Supporting Information for

Formulation of Small Test Sets Using Large Test Sets for Efficient Assessment of Quantum Chemistry Methods

Bun Chan

Graduate School of Engineering, Nagasaki University, Bunkyo 1-14, Nagasaki 852-8521, Japan

Additional Discussions and Information

Comparison Between MAD and MAFD for Method Assessment	S2-S5
Comparison Between MGCDB82 and GMTKN55	S6-S7
The Use of MAFD versus WTMAD2	S8-S9
An Alternate Set of Scale Coefficients for the MG8 Model	S10-S10
Description of the DFT Methods Used for Cross-Validating MG8	S11-S12
Assessment of DFT Methods with the SGB13 Model	S13-S16
The XE6 Model for Assessing Computed Excitation Energies	S17-S18
References	S19-S22

Section S1. Comparison Between MAD and MAFD for Method Assessment

An import question for assessing the accuracy of a method is: on what basis do we judge the its accuracy? In other words, what indicator(s) do we use? For instance, mean absolution deviation (MAD) and root-mean-square deviation (RMSD) are widely used. However, we do not consider a meaningful distinction exists between MAD and RMSD because, at the limit of normally distributed deviations, the two are related by a constant factor.¹ In this regard, we prefer the more straightforward interpretation of MAD.

A question with less trivial answer is whether to consider the magnitudes of all data points having equal importance. For instance, for the assessment with the GMTKN55 set,² weighted total MADs (WTMADs) were used instead. Two WTMADs, namely WTMAD1 and WTMAD2, have been defined, with WTMAD2 representing a more objective gauge. Regardless of the choice of WTMAD, both take into account that a particularly sized deviation would represent different fractional deviations for quantities of dissimilar magnitudes, which may lead to a disparity in practical implications. We will consider this aspect in some detail using the raw data provided for the MGCDB82 set.³ Specifically, in addition to obtaining unweighted MADs, we will also use a simple measure of dividing the magnitude of each deviation by its reference value. We will refer to the sum of these fractional deviations as the mean absolute fractional deviation (MAFD).

Let us first consider these two measures, namely MAD and MAFD, for gauging the overall accuracy of a computational chemistry method. Before we argue for or against one or the other, shall we first ask the question: does it matter? In other words, do the two metrics reveal the same information of interest, which in this case is the (relative) accuracy of a method for calculating thermochemical quantities. To attempt answering this question, we plot the MADs for all the (~200) methods assessed previously against the corresponding MAFDs (Figure S1). While the R^2 value for a linear regression is just 0.324, visualizing the plot shows a somewhat less discouraging picture. We see several key outliers in the plot above a trend that is

qualitatively reasonably clear, though the quantitative correlation seems to be quite poor. Among all the methods assessed, HF is one of the least accurate both in terms of MAD and MAFD. However, its MAD is disproportionally large when considering the general trend of most other methods.



Figure S1. Mean absolute deviations (MADs) for MGCDB82 versus the corresponding mean absolute fractional deviations (MAFDs) for \sim 200 methods assessed previously.

Several other outliers are related to HF. The dispersion-corrected⁴ variants [HF-D3(0), HF-D3(BJ) and HF-NL] show vastly improved MAFDs. This is not surprising because these corrections are designed to improve the description of non-covalent interactions. These quantities typically have smaller relative energies when compared with those for other properties. We can therefore expect them to contribute more significantly to the MAFDs. However, the improvement in MADs achieved by using dispersion corrections are not large. Overall, the inclusion of dispersion corrections brings the points further away from the general trend of the plot. In contrast, when one supplement HF with a correlation functional (PW92⁵ or LYP⁶), one sees improvements in MAD and MAFD to more balanced extents. These two methods are less glaring outliers in the plot.

Apart from HF and related methods, we also find the two local-density approximation methods (LDAs, specifically S-VWN5^{7,8} and S-PW92^{5,7}) to be notable outliers in the figure. An expanded version of Figure S1 is also given (Figure S2), with a focus on the region of small MADs and MAFDs. It shows that, while dispersion-corrected functionals are generally more accurate than non-corrected ones, some non-corrected ones are competitive with some of the most accurate DFTs being assessed, including the dispersion-corrected ones. We note that the use of dispersion corrections is becoming almost ubiquitous in recent years. While we appreciate their use for current generation of approximate functionals, the results in Figure S1 remind us of the importance of seeking a balanced improvement across the board, rather than relying too heavily on features designed with a narrow focus.



Figure S2. Mean absolute deviations (MADs, kJ mol⁻¹) for MGCDB82 versus the corresponding mean absolute fractional deviations (MAFDs).

If we remove the outliers [HF and related methods and the two LDAs] from the linear regression, the R^2 become 0.567, which is by no means satisfactory but does correspond to a semi-qualitatively usable trend (with a high level of uncertainty). When we further put the focus on just the non-dispersion-corrected methods, the R^2 become 0.674. Thus, when the aim is to compare different methods for computing a

range of main-group chemical quantities that are diverse and representative, the two measures seems to have some correspondence. The dilemma of using one or the other may therefore not be of critical importance. The two quantities MAD and MAFD provide us with different information regarding a particular method, and (in our opinion) it would not be wise to argue for or against one or the other with absolute conviction. In the present study, we use MAD as we prefer direct inspection of the actual magnitudes of the deviations.

Section S2. Comparison Between MGCDB82 and GMTKN55.

Another aspect that we will briefly discuss is whether the results of MGCDB82 and GMTKN55 sets resemble those of each other. After all, if the observations for a large and general compilation are not transferable to another extensive set assembled with very similar intended purposes, it may not be realistic to expect a substantially smaller set to correspond well with any of these sets in a robust manner. More importantly, if the conclusions drawn from one large set are very different to those from another, it would be a rather dubious decision to use one set as a target over another. Both MGCDB82 and GMTKN55 have been used to benchmark large collections of DFT methods.^{2,3} In each case, approximately 200 functionals have been tested. Among these, 75 are common to both studies. They are (in alphabetical order):

B3LYP, B3LYP-D3BJ, B3LYP-NL, B3P86, B3PW91, B3PW91-D3BJ, B97-1, B97-2, B97-D3BJ, BLYP, BLYP-D3BJ, BMK, BMK-D3BJ, BP86, BP86-D3BJ, BPBE, BPBE-D3BJ, HCTH407, M05, M05-D3, M052X, M052X-D3, M06, M06-D3, M062X, M062X-D3, M06L, M06L-D3, M08HX, M11, M11-D3BJ, M11L, M11L-D3, MN12L, MN12L-D3BJ, MN12SX, MN12SX-D3BJ, MN15, MN15L, MPWPW91, N12, N12-D3, N12SX, N12SX-D3BJ, OLYP, OLYP-D3BJ, PBE, PBE-D3BJ, PBE0, PBE0-D3BJ, PKZB, PW6B95, PW6B95-D3BJ, PW91, revPBE, revPBE-D3BJ, revTPSS, revTPSSh, RPBE, RPBE-D3BJ, rPW86PBE, rPW86PBE-D3BJ, SCAN, SCAN-D3BJ, SOGGA11X, SOGGA11X-D3BJ, tHCTH, tHCTHhyb, TPSS, TPSS-D3BJ, TPSSh, TPSSh-D3BJ, VV10, ωB97X-D3, ωB97X-V.

We will thus compare the statistical measures of these methods for the two sets. In the assessment with GMTKN55, weighted total MADs (termed WTMAD2) were used instead of MAD. To facilitate a straightforward comparison, for MGCDB82, in addition to MADs, we have also obtained WTMAD2 values using the same protocol for calculating those for the GMTKN55 set.

Figure S3 shows the correlation between the MADs and WTMAD2s for MGCDB82 versus the WTMAD2s for the GMTKN55 set. We can see that the trends for the two

S6

sets correspond reasonably well, with R^2 values of 0.901 and 0.935, respectively, for the MADs and WTMAD2s for MGCDB82. While the good R^2 values between these sets would suggest the observations on MGCDB82 are quite likely to be applicable also to GMTKN55, visual inspection of the figure indicates that poor correlations may occur. This can be illustrated by WTMAD2 values of ~ 50 kJ mol⁻¹ for GMTKN55 corresponding to a series of WTMAD2s for MGCDB82 spanning a range from ~20 to ~70 kJ mol⁻¹. We note that this is despite the R^2 value for WTMAD2 being larger than that for MAD. With that being said, all measures agree at the far ends of the spectrum, with O-LYP⁹ being inferred as the least accurate and ω B97X-V¹⁰ being indicated as the most accurate, though some other methods also come close to matching its rather remarkable accuracy.



Figure S3. MADs and WTMAD2s for MGCDB82 versus normalized WTMAD2 values for GMTKN55 for a set of 75 DFT methods that have been assessed with both data sets. All values are given in the unit of kJ mol⁻¹.

Section S3. The Use of MAFD versus WTMAD2

So far, our discussion has been based mainly on three statistical indicators for assessing the accuracy of a method (MAD, MAFD and WTMAD2). Among them, both MAFD and WTMAD2 are metrics intended to provide information about the accuracy of an assessed method in terms of deviations relative to the magnitudes of the quantities being determined. MAFD is defined as:

$$MAFD = \sum_{i=1}^{n} \left| \frac{E_{i,calc} - E_{i,ref}}{E_{i,ref}} \right|$$

in which *i* refers to individual data points. A more elaborate scheme² is used to obtain the WTMAD2 for an assessed method:

$$MAD = \sum_{i=1}^{m} \left| E_{i,calc} - E_{i,ref} \right|$$

WTMAD2 =
$$\frac{\overline{|E_{ref}|}_{overall}}{\sum_{j=1}^{n} N_j} \sum_{j=1}^{n} \frac{N_j}{\overline{|E_{ref}|}_j} MAD_j$$

in which *i* refers to data points within a subset and *j* refers to subsets within the overall collection of data sets; *N* is the number of data points. MAFD measures the relative deviation for each data point directly, whereas WTMAD2 takes a collection of data points in a subset as a single unit, and determines the relative deviations based on the number of contributing data points to the overall compendium and the average energetic magnitudes for these units of subsets. While the two metrics appear to be quite different in their forms, for the data obtained previously for MGCDB82, they correlate well for all the assessed methods (Figure S3). For gaining insights about deviations relative to the magnitudes of the quantities, we prefer MAFD for its simplicity and more straightforward interpretation.



Figure S4. Mean absolute fractional deviations (MAFDs) for MGCDB82 versus the corresponding weighted total mean absolute deviation 2 (WTMAD2, kJ mol⁻¹) values.

Section S4. An Alternate Set of Scale Coefficients for the MG8 Model

The eight data sets in the MG8 model provide, as a single set, the best statistical representation of their categories in MGCDB82. To arrive at a single measure, we can sum the weighted MADs of the eight subsets with the weighting coefficients determined by the number of data points in the groups that they contribute to the overall compendium. In this way and in conjunction with the regression formula for each set (see the MG8 worksheet in the Excel workbook of Supporting Information), one can arrive at a definition for EMAD_{MG8} alternative to the one used in the main text. The coefficients obtained by the two approaches are shown in Table S1. While the parameters differ quantitatively, they both yield good correlations between the obtained EMADs and the actual MADs for MGCDB82 (R^2 of 0.988 and 0.984 with the regression and derived coefficients, respectively).

Table S1. Data Sets in the MG8 Model, Their Scale Coefficients and Interceptsfor Obtaining EMAD Values, as Determined by Linear Regression andDerivation Based on Number of Contributing Data Points of the Categories

Group	Data set	regression	derived
NCEC	S66x8	0.4533	0.3261
NCED	H2O20Bind10	0.0078	0.0102
NCD	Bauza30	0.0580	0.0119
IE	YMPJ519	0.2084	0.1827
ID	C20C24	0.0116	0.0034
TCE	TAE140nonMR	0.0720	0.0762
TCD	TAE140MR	0.0232	0.0173
BH	DBH24	0.0278	0.0455
Intercept (kJ mol ⁻¹)		2.5258	3.6461

Section S5. Description of the DFT Methods Used for Cross-Validating MG8

We have assessed the accuracy of six DFT methods for the purpose of crossvalidating the MG8 model with the independent VM6 and VM9 models. These methods are HSEB-D, EDF2-D, W404-D, APF-D, MBh-D and PGTh. Here we provide a brief overview of the DFT methods and the rationale for their inclusion in our assessment.

The HSEB-D method¹¹ contains a screened-exchange functional based on the HSE^{12,13} method but with the PBEc correlation component¹⁴ in HSE replaced by a modified B97c functional,¹⁵ and it also contains a D2¹⁶ dispersion term. This method has been shown to provide a better accuracy than that by HSE for a range of test sets with a total of more than 800 data points. In the present study, we use an alternative s_6 value of 0.8 for the D2 component to yield the smallest EMAD_{MG8}. The EDF2 method¹⁷ is devised for obtaining accurate vibrational frequencies, while WP04¹⁸ is designed for the accurate computation of proton chemical shifts. We supplement these methods with D2 dispersion corrections optimized for EMAD_{MG8} with s_6 values of 0.6 and 1.2, respectively. The APF-D method¹⁹ comprises the APF exchange–correlation component that is formulated to be "dispersion-free", and a complementary dispersion term developed with a spherical atom model. We retain this original dispersion component in the present study.

We formulate the MBh-D method by examining the combination of several exchange functionals (B88,²⁰ MPW²¹ and PBE¹⁴) with the revised B97c correlation component used in HSEB-D, within the context of a global-hybrid DFT with a D2 dispersion term. The method that shows the best accuracy, as judged with EMAD_{MG8}, employs the MPW exchange with 25% Hartree–Fock exchange, and an s_6 value of 1.1. We devised the PGTh-D method in a similar manner. In this case, we employ the PBE-GX exchange,²² which is based on a finite uniform electron gas and is thus distinct from all other commonly used exchange functionals. We examine its combination with two correlation functionals, namely SCANc²³ and regTPSSc²⁴, as they have been found to yield good accuracy when used in conjunction with PBE-GX

for a small set of systems.²² Based on the results of our assessment, we define the PGTh method as a global-hybrid DFT with 30% Hartree–Fock exchange using the regTPSSc correlation functional. We have explored the addition of a D2 term to PGTh but find that it does not lead to a smaller EMAD_{MG8}.

Section S6. Assessment of DFT Methods with the SGB13 Model

The SBG31 set²⁵ has been used regularly for assessing computational chemistry methods for the accuracies in calculating semiconductor band gaps. To down-size SBG31 and devise our SBG13 model, we have benchmarked 26 pure and screened-exchange DFT methods. The results are shown in Table S2. With the exception of HISS, all methods considered here underestimate these band gaps. In general, we find that the screened-exchange methods provide the best quantitative agreements, with typical MADs of ~ 0.3 eV. In comparison, the MADs for other methods are generally in the range of ~ 0.7–1.2 eV.

	AlAs	AlP	AlSb	B-As	BP	BaS	BaSe
Ref	2.23	2.51	1.68	1.46	2.40	3.88	3.58
SVWN5	1.35	1.56	1.21	1.13	1.31	2.08	1.77
BLYP	1.71	1.96	1.42	1.42	1.63	2.42	2.06
BP86	1.53	1.76	1.34	1.24	1.42	2.33	1.98
BPW91	1.49	1.73	1.32	1.21	1.39	2.38	2.00
HCTH147	1.77	2.06	1.52	1.40	1.59	2.61	2.21
HCTH407	1.88	2.21	1.59	1.46	1.59	2.82	2.36
НСТН93	1.80	2.09	1.55	1.40	1.55	2.72	2.30
M06L	1.83	2.10	1.43	1.48	1.80	2.61	2.18
M11L	2.22	2.41	1.87	1.79	1.68	2.33	2.35
MN12L	1.72	1.77	1.38	1.41	1.38	2.24	2.01
MN15L	1.72	1.89	1.32	1.37	1.47	2.38	2.09
mPW91	1.49	1.73	1.32	1.22	1.39	2.36	1.99
N12	1.52	1.88	1.28	1.25	1.46	2.42	1.98
OLYP	1.76	2.02	1.57	1.32	1.44	2.83	2.38
PBE	1.47	1.70	1.29	1.19	1.36	2.35	1.98
PKZB	1.50	1.69	1.34	1.12	1.24	2.52	2.21
PW91	1.48	1.71	1.31	1.21	1.39	2.33	1.97
revTPSS	1.53	1.69	1.37	1.11	1.28	2.44	2.10
SOGGA	0.96	1.14	0.95	0.73	1.04	3.02	2.33
tHCTH	1.81	2.10	1.59	1.41	1.61	2.66	2.25
TPSS	1.58	1.78	1.43	1.21	1.40	2.50	2.11
VSXC	1.67	1.75	1.21	1.44	1.63	1.69	1.69
N12SX	2.25	2.54	1.96	1.93	2.10	3.41	2.95
MN12SX	2.20	2.36	1.89	1.86	1.95	2.90	2.68

Table S2. Reference and Calculated Band Gaps (eV) for the SBG31 Set

HSE	2.13	2.42	1.82	1.88	2.13	3.19	2.74
HISS	2.40	2.72	2.05	2.15	2.45	3.60	3.09
	ВаТе	С	CdS	CdSe	CdTe	GaAs	GaN
Ref	3.08	5.48	2.55	1.90	1.92	1.52	3.50
SVWN5	1.40	4.17	0.97	0.31	0.54	0.04	1.83
BLYP	1.69	4.50	1.21	0.55	0.70	0.15	1.96
BP86	1.60	4.24	1.19	0.56	0.73	0.26	1.85
BPW91	1.62	4.23	1.26	0.62	0.79	0.32	1.86
HCTH147	1.80	4.33	1.48	0.78	0.86	0.41	2.06
HCTH407	1.97	4.34	1.71	0.96	1.00	0.52	2.16
НСТН93	1.90	4.29	1.62	0.93	0.98	0.53	2.12
M06L	1.51	4.75	1.82	1.22	1.28	1.01	2.09
M11L	1.98	4.02	1.83	1.91	2.61	1.08	2.23
MN12L	1.45	4.24	1.46	0.76	1.02	0.78	1.79
MN15L	1.44	4.14	1.79	1.08	1.12	1.33	1.97
mPW91	1.60	4.24	1.24	0.60	0.78	0.30	1.86
N12	1.56	4.33	1.42	0.55	0.68	0.13	1.96
OLYP	2.01	4.18	1.73	1.06	1.10	0.63	2.06
PBE	1.59	4.20	1.26	0.63	0.81	0.36	1.86
PKZB	1.79	3.99	1.53	0.96	1.13	0.76	1.87
PW91	1.58	4.24	1.22	0.58	0.77	0.29	1.86
revTPSS	1.69	4.09	1.41	0.86	1.08	0.64	1.70
SOGGA	1.87	3.95	2.40	1.68	1.73	0.72	1.99
tHCTH	1.87	4.33	1.50	0.76	0.84	0.33	2.08
TPSS	1.70	4.24	1.47	0.85	1.05	0.60	1.78
VSXC	1.34	4.40	1.78	1.29	1.44	0.88	2.13
N12SX	2.44	5.50	2.50	1.63	1.75	1.06	3.36
MN12SX	2.17	5.05	2.20	1.62	1.86	1.03	2.95
HSE	2.21	5.43	2.21	1.48	1.64	1.11	3.08
HISS	2.49	6.08	2.79	1.98	2.08	1.58	3.97
	GaNwu	GaP	GaSb	Ge	InAs	InN	InP
Ref	3.30	2.35	0.73	0.74	0.41	0.69	1.42
SVWN5	2.18	1.56	0.00	0.00	0.00	0.00	0.74
BLYP	2.31	1.85	0.00	0.00	0.00	0.00	0.83
BP86	2.20	1.74	0.08	0.01	0.00	0.00	0.89
BPW91	2.22	1.75	0.14	0.07	0.00	0.00	0.96
HCTH147	2.42	1.98	0.16	0.13	0.00	0.00	1.04
HCTH407	2.53	2.15	0.26	0.26	0.00	0.10	1.22
НСТН93	2.48	2.05	0.27	0.25	0.00	0.06	1.16
M06L	2.47	2.11	0.75	0.60	0.59	0.00	1.72
M11L	2.57	2.05	1.15	0.67	1.07	0.37	1.34
MN12L	2.14	1.78	0.58	0.56	0.20	0.00	1.60
MN15L	2.33	1.82	0.79	0.73	0.52	0.00	1.87

mPW91	2.22	1.75	0.13	0.05	0.00	0.00	0.95
N12	2.32	1.93	0.00	0.00	0.00	0.00	1.14
OLYP	2.43	2.07	0.37	0.34	0.12	0.07	1.24
PBE	2.22	1.74	0.19	0.13	0.00	0.00	0.99
PKZB	2.22	1.76	0.54	0.42	0.22	0.00	1.31
PW91	2.21	1.73	0.12	0.04	0.00	0.00	0.94
revTPSS	2.05	1.73	0.44	0.33	0.13	0.00	1.18
SOGGA	2.34	1.10	0.76	0.16	0.44	0.15	1.19
tHCTH	2.44	1.99	0.08	0.02	0.00	0.01	0.98
TPSS	2.15	1.83	0.39	0.32	0.08	0.00	1.19
VSXC	2.45	1.58	0.59	0.53	0.33	0.07	1.59
N12SX	3.76	2.59	0.87	0.74	0.62	1.06	2.07
MN12SX	3.33	2.29	0.80	0.64	0.57	0.63	1.69
HSE	3.48	2.39	0.90	0.80	0.57	0.72	1.77
HISS	4.37	2.67	1.32	1.10	0.98	1.43	2.24
	InSb	MgS	MgSe	MgTe	SiC	Si	ZnO
Ref	0.23	5.40	2.47	3.60	2.42	1.17	3.40
SVWN5	0.00	3.37	1.64	2.41	1.34	0.53	0.75
BLYP	0.00	3.75	1.86	2.67	1.72	0.85	1.00
BP86	0.00	3.65	1.86	2.66	1.47	0.68	0.87
BPW91	0.00	3.68	1.88	2.69	1.42	0.65	0.90
HCTH147	0.00	4.01	2.16	2.87	1.71	0.89	1.29
HCTH407	0.00	4.29	2.19	3.01	1.70	0.95	1.46
НСТН93	0.00	4.15	2.18	2.96	1.63	0.89	1.38
M06L	0.36	4.27	2.46	3.18	1.62	1.04	0.93
M11L	1.31	3.92	3.33	3.32	2.30	1.21	1.45
MN12L	0.11	3.86	3.28	2.73	1.78	0.74	0.56
MN15L	0.14	4.03	2.96	2.73	1.58	0.85	0.87
mPW91	0.00	3.65	1.86	2.67	1.43	0.65	0.89
N12	0.00	3.86	1.89	2.56	1.39	0.76	1.19
OLYP	0.00	4.22	2.08	3.03	1.43	0.82	1.33
PBE	0.00	3.64	1.81	2.65	1.38	0.62	0.87
PKZB	0.08	3.97	2.35	3.06	1.25	0.61	0.96
PW91	0.00	3.63	1.83	2.64	1.42	0.64	0.88
revTPSS	0.04	3.88	2.12	3.09	1.22	0.61	0.77
SOGGA	0.66	3.77	2.09	3.43	1.22	0.26	0.93
tHCTH	0.00	4.10	2.39	2.99	1.75	0.94	1.44
TPSS	0.00	3.90	2.12	3.07	1.34	0.71	0.90
VSXC	0.09	3.33	3.32	2.35	2.01	0.87	1.20
N12SX	0.52	5.11	2.84	3.73	2.12	1.30	2.43
MN12SX	0.47	4.61	3.36	3.55	2.37	1.13	2.43
HSE	0.47	4.67	2.69	3.54	2.32	1.21	2.44
HISS	0.83	5.31	3.14	3.91	2.75	1.45	3.48

	ZnS	ZnSe	ZnTe
Ref	3.66	2.70	2.38
SVWN5	2.02	1.05	1.11
BLYP	2.26	1.28	1.25
BP86	2.24	1.30	1.31
BPW91	2.29	1.35	1.36
HCTH147	2.59	1.58	1.49
HCTH407	2.81	1.73	1.60
НСТН93	2.72	1.71	1.61
M06L	2.97	2.05	1.89
M11L	2.77	2.67	2.94
MN12L	2.73	1.68	1.75
MN15L	3.11	2.09	1.93
mPW91	2.28	1.33	1.35
N12	2.50	1.27	1.22
OLYP	2.77	1.79	1.68
PBE	2.30	1.37	1.39
PKZB	2.65	1.78	1.79
PW91	2.26	1.32	1.34
revTPSS	2.48	1.64	1.69
SOGGA	2.23	1.53	1.87
tHCTH	2.63	1.58	1.49
TPSS	2.53	1.62	1.65
VSXC	2.86	2.13	2.02
N12SX	3.68	2.41	2.34
MN12SX	3.50	2.57	2.59
HSE	3.44	2.38	2.34
HISS	4.13	2.97	2.88

Section S6. The XE6 Model for Assessing Computed Excitation Energies

In addition to the MOR13, SBG5 and MB13 models discussed in the main text, we have further formulated the XE6 model for evaluating the accuracy of theoretical methods for the calculation of excitation energies. The XE6 model is based on the set of excitation energies of Schreiber et al.,²⁶ which contains 28 molecules and a variety of singlet and triplet excitation energies. Among these, 103 singlet excitation energies have been used subsequently for the benchmark of 30 quantum chemistry methods.²⁷ We use standard lasso regularization regression to formulate a linear combination model of a minimal subset of the 103 excitation energies, for which the estimated MADs for the 30 methods correlate with the actual MADs with a criterion of $R^2 > 0.995$. The resulting model, which we term XE6 (Excitation Energy 6), contains just six excitation energies from five molecules. The linear fit has an R^2 of 0.997. The parameters for obtaining the estimated MAD for the complete set of 103 energies using the XE6 model (EMAD_{XE6}) are shown in Table S3.

Table S3.	Details for	Obtaining	EMAD _{XE6}	as an	Estimation	of the N	MAD f	or a	Set
of 103 Sin	glet Excitat	ion Energie	es						

Molecule	Excitation	Ref (eV)	Coefficient
<i>E</i> -Butadiene	1 A_g $(\pi \rightarrow \pi^*)$	6.55	0.1436858
Pyrazine	1 B_{2u} $(\pi \rightarrow \pi^*)$	4.64	0.0977290
Pyrazine	1 B_{1g} $(n \rightarrow \pi^*)$	6.60	0.1055256
<i>s</i> -Tetrazine	2 A_u $(n \rightarrow \pi^*)$	5.50	0.0922092
Cytosine	1 A'' $(n \rightarrow \pi^*)$	4.87	0.0070462
Uracil	2 A'' $(n \rightarrow \pi^*)$	6.10	0.1499667
Intercept (eV)			0.0298189

As case studies, we use the XE6 model to assess the accuracy of time-dependent calculations with two screened-exchange (SX) DFT methods that we have recently formulated. They are HSEB²⁸ and reHISSB.²⁹ It is noteworthy that the set of 30 DFT

methods used for deriving XE6 does not contain SX-DFT. However, the HSE SX-DFT method,³⁰ which is closely related to HSEB (and to reHISSB but to a somewhat lesser degree), has been independently assessed with the full set of 103 excitation energies,³¹ for which its MAD is 0.23 eV. Such a performance is on par with some of the most accurate hybrid functionals (~ 0.2–0.3 eV) among the 30 methods for this set of singlet excitation energies, and is substantially better than typical pure DFT methods (> 0.5 eV). In comparison, the EMAD_{XE6} values for HSEB and reHISSB are 0.19 and 0.21 eV, respectively, which is what one might expect given their similarity to HSE. This again lends support to our premise that statistically down-sized data sets can provide an accurate means for evaluating theoretical methodologies.

References

- 1 Geary, R. C. *Biometrika* **1935**, *27*, 310–332.
- 2 Goerigk, L.; Hansen. A.; Bauer, C. A.; Ehrlich, S.; Najibi, A.; Grimme, S. A Look at the Density Functional Theory Zoo with the Advanced GMTKN55 Database for General Main Group Thermochemistry, Kinetics and Noncovalent Interactions. *Phys. Chem. Chem. Phys.* **2017**, *19*, 32184–32215.
- 3 Mardirossian, N.; Head-Gordon, M. Thirty Years of Density Functional Theory in Computational Chemistry: An Overview and Extensive Assessment of 200 Density Functionals. *Mol. Phys.* 2017, 115, 2315–2372.
- 4 Grimme, S. Density Functional Theory with London Dispersion Corrections. *WIREs Comput. Mol. Sci.* **2011**, *1*, 211–228.
- 5 Perdew, J. P.; Wang, Y. Accurate and Simple Analytic Representation of the Electron-Gas Correlation Energy. *Phys. Rev. B* **1992**, *45*, 13244–13249.
- 6 Lee, C.; Yang, W.; Parr, R. G. Development of the Colic-Salvetti Correlation-Energy Formula into a Functional of the Electron Density. *Phys. Rev. B* **1988**, *37*, 785–789.
- 7 Slater, J. C. *The Self-Consistent Field for Molecular and Solids, Quantum Theory of Molecular and Solids,* Vol. 4; McGraw-Hill: New York, 1974.
- 8 Vosko, S. H.; Wilk, L.; Nusair, M. Accurate Spin-Dependent Electron Liquid Correlation Energies for Local Spin Density Calculations: A Critical Analysis. *Can. J. Phys.* **1980**, *58*, 1200-1211.
- Handy, N. C.; Cohen, A, J. Left-Right Correlation Energy. *Mol. Phys.* 2001, 99, 403–412.
- 10 Mardirossian, N.; Head-Gordon, M. *ω*B97X-V: A 10-Parameter, Range-Separated Hybrid, Generalized Gradient Approximation Density Functional with Nonlocal

Correlation, Designed by a Survival-of-the-Fittest Strategy. *Phys. Chem. Chem. Phys.* **2014**, *16*, 9904–9924.

- 11 Chan, B.; Kawashima, Y.; Hirao, K. Correlation Functional in Screened-Exchange Density Functional Theory Procedures. *J. Comput. Chem.* **2017**, *38*, 2307–2315.
- 12 Heyd, J.; Scuseria, G. E.; Ernzerhof, M. Hybrid Functionals Based on a Screened Coulomb Potential. *J. Chem. Phys.* **2003**, *118*, 8207–8215.
- Izmaylov, A. F.; Scuseria, G.; Frisch, M. J. Efficient Evaluation of Short-Range Hartree-Fock Exchange in Large Molecules and Periodic Systems. *J. Chem. Phys.* 2006, 125, 104103-1–8.
- 14 Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.
- 15 Becke, A. D. Density-Functional Thermochemistry. V. Systematic Optimization of Exchange–Correlation Functionals. *J. Chem. Phys.* **1997**, *107*, 8554–8560.
- 16 Grimme, S. Semiempirical GGA-Type Density Functional Constructed with a Long-Range Dispersion Correction. *J. Comput. Chem.* **2006**, *27*, 1787–1799.
- Lin, C. Y.; George, M. W.; Gill, P. M. W. EDF2: A Density Functional for Predicting Molecular Vibrational Frequencies. *Aust. J. Chem.* 2004, *57*, 365–370.
- 18 Wiitala, K. W.; Hoye, T. R.; Cramer, C. J. Hybrid Density Functional Methods Empirically Optimized for the Computation of 13C and 1H Chemical Shifts in Chloroform Solution. J. Chem. Theory Comput. 2006, 2, 1085–1092.
- 19 Austin, A.; Petersson, G.; Frisch, M. J.; Dobek, F. J.; Scalmani, G.; Throssell, K. A Density Functional with Spherical Atom Dispersion Terms. *J. Chem. Theory Comput.* 2012, *8*, 4989–5007.
- 20 Becke, A. D. Density-Functional Exchange-Energy Approximation with Correct Asymptotic Behavior. *Phys. Rev. A.* **1988**, *38*, 3098–3100.

- 21 Adamo, C.; Barone, V. Exchange Functionals with Improved Long-Range Behavior and Adiabatic Connection Methods Without Adjustable Parameters: The MPW and MPW1PW Models. J. Chem. Phys. 1998, 108, 664–675.
- 22 Loos, P.-F. Exchange Functionals Based on Finite Uniform Electron Gases. *J. Chem. Phys.* **2017**, *146*, 114108-1–7.
- 23 Sun, J.; Ruzsinszky, A.; Perdew, J. P. Strongly Constrained and Appropriately Normed Semilocal Density Functional. *Phys. Rev. Lett.* **2015**, *115*, 036402-1–6.
- 24 Perdew, J. P.; Ruzsinszky, A.; Csonka, G. I.; Constantin, L. A.; Sun, J. Workhorse Semilocal Density Functional for Condensed Matter Physics and Quantum Chemistry. *Phys. Rev. Lett.* **2009**, *103*, 026403-1–4.
- 25 Peverati, R.; Truhlar, D. G. Performance of the M11-L Density Functional for Bandgaps and Lattice Constants of Unary and Binary Semiconductors. *J. Chem. Phys.* **2012**, *136*, 134704-1–10.
- 26 Schreiber, M.; Silva-Junior, M. R.; Sauer, S. P. A.; Thiel, W. Benchmarks for Electronically Excited States: CASPT2, CC2, CCSD, and CC3. *J. Chem. Phys.* 2008, *128*, 134110-1–25.
- 27 Jacquemin, D.; Wathelet, V.; Perpète, E. A.; Adamo, C Extensive TD-DFT Benchmark: Singlet-Excited States of Organic Molecules. *J. Chem. Theory Comput.* 2009, 5, 2420–2435
- 28 Chan, B.; Kawashima, Y.; Hirao, K. Correlation Functional in Screened-Exchange Density Functional Theory Procedures. *J. Comput. Chem.* **2017**, *38*, 2307–2315.
- 29 Chan, B.; Kawashima, Y.; Hirao, K. The reHISS Three-Range Exchange Functional with an Optimal Variation of Hartree–Fock and Its Use in the reHISSB-D Density Functional Theory Method. *J. Comput. Chem.* DOI:10.1002/jcc.25383.

- 30 Izmaylov, A. F.; Scuseria, G.; Frisch, M. J. Efficient Evaluation of Short-Range Hartree-Fock Exchange in Large Molecules and Periodic Systems, *J. Chem. Phys.* 2006, *125*, 104103-1–8.
- 31 Peverati, R.; Truhlar, D. G. Performance of the M11 and M11-L Density Functionals for Calculations of Electronic Excitation Energies by Adiabatic Time-Dependent Density Functional Theory. *Phys. Chem. Chem. Phys.* 2012, 14, 11363–11370.