| \quad (3)$$

assigned to each species. Here $\text{MAD}_S(d) = \frac{1}{N_S} \sum_{I \in S} w_I |\Delta E_I^d - \Delta E_I|$ is the MAD for the subset. The full set $\text{MAD}(d)$ is defined in (1), with w_I defined in (2). The MAD of MADs is then taken over all DFAs d , to obtain a single number.

If S was lower in error than the previous lowest error system, it was allowed to live, and joined a set $L = \{S_1, S_2, \dots\}$ of live contenders, where $\text{Err}(S_{i+1}) < \text{Err}(S_i)$ by construction. Typically 50,000 random selection yielded about 10-20 elements in L .

2. In the next phase, breeding and survival of the fittest were introduced. Two elements, P_1 and P_2 of the live species L were selected at random to serve as parents. These produced a child, C , composed of half the elements I of P_1 , selected at random, and half the elements of P_2 , ensuring no duplication. This process is analogous to sharing genes.

Fig. 1 Comparison of errors found using $N_s = 30, 100, 150$ (top to bottom) subsets compared with the full database. Here, results are divided into four categories: double-hybrid DFAs (reds), hybrid DFAs (greens), GGAs (blues) and meta-GGAs (earth tones). The paler colour in each category represent MADs for the subset, and the darker colour is for the full database. The blended colour represents overlap. Only the ten best methods are shown in each category.

Errors [using (3)] were then calculated for both parents and the child.

In the event that the child had a lower error than one parent, the parent with the highest error was discarded, and replaced by the child, e.g., $L = \{\dots P_1, \dots P_2, \dots\} \rightarrow \{\dots C, \dots P_2, \dots\}$ when $C < P_1$. Otherwise the child died and L was left unchanged. This cycle was repeated over 50,000 generations, allowing the population to evolve to lower errors. Eventually, the species with the highest fitness (overall lowest error) was selected, leaving the error quite small.

The optimisation was thus able to evolve to a subset of N_s systems, specifically, the most fit species found above, that could reproduce $\text{MAD}(d)$ with good accuracy across all DFAs d . Calculations take a matter of minutes on a laptop, and the best runs out of multiple trials were taken as the final lists. The accuracy of the sets reported here, with $N_s = 30, 100, 150$ are respectively $\text{Err}(S_{30}) = 0.362$, $\text{Err}(S_{100}) = 0.165$ and $\text{Err}(S_{150}) = 0.126$, or 3.9%, 1.8% and 1.4% of the average MAD, $\frac{1}{N_{\text{DFAs}}} \sum_{d \in \text{DFAs}} \text{MAD}(d)$.

Figure 1 shows WTMA2 errors for the full database and the various subsets found via the evolutionary algorithm. As in the

Table 1 Kendall rank correlation coefficients τ [eq. (4)] for each subset, separated by approach. Values close to one indicate good agreement between rankings from the full database, and the chosen subset.

N_s	Double-hybrid	Hybrid	GGA	meta-GGA
30	0.883	0.854	0.901	0.789
100	0.860	0.933	0.948	0.918
150	0.930	0.938	0.967	0.906

original GMTKN55 paper, we separate the approaches into: i) double-hybrids involving mixtures of Hartree-Fock (HF), post-HF correlation and density functional (DF) terms; ii) hybrids involving HF and DF terms; iii) generalised gradient approximations (GGAs) involving the density and its gradient only; iv) meta-GGAs which involve densities, gradients and kinetic energy density terms. This division reflects the broad families of low-cost modern quantum chemistry approaches.

The high quality of the subsets is evident in the plot, which shows the subset database and full database giving very similar values and rankings across all types of methods. It is clear that rankings from the reduced subsets can be used as a substitute for the full set, albeit with some errors, especially for the small 30 element set. Additional plots showing the worst cases are included in the ESI, and are of similar quality. Kendall rank correlation coefficients⁵²

$$\tau = \frac{\sum_{d \neq d'} \text{sgn}[\text{MAD}(d) - \text{MAD}(d')] \text{sgn}[\text{MAD}_S(d) - \text{MAD}_S(d')]}{n(n-1)}, \quad (4)$$

are reported in Table 1 to quantify agreement between the full database and the subsets. These highlight the general accuracy of the subsets as a ranking tool, as the worst case is 79% for $N_s = 30$, which still indicates a strong degree of correlation in the rankings from the subsets and the full database.

Finally, Table 2 shows the full set of systems used in the $N_s = 100$ data set. The systems included come from 41 benchmark sets, 75% of the full list of 55, and involve between one and nine elements from each set. The high fraction of sets that are sampled is interesting, and suggests that the evolutionary algorithm finds key systems that exemplify the different chemical physics scrutinised in the various benchmark sets that form the full GMTKN55 database. The fraction of benchmark sets sampled for $N_s = 30$ is 20%, while 89% of sets are sampled for $N_s = 150$. This result may indicate a direction to explore for future improvements in benchmarking, as it suggests that selected case studies can yield almost as much information as results for the full set, at least when combined with other tests. Furthermore, some benchmark sets are never sampled in the reduced subsets, suggesting they might duplicate information from other sets.

Due to the reduced number of systems, an assessment using one of the subsets reported here would involve, at worst, a few hundred calculations (specifically, 82/240/336 calculations for 30/100/150 systems), a small fraction of the 2462 calculations required for the full GMTKN55 database. This is a substantial saving in terms of time, with little cost to accuracy if only rankings and overall quality estimates are required. The 30 element

subset is probably too inaccurate to be used on its own for detailed assessment, however, but may be sufficient when used on combination with other tests.

To conclude, this work derives and presents ‘diet’ versions of the GMTKN55 database for use in future benchmarking studies. The reduced sets, involving 30, 100 and 150 energy differences, are able to reproduce not just quantitative errors, but also rankings, with a generally good degree of accuracy. They are thus able to serve as reliable proxies for the full GMTKN55 database, to be used when, e.g., calculations are very expensive. Density functional developers who cannot afford to test on the full GMTKN55 set are thus advised to use the reduced sets reported here, to establish a metric of relative success that is comparable to best practice.

That said, it must be stressed that GMTKN55 has far greater analytic power than any subset thereof, and its use is certainly advised whenever possible. The reduced sets introduced here are not intended to replace GMTKN55, but to supplement it. One possible application would be to use a subset for pre-screening, e.g. while developing a new approach, followed by tests on GMTKN55 once best candidates have been established. Another possibility might be to use a subset for optimisation of empirical methods.

Another important benefit of the reduced sets, and in fact the initial driving motivation behind this work, is that it offers a route to a *truly comprehensive benchmarking protocol* for general chemical physics that covers materials science and solid state problems, in addition to the comprehensive gas phase chemistry results in GMTKN55. To understand why the ‘diet’ sets will help here, note that it is unlikely that we will have 1500 solid phase benchmark standard energy differences in the near future, due to: a) the increased methodological challenges posed by such systems (e.g., many methods fail for small and zero gap systems), and b) the significantly higher numerical cost of calculating such systems to benchmark accuracy. One hundred diverse systems (or perhaps just thirty, initially) is a much more feasible size for a solid state benchmark set. To avoid sample size bias, such a set must thus be paired with a comparably sized, but representative, sample of standard chemistry, such as the ones presented here. Work along these lines is being pursued.

References

- 1 K. Burke, *J. Chem. Phys.*, 2012, **136**, 150901.
- 2 A. D. Becke, *J. Chem. Phys.*, 2014, **140**, 18A301.
- 3 R. O. Jones, *Rev. Mod. Phys.*, 2015, **87**, 897.
- 4 S. Grimme and P. R. Schreiner, *Angew. Chem. Int. Ed.*, 2018, **57**, 4170–4176.
- 5 L. Goerigk, A. Hansen, C. Bauer, S. Ehrlich, A. Najibi and S. Grimme, *Phys. Chem. Chem. Phys.*, 2017, **19**, 32184–32215.
- 6 G. Román-Pérez and J. M. Soler, *Phys. Rev. Lett.*, 2009, **103**, 096102.
- 7 D. C. Langreth, M. Dion, H. Rydberg, E. Schröder, P. Hyldgaard and B. I. Lundqvist, *Int. J. Quantum Chem.*, 2005, **101**, 599–610.
- 8 GMTKN55 - A database for general main group thermochemistry, kinetics, and non-covalent interactions, 2018, <https://www.chemie.uni-bonn.de/pctc/mulliken-center/software/GMTKN/gmtkn55>.
- 9 K. U. Lao, R. Schäffer, G. Jansen and J. M. Herbert, *J. Chem. Theory Comput.*, 2015, **11**, 2473–2486.
- 10 H. Yu and D. G. Truhlar, *J. Chem. Theory Comput.*, 2015, **11**, 2968–2983.
- 11 Y. Zhao, B. J. Lynch and D. G. Truhlar, *Phys. Chem. Chem. Phys.*, 2005, **17**, 43–52.
- 12 Y. Zhao, N. González-García and D. G. Truhlar, *J. Phys. Chem. A*, 2005, **109**, 2012–2018.
- 13 A. Karton and L. Goerigk, *J. Comput. Chem.*, 2015, **36**, 622–632.

- 14 S. N. Steinmann, G. Csonka and C. Carminboeuf, *J. Chem. Theory Comput.*, 2009, **5**, 2950–2958.
- 15 S. Kozuch, S. M. Bachrach and J. M. Martin, *J. Phys. Chem. A*, 2014, **118**, 293–303.
- 16 L.-J. Yu and A. Karton, *Chem. Phys.*, 2014, **441**, 166–177.
- 17 J. Friedrich and J. Hähnchen, *J. Chem. Theory Comput.*, 2013, **9**, 5381–5394.
- 18 L. A. Curtiss, K. Raghavachari, G. W. Trucks and J. A. Pople, *J. Chem. Phys.*, 1991, **94**, 7221–7230.
- 19 L. Goerigk and S. Grimme, *J. Chem. Theory Comput.*, 2010, **6**, 107–126.
- 20 L. A. Curtiss, K. Raghavachari, P. C. Redfern and J. A. Pople, *J. Chem. Phys.*, 1997, **106**, 1063–1079.
- 21 S. Kozuch and J. M. L. Martin, *J. Chem. Theory Comput.*, 2013, **9**, 1918–1931.
- 22 J. Rezac, K. E. Riley and P. Hobza, *J. Chem. Theory Comput.*, 2012, **8**, 4285–4292.
- 23 S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *J. Chem. Phys.*, 2010, **132**, 154104.
- 24 S. Zahn, D. R. MacFarlane and E. I. Izgorodina, *Phys. Chem. Chem. Phys.*, 2013, **15**, 13664–13675.
- 25 L. Goerigk and R. Sharma, *Can. J. Chem.*, 2016, **94**, 1133–1143.
- 26 S. Grimme, M. Steinmetz and M. Korth, *J. Org. Chem.*, 2007, **72**, 2118–2126.
- 27 R. Huenerbein, B. Schirmer, J. Moellmann and S. Grimme, *Phys. Chem. Chem. Phys.*, 2010, **12**, 6940–6948.
- 28 M. Korth and S. Grimme, *J. Chem. Theory Comput.*, 2009, **5**, 993–1003.
- 29 U. R. Fogueri, S. Kozuch, A. Karton and J. M. L. Martin, *J. Phys. Chem. A*, 2013, **117**, 2269–2277.
- 30 D. Řeřha, H. Valdes, J. Vondrasek, P. Hobza, A. Abu-Riziq, B. Crews and M. S. de Vries, *Chem. - Eur. J.*, 2005, **11**, 6803–6817.
- 31 D. Setiawan, E. Kraka and D. Cremer, *J. Phys. Chem. A*, 2015, **119**, 1642–1656.
- 32 S. Grimme, *Angew. Chem. Int. Ed.*, 2013, **52**, 6306–6312.
- 33 F. Neese, T. Schwabe, S. Kossmann, B. Schirmer and S. Grimme, *J. Chem. Theory Comput.*, 2009, **5**, 3060–3073.
- 34 J. Řezáč, K. E. Riley and P. Hobza, *J. Chem. Theory Comput.*, 2011, **7**, 2427–2438.
- 35 G. I. Csonka, A. D. French, G. P. Johnson and C. A. Stortz, *J. Chem. Theory Comput.*, 2009, **5**, 679–692.
- 36 H. Kruse, A. Mladek, K. Gkionis, A. Hansen, S. Grimme and J. Sponer, *J. Chem. Theory Comput.*, 2015, **11**, 4972–4991.
- 37 A. Karton, S. Daon, J. M. L. Martin and B. Ruscic, *Chem. Phys. Lett.*, 2011, **510**, 165–178.
- 38 V. S. Bryantsev, M. S. Diallo, A. C. T. van Duin and W. A. Goddard III, *J. Chem. Theory Comput.*, 2009, **5**, 1016–1026.
- 39 Y. Zhao, H. T. Ng, R. Peverati and D. G. Truhlar, *J. Chem. Theory Comput.*, 2012, **8**, 2824–2834.
- 40 L. Goerigk and S. Grimme, *J. Chem. Theory Comput.*, 2009, **6**, 107–126.
- 41 L. Goerigk and S. Grimme, *J. Chem. Theory Comput.*, 2010, **7**, 291–309.
- 42 L. Goerigk and S. Grimme, *Phys. Chem. Chem. Phys.*, 2011, **13**, 6670–6688.
- 43 P. Jurečka, J. Sponer, J. Cerny and P. Hobza, *Phys. Chem. Chem. Phys.*, 2006, **8**, 1985–1993.
- 44 L. Goerigk and S. Grimme, *J. Chem. Theory Comput.*, 2011, **7**, 291–309.
- 45 S. Grimme, H. Kruse, L. Goerigk and G. Erker, *Angew. Chem. Int. Ed.*, 2010, **49**, 1402–1405.
- 46 R. Sure, A. Hansen, P. Schwerdtfeger and S. Grimme, *Phys. Chem. Chem. Phys.*, 2017, **19**, 14296–14305.
- 47 A. Karton, R. J. O'Reilly, B. Chan and L. Radom, *J. Chem. Theory Comput.*, 2012, **8**, 3128–3136.
- 48 A. Karton, R. J. O'Reilly and L. Radom, *J. Phys. Chem. A*, 2012, **116**, 4211–4221.
- 49 T. Schwabe and S. Grimme, *Phys. Chem. Chem. Phys.*, 2007, **9**, 3397–3406.
- 50 S. Grimme, *Angew. Chem. Int. Ed.*, 2006, **45**, 4460–4464.
- 51 T. Takatani, E. G. Hohenstein, M. Malagoli, M. S. Marshall and C. D. Sherrill, *J. Chem. Phys.*, 2010, **132**, 144104.
- 52 M. G. Kendall, *Biometrika*, 1938, **30**, 81–93.

Table 2 Full list of benchmark sets and specific elements used in the $N_s = 100$ 'diet GMTKN55' ranking set. These structures are sufficient to reproduce the full WTMAD-2 ranking with good accuracy, using the weights W listed here. Results for $N_s = 30$ and 150 are presented in the electronic supplementary information (ESI)[‡].

Set	W	IDs	Set	W	IDs	Set	W	IDs
ACONF [†]	30.99	2, 5, 8	ADIM6 [†]	16.93	2	AHB21 ⁹	2.53	1, 15
ALKBDE10 ¹⁰	0.56	1	Amino20x4 [†]	23.31	23, 48, 56	BH76 ^{11,12}	3.05	6, 18, 54, 74
BH76RC [†]	2.66	16	BHPERI ¹³	2.72	1, 17, 23, 26	BHROT27 [†]	9.06	1, 10, 20
BSR36 ¹⁴	3.51	4, 21	BUT14DIOL ¹⁵	20.30	28, 39, 63	CARBHB12 [†]	9.42	3, 4, 6, 11
CDIE20 ¹⁶	14.02	2, 14	CHB6 ⁹	2.12	1	DC13 [†]	1.03	6, 12
FH51 ¹⁷	1.83	25	G21EA ¹⁸	1.69	5, 21, 25	G21IP ^{18,19}	0.22	28
G2RC ²⁰	1.11	15	HAL59 ^{21,22†}	12.38	1, 8, 21, 38, 43, 56	HEAVY28 ²³	45.79	5, 24
ICONF [†]	17.40	16	IL16 ^{9,24}	0.52	3	INV24 ²⁵	1.78	12, 19, 22
ISO34 ²⁶	3.90	7, 12	ISOL24 ²⁷	2.59	13	MB16-43 ²⁸	0.12	4, 31
MCONF ²⁹	11.43	7, 22, 30, 42	PArel [†]	12.28	3, 5, 10, 20	PCONF21 ³⁰	35.05	7, 8, 18
PNICO23 ³¹	13.30	13, 17	RC21 ³²	1.59	4	RG18 [†]	98.00	1, 15
RSE43 ³³	7.48	28, 30, 43	S66 ³⁴	10.40	11, 17, 24, 31, 45	SCONF ³⁵	12.36	10
SIE4x4 [†]	1.69	7, 10	UPU23 ³⁶	9.93	5	W4-11 ³⁷	0.19	12, 37, 39, 58, 85, 92, 104, 136, 138
WATER27 ³⁸	0.70	17, 20	YBDE18 ³⁹	1.15	1			

[†] Contains systems from one of the GMTKN databases^{5,40–42}

[‡] These include additional benchmark sets^{43–51}

Supplementary Table 1 Full list of benchmark sets and specific elements used in the $N_s = 30$ diet-GMTKN55 ranking set. These structures are sufficient to reproduce the full WTMAD-2 ranking with good accuracy, using the weights W listed here..

Set	W	IDs	Set	W	IDs	Set	W	IDs	Set	W	IDs
ACONF [†]	30.99	10	Amino20x4 [†]	23.31	28, 54	BH76 ^{11,12}	3.05	5	BHPERI ¹³	2.72	11
BHROT27 [†]	9.06	16, 26	BSR36 ¹⁴	3.51	31	BUT14DIOL ¹⁵	20.30	13	CDIE20 ¹⁶	14.02	9
DC13 [†]	1.03	1	DIPCS10 [†]	0.09	7	FH51 ¹⁷	1.83	24, 30	G21EA ¹⁸	1.69	14, 25
HAL59 ^{21,22,†}	12.38	40, 57	HEAVY28 ²³	45.79	16	MB16-43 ²⁸	0.12	10	MCONF ²⁹	11.43	1
PNICO23 ³¹	13.30	16	PX13 ⁴⁷	1.70	9	S66 ³⁴	10.40	6, 50	SIE4x4 [†]	1.69	15
W4-11 ³⁷	0.19	30, 57, 132	WCPT18 ⁴⁸	1.62	15						

[†] Contains systems from one of the GMTKN databases^{5,40–42}

Supplementary Table 2 As per Supplementary Table 1, but for $N_s = 150$.

Set	W	IDs	Set	W	IDs
ACONF [†]	30.99	11, 14	AHB21 ⁹	2.53	6, 10, 15
AL2X6 [†]	1.58	5	Amino20x4 [†]	23.31	4, 6, 15, 25, 39, 41, 46, 68
BH76 ^{11,12}	3.05	8, 36, 37, 47, 50, 53, 66	BH76RC [†]	2.66	5
BHPERI ¹³	2.72	2, 8, 25	BHROT27 [†]	9.06	3, 11, 14, 21
BSR36 ¹⁴	3.51	17, 22	BUT14DIOL ¹⁵	20.30	2, 29, 32, 38, 39, 42, 43, 53
C60ISO ⁴⁶	0.58	7, 9	CDIE20 ¹⁶	14.02	1, 18, 20
CHB6 ⁹	2.12	2	DARC [†]	1.75	11
DC13 [†]	1.03	10, 13	DIPCS10 [†]	0.09	2
FH51 ¹⁷	1.83	26, 27, 40, 48	G21EA ¹⁸	1.69	17, 20, 23
G21IP ^{18,19}	0.22	6, 15, 31	G2RC ²⁰	1.11	16, 24, 25
HAL59 ^{21,22,†}	12.38	7, 11, 16, 28, 32, 56	HEAVY28 ²³	45.79	3, 16, 24
ICONF [†]	17.40	16	IDISP ^{19,26,44,49,50}	4.00	3
IL16 ^{9,24}	0.52	11, 14	INV24 ²⁵	1.78	1, 14, 15
ISO34 ²⁶	3.90	21	ISOL24 ²⁷	2.59	2, 4, 20
MB16-43 ²⁸	0.12	14	MCONF ²⁹	11.43	11, 20, 30
NBPRC ^{19,45}	2.30	1	PA26 [†]	0.30	20, 25
PArel [†]	12.28	13	PCONF21 ³⁰	35.05	3
PNICO23 ³¹	13.30	7, 11, 13, 21	PX13 ⁴⁷	1.70	8
RC21 ³²	1.59	5, 6, 8, 16, 20	RG18 [†]	98.00	11, 13, 15
RSE43 ³³	7.48	14, 40	S22 ^{43,51}	7.78	3, 8, 11
S66 ³⁴	10.40	8, 17, 25, 28, 42, 47, 61	SCONF ³⁵	12.36	16
SIE4x4 [†]	1.69	15	TAUT15 [†]	18.66	5, 13
UPU23 ³⁶	9.93	2, 4, 9, 15	W4-11 ³⁷	0.19	2, 7, 30, 31, 38, 58, 63, 73, 77, 85, 86, 95, 102, 106, 129
WATER27 ³⁸	0.70	2, 17	WCPT18 ⁴⁸	1.62	6, 8, 9, 18
YBDE18 ³⁹	1.15	6, 8, 9, 13, 16			

[†] Contains systems from one of the GMTKN databases^{5,40–42}

Supplementary References

- [1] K. Burke, *J. Chem. Phys.*, 2012, **136**, 150901.
- [2] A. D. Becke, *J. Chem. Phys.*, 2014, **140**, 18A301.
- [3] R. O. Jones, *Rev. Mod. Phys.*, 2015, **87**, 897.
- [4] S. Grimme and P. R. Schreiner, *Angew. Chem. Int. Ed.*, 2018, **57**, 4170–4176.
- [5] L. Goerigk, A. Hansen, C. Bauer, S. Ehrlich, A. Najibi and S. Grimme, *Phys. Chem. Chem. Phys.*, 2017, **19**, 32184–32215.
- [6] G. Román-Pérez and J. M. Soler, *Phys. Rev. Lett.*, 2009, **103**, 096102.
- [7] D. C. Langreth, M. Dion, H. Rydberg, E. Schröder, P. Hyldgaard and B. I. Lundqvist, *Int. J. Quantum Chem.*, 2005, **101**, 599–610.
- [8] GMTKN55 - A database for general main group thermochemistry, kinetics, and non-covalent interactions, 2018, <https://www.chemie.uni-bonn.de/pctc/mulliken-center/software/GMTKN/gmtkn55>.
- [9] K. U. Lao, R. Schäffer, G. Jansen and J. M. Herbert, *J. Chem. Theory Comput.*, 2015, **11**, 2473–2486.
- [10] H. Yu and D. G. Truhlar, *J. Chem. Theory Comput.*, 2015, **11**, 2968–2983.
- [11] Y. Zhao, B. J. Lynch and D. G. Truhlar, *Phys. Chem. Chem. Phys.*, 2005, **17**, 43–52.
- [12] Y. Zhao, N. González-García and D. G. Truhlar, *J. Phys. Chem. A*, 2005, **109**, 2012–2018.
- [13] A. Karton and L. Goerigk, *J. Comput. Chem.*, 2015, **36**, 622–632.

- [14] S. N. Steinmann, G. Csonka and C. Carminboeuf, *J. Chem. Theory Comput.*, 2009, **5**, 2950–2958.
- [15] S. Kozuch, S. M. Bachrach and J. M. Martin, *J. Phys. Chem. A*, 2014, **118**, 293–303.
- [16] L.-J. Yu and A. Karton, *Chem. Phys.*, 2014, **441**, 166–177.
- [17] J. Friedrich and J. Hänchen, *J. Chem. Theory Comput.*, 2013, **9**, 5381–5394.
- [18] L. A. Curtiss, K. Raghavachari, G. W. Trucks and J. A. Pople, *J. Chem. Phys.*, 1991, **94**, 7221–7230.
- [19] L. Goerigk and S. Grimme, *J. Chem. Theory Comput.*, 2010, **6**, 107–126.
- [20] L. A. Curtiss, K. Raghavachari, P. C. Redfern and J. A. Pople, *J. Chem. Phys.*, 1997, **106**, 1063–1079.
- [21] S. Kozuch and J. M. L. Martin, *J. Chem. Theory Comput.*, 2013, **9**, 1918–1931.
- [22] J. Rezac, K. E. Riley and P. Hobza, *J. Chem. Theory Comput.*, 2012, **8**, 4285–4292.
- [23] S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *J. Chem. Phys.*, 2010, **132**, 154104.
- [24] S. Zahn, D. R. MacFarlane and E. I. Izgorodina, *Phys. Chem. Chem. Phys.*, 2013, **15**, 13664–13675.
- [25] L. Goerigk and R. Sharma, *Can. J. Chem.*, 2016, **94**, 1133–1143.
- [26] S. Grimme, M. Steinmetz and M. Korth, *J. Org. Chem.*, 2007, **72**, 2118–2126.
- [27] R. Huenerbein, B. Schirmer, J. Moellmann and S. Grimme, *Phys. Chem. Chem. Phys.*, 2010, **12**, 6940–6948.
- [28] M. Korth and S. Grimme, *J. Chem. Theory Comput.*, 2009, **5**, 993–1003.
- [29] U. R. Fogueri, S. Kozuch, A. Karton and J. M. L. Martin, *J. Phys. Chem. A*, 2013, **117**, 2269–2277.
- [30] D. Řeřha, H. Valdes, J. Vondrasek, P. Hobza, A. Abu-Riziq, B. Crews and M. S. de Vries, *Chem. - Eur. J.*, 2005, **11**, 6803–6817.
- [31] D. Setiawan, E. Kraka and D. Cremer, *J. Phys. Chem. A*, 2015, **119**, 1642–1656.
- [32] S. Grimme, *Angew. Chem. Int. Ed.*, 2013, **52**, 6306–6312.
- [33] F. Neese, T. Schwabe, S. Kossmann, B. Schirmer and S. Grimme, *J. Chem. Theory Comput.*, 2009, **5**, 3060–3073.
- [34] J.Řezáč, K. E. Riley and P. Hobza, *J. Chem. Theory Comput.*, 2011, **7**, 2427–2438.
- [35] G. I. Csonka, A. D. French, G. P. Johnson and C. A. Stortz, *J. Chem. Theory Comput.*, 2009, **5**, 679–692.
- [36] H. Kruse, A. Mladek, K. Gkionis, A. Hansen, S. Grimme and J. Sponer, *J. Chem. Theory Comput.*, 2015, **11**, 4972–4991.
- [37] A. Karton, S. Daon, J. M. L. Martin and B. Ruscic, *Chem. Phys. Lett.*, 2011, **510**, 165–178.
- [38] V. S. Bryantsev, M. S. Diallo, A. C. T. van Duin and W. A. Goddard III, *J. Chem. Theory Comput.*, 2009, **5**, 1016–1026.
- [39] Y. Zhao, H. T. Ng, R. Peverati and D. G. Truhlar, *J. Chem. Theory Comput.*, 2012, **8**, 2824–2834.
- [40] L. Goerigk and S. Grimme, *J. Chem. Theory Comput.*, 2009, **6**, 107–126.
- [41] L. Goerigk and S. Grimme, *J. Chem. Theory Comput.*, 2010, **7**, 291–309.
- [42] L. Goerigk and S. Grimme, *Phys. Chem. Chem. Phys.*, 2011, **13**, 6670–6688.
- [43] P. Jurečka, J. Sponer, J. Cerny and P. Hobza, *Phys. Chem. Chem. Phys.*, 2006, **8**, 1985–1993.
- [44] L. Goerigk and S. Grimme, *J. Chem. Theory Comput.*, 2011, **7**, 291–309.
- [45] S. Grimme, H. Kruse, L. Goerigk and G. Erker, *Angew. Chem. Int. Ed.*, 2010, **49**, 1402–1405.
- [46] R. Sure, A. Hansen, P. Schwerdtfeger and S. Grimme, *Phys. Chem. Chem. Phys.*, 2017, **19**, 14296–14305.
- [47] A. Karton, R. J. O'Reilly, B. Chan and L. Radom, *J. Chem. Theory Comput.*, 2012, **8**, 3128–3136.
- [48] A. Karton, R. J. O'Reilly and L. Radom, *J. Phys. Chem. A*, 2012, **116**, 4211–4221.
- [49] T. Schwabe and S. Grimme, *Phys. Chem. Chem. Phys.*, 2007, **9**, 3397–3406.
- [50] S. Grimme, *Angew. Chem. Int. Ed.*, 2006, **45**, 4460–4464.
- [51] T. Takatani, E. G. Hohenstein, M. Malagoli, M. S. Marshall and C. D. Sherrill, *J. Chem. Phys.*, 2010, **132**, 144104.
- [52] M. G. Kendall, *Biometrika*, 1938, **30**, 81–93.