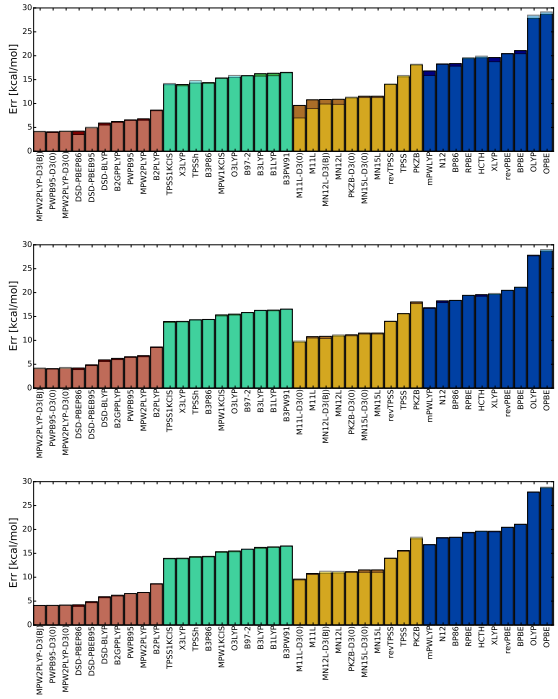
Electronic Supplementary Material (ESI) for Physical Chemistry Chemical Physics.

Electronic supplementary information for "'Diet GMKTN55' offers accelerated benchmarking through a representative subset approach"

This electronic supplementary information contains:

1. Supplementary Figure 1 showing the worst cases using the $N_s = 30$, 100 and 150 element subsets.
2. Supplementary Tables 1,2 listing systems included in the $N_s = 30$ and 150 subsets, to supplement results for $N_s = 100$ in the main text.
3. Supplementary Figure 2 showing the best cases for MGCDB84 using $N_s = 30$, 100 and 150 element subsets tailored to the alternative database.
4. Description of results and files provided on Github.



Supplementary Figure 1 Comparison of errors found using $N_s = 30, 100, 150$ (top to bottom) subsets compared with the full database. Here, results are divivided into four categories: double-hybrid DFAs (reds), hybrid DFAs (greens), meta-GGAs (earth tones) and GGAs (blues). 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 worst methods are shown in each category.

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.

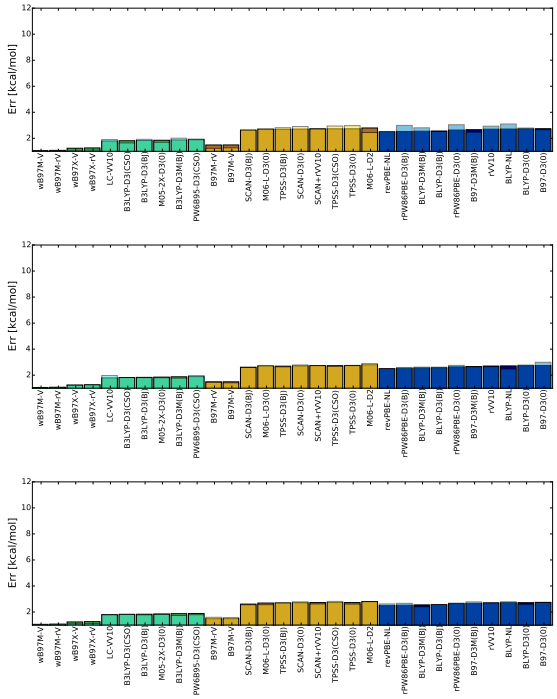
Set	W	IDs	Set	W	IDs	Set	W	IDs	Set	W	IDs
ACONF[†]	30.99	10	Amino20x4[†]	23.31	28, 54	BH76[14,15]	3.05	5	BHPERI[16]	2.72	11
BHROT27[†]	9.06	16, 26	BSR36[17]	3.51	31	BUT14DIOL[18]	20.30	13	CDIE20[19]	14.02	9
DC13[†]	1.03	1	DIPCS10[†]	0.09	7	FH51[20]	1.83	24, 30	G21EA[21]	1.69	14, 25
HAL59[24,25†]	12.38	40, 57	HEAVY28[26]	45.79	16	MB16-43[31]	0.12	10	MCONF[32]	11.43	1
PNICO23[34]	13.30	16	PX13[50]	1.70	9	S66[37]	10.40	6, 50	SIE4x4[†]	1.69	15
W4-11[40]	0.19	30, 57, 132	WCPT18[51]	1.62	15						

[†] Contains systems from one of the GMTKN databases[5,43–45]

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[12]	2.53	6, 10, 15
AL2X6[†]	1.58	5	Amino20x4[†]	23.31	4, 6, 15, 25, 39, 41, 46, 68
BH76[14,15]	3.05	8, 36, 37, 47, 50, 53, 66	BH76RC[†]	2.66	5
BHPERI[16]	2.72	2, 8, 25	BHROT27[†]	9.06	3, 11, 14, 21
BSR36[17]	3.51	17, 22	BUT14DIOL[18]	20.30	2, 29, 32, 38, 39, 42, 43, 53
C60ISO[49]	0.58	7, 9	CDIE20[19]	14.02	1, 18, 20
CHB6[12]	2.12	2	DARC[†]	1.75	11
DC13[†]	1.03	10, 13	DIPCS10[†]	0.09	2
FH51[20]	1.83	26, 27, 40, 48	G21EA[21]	1.69	17, 20, 23
G21IP[21,22]	0.22	6, 15, 31	G2RC[23]	1.11	16, 24, 25
HAL59[24,25†]	12.38	7, 11, 16, 28, 32, 56	HEAVY28[26]	45.79	3, 16, 24
ICONF[†]	17.40	16	IDISP[22,29,47,52,53]	4.00	3
IL16[12,27]	0.52	11, 14	INV24[28]	1.78	1, 14, 15
ISO34[29]	3.90	21	ISOL24[30]	2.59	2, 4, 20
MB16-43[31]	0.12	14	MCONF[32]	11.43	11, 20, 30
NBPRC[22,48]	2.30	1	PA26[†]	0.30	20, 25
PArel[†]	12.28	13	PCONF21[33]	35.05	3
PNICO23[34]	13.30	7, 11, 13, 21	PX13[50]	1.70	8
RC21[35]	1.59	5, 6, 8, 16, 20	RG18[†]	98.00	11, 13, 15
RSE43[36]	7.48	14, 40	S22[46,54]	7.78	3, 8, 11
S66[37]	10.40	8, 17, 25, 28, 42, 47, 61	SCONF[38]	12.36	16
SIE4x4[†]	1.69	15	TAUT15[†]	18.66	5, 13
UPU23[39]	9.93	2, 4, 9, 15	W4-11[40]	0.19	2, 7, 30, 31, 38, 58, 63, 73, 77, 85, 86, 95, 102, 106, 129
WATER27[41]	0.70	2, 17	WCPT18[51]	1.62	6, 8, 9, 18
YBDE18[42]	1.15	6, 8, 9, 13, 16			

[†] Contains systems from one of the GMTKN databases[5,43–45]



Supplementary Figure 2 Comparison of errors found using $N_s = 30, 100, 150$ (top to bottom) subsets compared with the full database for MGCDB84. Here, results are divivided into three categories: hybrid DFAs (greens), meta-GGAs (earth tones) and GGAs (blues). 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. Scale is kept consistent with plots in the main text.

Description of results and files provided on Github

In addition to the tabulated data presented here, key results were collated into "yaml" files (`http://yaml.org/`) containing the structural data required to use the diet databases reported in the text. These files, are available at `https://github.com/gambort/DietGMTKN55`.

As an example of how to use these files, a python script is also provided that generates the structural part of Gaussian Input Files. This file is primarily designed to serve as a software-ambivalent template for generating structures for any quantum chemistry code, and for processing the outputs therefrom.

Additional files are provided for the diet versions of MGCDB84[9], which were used to generate Supplementary Figure 2.

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] B. Chan, *J. Chem. Theory Comput.*, 2018, **14**, 4254–4262.

[9] N. Mardirossian and M. Head-Gordon, *Mol. Phys.*, 2017, **115**, 2315–2372.

[10] B. J. Lynch and D. G. Truhlar, *J. Phys. Chem. A*, 2003, **107**, 8996–8999.

[11] *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`.

[12] K. U. Lao, R. Schäffer, G. Jansen and J. M. Herbert, *J. Chem. Theory Comput.*, 2015, **11**, 2473–2486.

[13] H. Yu and D. G. Truhlar, *J. Chem. Theory Comput.*, 2015, **11**, 2968–2983.

[14] Y. Zhao, B. J. Lynch and D. G. Truhlar, *Phys. Chem. Chem. Phys.*, 2005, **17**, 43–52.

[15] Y. Zhao, N. González-García and D. G. Truhlar, *J. Phys. Chem. A*, 2005, **109**, 2012–2018.

[16] A. Karton and L. Goerigk, *J. Comput. Chem.*, 2015, **36**, 622–632.

[17] S. N. Steinmann, G. Csonka and C. Corminboeuf, *J. Chem. Theory Comput.*, 2009, **5**, 2950–2958.

[18] S. Kozuch, S. M. Bachrach and J. M. Martin, *J. Phys. Chem. A*, 2014, **118**, 293–303.

[19] L.-J. Yu and A. Karton, *Chem. Phys.*, 2014, **441**, 166–177.

[20] J. Friedrich and J. Hänchen, *J. Chem. Theory Comput.*, 2013, **9**, 5381–5394.

[21] L. A. Curtiss, K. Raghavachari, G. W. Trucks and J. A. Pople, *J. Chem. Phys.*, 1991, **94**, 7221–7230.

[22] L. Goerigk and S. Grimme, *J. Chem. Theory Comput.*, 2010, **6**, 107–126.

[23] L. A. Curtiss, K. Raghavachari, P. C. Redfern and J. A. Pople, *J. Chem. Phys.*, 1997, **106**, 1063–1079.

[24] S. Kozuch and J. M. L. Martin, *J. Chem. Theory Comput.*, 2013, **9**, 1918–1931.

[25] J. Rezac, K. E. Riley and P. Hobza, *J. Chem. Theory Comput.*, 2012, **8**, 4285–4292.

[26] S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *J. Chem. Phys.*, 2010, **132**, 154104.

[27] S. Zahn, D. R. MacFarlane and E. I. Izgorodina, *Phys. Chem. Chem. Phys.*, 2013, **15**, 13664–13675.

[28] L. Goerigk and R. Sharma, *Can. J. Chem.*, 2016, **94**, 1133–1143.

[29] S. Grimme, M. Steinmetz and M. Korth, *J. Org. Chem.*, 2007, **72**, 2118–2126.

[30] R. Huenerbein, B. Schirmer, J. Moellmann and S. Grimme, *Phys. Chem. Chem. Phys.*, 2010, **12**, 6940–6948.

[31] M. Korth and S. Grimme, *J. Chem. Theory Comput.*, 2009, **5**, 993–1003.

[32] U. R. Fogueri, S. Kozuch, A. Karton and J. M. L. Martin, *J. Phys. Chem. A*, 2013, **117**, 2269–2277.

[33] D. Rěha, H. Valdes, J. Vondrasek, P. Hobza, A. Abu-Riziq, B. Crews and M. S. de Vries, *Chem. - Eur. J.*, 2005, **11**, 6803–6817.

[34] D. Setiawan, E. Kraka and D. Cremer, *J. Phys. Chem. A*, 2015, **119**, 1642–1656.

[35] S. Grimme, *Angew. Chem. Int. Ed.*, 2013, **52**, 6306–6312.

[36] F. Neese, T. Schwabe, S. Kossmann, B. Schirmer and S. Grimme, *J. Chem. Theory Comput.*, 2009, **5**, 3060–3073.

[37] J.Řezáč, K. E. Riley and P. Hobza, *J. Chem. Theory Comput.*, 2011, **7**, 2427–2438.

[38] G. I. Csonka, A. D. French, G. P. Johnson and C. A. Stortz, *J. Chem. Theory Comput.*, 2009, **5**, 679–692.

[39] H. Kruse, A. Mladek, K. Gkionis, A. Hansen, S. Grimme and J. Sponer, *J. Chem. Theory Comput.*, 2015, **11**, 4972–4991.

[40] A. Karton, S. Daon, J. M. L. Martin and B. Ruscic, *Chem. Phys. Lett.*, 2011, **510**, 165–178.

[41] V. S. Bryantsev, M. S. Diallo, A. C. T. van Duin and W. A. Goddard III, *J. Chem. Theory Comput.*, 2009, **5**, 1016–1026.

[42] Y. Zhao, H. T. Ng, R. Peverati and D. G. Truhlar, *J. Chem. Theory Comput.*, 2012, **8**, 2824–2834.

[43] L. Goerigk and S. Grimme, *J. Chem. Theory Comput.*, 2009, **6**, 107–126.

[44] L. Goerigk and S. Grimme, *J. Chem. Theory Comput.*, 2010, **7**, 291–309.

[45] L. Goerigk and S. Grimme, *Phys. Chem. Chem. Phys.*, 2011, **13**, 6670–6688.

[46] P. Jurečka, J. Sponer, J. Cerny and P. Hobza, *Phys. Chem. Chem. Phys.*, 2006, **8**, 1985–1993.

[47] L. Goerigk and S. Grimme, *J. Chem. Theory Comput.*, 2011, **7**, 291–309.

[48] S. Grimme, H. Kruse, L. Goerigk and G. Erker, *Angew. Chem. Int. Ed.*, 2010, **49**, 1402–1405.

[49] R. Sure, A. Hansen, P. Schwerdtfeger and S. Grimme, *Phys. Chem. Chem. Phys.*, 2017, **19**, 14296–14305.

[50] A. Karton, R. J. O'Reilly, B. Chan and L. Radom, *J. Chem. Theory Comput.*, 2012, **8**, 3128–3136.

[51] A. Karton, R. J. O'Reilly and L. Radom, *J. Phys. Chem. A*, 2012, **116**, 4211–4221.

[52] T. Schwabe and S. Grimme, *Phys. Chem. Chem. Phys.*, 2007, **9**, 3397–3406.

[53] S. Grimme, *Angew. Chem. Int. Ed.*, 2006, **45**, 4460–4464.

[54] T. Takatani, E. G. Hohenstein, M. Malagoli, M. S. Marshall and C. D. Sherrill, *J. Chem. Phys.*, 2010, **132**, 144104.

[55] J. P. Perdew and K. Schmidt, AIP Conference Proceedings, 2001, pp. 1–20.

[56] M. G. Kendall, *Biometrika*, 1938, **30**, 81–93.