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

Supporting information for paper Non-Covalent Interactions Atlas Benchmark Data Sets 5: London Dispersion in an Extended Chemical Space

Jan Řezáč*

Institute of Organic Chemistry and Biochemistry, Czech Academy of Sciences, 166 10 Prague, Czech Republic

E-mail: rezac@uochb.cas.cz

May 30, 2022

Contents

1 Data provided in separate files

The following ZIP archives are provided as a part of the Supporting Information:

- **ncia_dispersion_SI_002.zip, ncia_dispersion_SI_003.zip** contain geometries of all the systems as .xyz files, and tables of interaction energies and their components (tab-delimited text files) for D1200 and D442×10 data sets, respectively. File names use the numbering defined in the paper, and include also additional information. The headers of the .xyz files contain the definition of the two monomers in the complex, the benchmark interaction energy and selected other metadata.

- **ncia_dispersion_SI_004.zip, ncia_dispersion_SI_005.zip** contain structured, machine-readable but human friendly data file in YAML format (https://yaml.org/) defining and describing the data sets. This file format is used by the Cuby framework (http://cuby4.molecular.cz), but it can be easily used on its own. The file contains all the metadata describing the systems, such as their assignment to groups. The last part of the file lists all the intermediate interaction energy components used to construct the benchmark, and results of some additional calculations discussed in the paper. (Please note that up to date version of the file will be bundled with the Cuby framework.)

Up to date version of all these files are also available at the GitHub repository `https://github.com/Honza-R/NCIAtlas`.

2 Composition of the D1200 and D442×10 data sets and its subsets

Table S1: Monomers used in the construction of the D1200 data set, listed by the element and monomer class they are representing.

Element	Class	Molecule name
H	H	Hydrogen molecule
B	B, saturated	Arachnopentaborane
	B, saturated	Diborane
	B, saturated	Nidopentaborane
	B, aromatic	Borazine
C	C, alkane	2,3,-dimethylbutane
	C, alkane	Butane
	C, alkane	Ethane
	C, alkane	Hexane
	C, alkane	Isopentane
	C, alkane	Pentane
	C, alkane	Propane
	C, alkane, cyclic	Cyclohexane
	C, alkane, cyclic	Cyclopentane
	C, alkane, 2-branched	Neohexane
	C, alkane, 2-branched	Neopentane
	C, alkene	Butadiene
	C, alkene	Cyclobutadiene
	C, alkene	Cyclopentadiene
	C, alkene	Ethene
	C, alkene	Nonbornadiene
	C, alkene	Propene
	C, alkyne	Butadyine
	C, alkyne	Ethyne
	C, alkyne	Propyne
	C, aromatic	Benzene
	C, aromatic	Toluene
	C, aromatic	Naphthalene
N	N, single bonds	Ammonia
	N, single bonds	Hydrazine
	N, single bonds	Methylamine
	N, with double bond	1,2,4-triazole
	N, with double bond	Diazene
	N, with double bond	Hydrogen azide
	N, with double bond	Methyl azide
	N, with triple bond	Diazomethane
	N, with triple bond	Nitrogen molecule
	N, aromatic	1,3,5-triazine
O	O, single bonds	Diacetonediperoxide
	O, single bonds	Dimethylether
	O, single bonds	Dimethylperoxide

	O, single bonds	Hydogen peroxide
	O, single bonds	Trioxane
	O, with double bond	Acetone
	O, with double bond	Carbon dioxide
	O, with double bond	Formaldehyde
P	P, single bonds	Diphosphine
	P, single bonds	Phosphine
	P, single bonds	Phosphorus molecule, P_4
	P, with double bond	Diphosphene
	P, aromatic	Phosphorine
S	S, single bonds	Dimethyldisulfide
	S, single bonds	Dimethylsulfide
	S, single bonds	Dithiethane
	S, single bonds	Sulfur molecule, S_8
	S, with double bond	Carbon disulfide
	S, with double bond	Thioacetone
	S, with double bond	Trithiane
	S, aromatic	Thiophene
F	F, monovalent	Difluoromethane
	F, monovalent	Fluorine molecule
	F, monovalent	Fluoromethane
	F, monovalent	Hexafluorobenzene
	F, monovalent	Tetrafluoroethylene
	F, monovalent	Tetrafluoromethane
	F, monovalent	Trifluoromethane
Cl	Cl, monovalent	Chlorine molecule
	Cl, monovalent	Chloromethane
	Cl, monovalent	Dichloromethane
	Cl, monovalent	Tetrachloroethylene
	Cl, monovalent	Tetrachloromethane
	Cl, monovalent	Trichloromethane
Br	Br, monovalent	Bromine molecule
	Br, monovalent	Bromomethane
	Br, monovalent	Dibromomethane
	Br, monovalent	Tetrabromomethane
	Br, monovalent	Tribromomethane
I	Br, monovalent	Diiodomethane
	Br, monovalent	Iodine molecule
	Br, monovalent	Iodomethane
	Br, monovalent	Tetraiodomethane
	Br, monovalent	Triiodomethane
He	He, free atom	Helium
Ne	Ne, free atom	Neon
Ar	Ar, free atom	Argon
Kr	Kr, free atom	Krypton
Xe	Xe, free atom	Xenon

Table S2: Counts of the element combinations covered by the D1200 data set.

H	1															
C	18	64														
N	4	85	11													
O	1	40	12	3												
B	4	35	15	7	8											
P	4	58	20	10	15	7										
S	7	71	24	14	18	24	21									
F	4	14	11	5	8	10	10	3								
Cl	4	18	14	4	8	12	12	3	4							
Br	3	17	10	5	7	12	10	3	4	3						
I	4	18	15	8	8	12	12	3	4	4	4					
He	1	12	8	5	4	4	7	3	3	3	3	1				
Ne	1	12	9	6	4	5	7	3	3	3	3	1	1			
Ar	1	12	9	6	4	5	7	3	3	3	3	1	1	1		
Kr	1	12	9	6	4	5	7	3	3	3	3	1	1	1	1	
Xe	1	11	8	6	3	5	7	3	3	3	3	1	1	1	1	1
	H	C	N	O	B	P	S	F	Cl	Br	I	He	Ne	Ar	Kr	Xe

Table S3: Counts of the element combinations covered by the D442×10 data set.

H	1															
C	6	21														
N	2	23	7													
O	1	11	5	2												
B	2	12	6	4	2											
P	3	18	10	5	6	3										
S	3	18	10	6	6	9	6									
F	1	6	4	2	2	3	3	1								
Cl	1	6	4	2	2	3	3	1	1							
Br	1	6	4	2	2	3	3	1	1	1						
I	1	6	4	2	2	3	3	1	1	1	1					
He	1	6	4	2	2	3	3	1	1	1	1	1				
Ne	1	6	4	2	2	3	3	1	1	1	1	1	1			
Ar	1	6	4	2	2	3	3	1	1	1	1	1	1	1		
Kr	1	6	4	2	2	3	3	1	1	1	1	1	1	1	1	
Xe	1	6	4	2	2	3	3	1	1	1	1	1	1	1	1	1
	H	C	N	O	B	P	S	F	Cl	Br	I	He	Ne	Ar	Kr	Xe

Table S4: Assignment of the systems to the smaller subsets obtained by clustering analysis (20, 50, 100 and 200 systems). (Note that this information is provided also in the data files.)

Subset size	Systems
20	1.02.18, 1.06.37, 1.06.58, 2.03.52, 2.04.55, 2.05.14, 2.12.24, 2.13.11, 3.09.04, 3.13.02, 4.03.01, 4.07.12, 4.08.01, 4.19.05, 4.29.01, 4.31.05, 4.34.01, 4.35.01, 4.44.01, 4.56.01
50	1.02.18, 1.06.43, 1.06.58, 1.07.45, 1.07.78, 1.09.35, 1.15.01, 2.02.06, 2.03.10, 2.03.18, 2.03.57, 2.04.14, 2.04.38, 2.06.20, 2.12.14, 2.12.24, 2.13.08, 3.04.02, 3.05.14, 3.09.04, 3.09.08, 3.09.10, 3.11.04, 3.13.02, 3.17.03, 3.20.04, 3.23.06, 3.27.10, 3.29.03, 3.30.01, 4.03.01, 4.07.12, 4.08.01, 4.09.01, 4.09.06, 4.12.06, 4.19.05, 4.27.02, 4.27.05, 4.28.01, 4.29.01, 4.29.04, 4.31.05, 4.33.02, 4.35.07, 4.42.03, 4.44.01, 4.56.01, 4.62.01, 4.64.01
100	1.01.01, 1.02.01, 1.03.03, 1.05.01, 1.06.58, 1.07.21, 1.07.56, 1.07.68, 1.07.78, 1.08.05, 1.08.11, 1.08.20, 1.08.33, 1.09.13, 1.10.02, 1.11.04, 1.11.11, 1.11.12, 1.14.02, 1.15.01, 2.01.02, 2.01.04, 2.02.04, 2.02.06, 2.03.10, 2.03.13, 2.03.34, 2.04.22, 2.04.38, 2.04.55, 2.05.05, 2.05.11, 2.05.17, 2.06.08, 2.06.15, 2.07.09, 2.09.10, 2.11.01, 2.11.05, 2.12.16, 2.12.20, 2.12.24, 2.13.08, 3.01.04, 3.02.04, 3.04.02, 3.05.04, 3.05.11, 3.05.14, 3.06.08, 3.07.07, 3.08.05, 3.08.15, 3.09.04, 3.09.08, 3.10.01, 3.10.05, 3.11.06, 3.11.08, 3.13.02, 3.15.04, 3.17.03, 3.23.09, 3.24.04, 3.28.08, 3.29.03, 3.30.01, 3.37.04, 4.01.01, 4.02.01, 4.07.12, 4.09.01, 4.09.06, 4.12.06, 4.12.09, 4.13.01, 4.13.04, 4.14.02, 4.14.07, 4.19.05, 4.24.01, 4.25.01, 4.26.03, 4.27.02, 4.27.05, 4.28.01, 4.28.04, 4.30.01, 4.31.05, 4.32.04, 4.34.02, 4.35.03, 4.35.07, 4.42.03, 4.48.02, 4.56.01, 4.61.01, 4.64.01, 4.66.01, 4.70.01
200	1.01.01, 1.02.01, 1.02.10, 1.02.13, 1.03.01, 1.03.03, 1.04.01, 1.05.01, 1.06.01, 1.06.17, 1.06.40, 1.06.50, 1.06.64, 1.07.13, 1.07.21, 1.07.32, 1.07.38, 1.07.45, 1.07.56, 1.07.68, 1.07.70, 1.07.78, 1.07.83, 1.08.05, 1.08.11, 1.08.20, 1.08.33, 1.09.13, 1.09.16, 1.09.21, 1.09.29, 1.10.02, 1.10.10, 1.10.11, 1.11.03, 1.11.04, 1.11.09, 1.11.11, 1.11.12, 1.12.05, 1.12.15, 1.13.02, 1.13.03, 1.15.01, 1.15.08, 2.01.01, 2.01.02, 2.01.04, 2.02.04, 2.03.10, 2.03.13, 2.03.23, 2.03.31, 2.03.34, 2.03.49, 2.04.09, 2.04.20, 2.04.35, 2.04.48, 2.05.05, 2.05.06, 2.05.11, 2.05.14, 2.05.17, 2.05.19, 2.05.20, 2.06.15, 2.06.18, 2.06.20, 2.06.22, 2.07.07, 2.07.09, 2.07.10, 2.08.06, 2.08.11, 2.08.14, 2.09.10, 2.10.02, 2.10.07, 2.11.01, 2.11.05, 2.12.04, 2.12.07, 2.12.14, 2.12.16, 2.12.20, 2.13.02, 2.13.08, 3.01.04, 3.02.04, 3.04.02, 3.05.04, 3.05.09, 3.06.08, 3.07.02, 3.07.05, 3.07.07, 3.07.16, 3.08.01, 3.08.15, 3.09.04, 3.09.08, 3.10.01, 3.10.05, 3.10.13, 3.11.02, 3.11.04, 3.11.06, 3.11.08, 3.12.08, 3.12.09, 3.12.14, 3.13.02, 3.13.03, 3.15.04, 3.16.03, 3.17.03, 3.17.06, 3.18.07, 3.19.03, 3.19.07, 3.21.01, 3.21.06, 3.21.10, 3.22.02, 3.22.10, 3.23.03, 3.23.06, 3.24.05, 3.24.10, 3.25.01, 3.25.05, 3.26.05, 3.27.10, 3.28.08, 3.28.10, 3.29.03, 3.30.01, 3.31.02, 3.37.04, 4.01.01, 4.02.01, 4.03.01, 4.04.01, 4.07.04, 4.07.05, 4.08.01, 4.08.03, 4.08.06, 4.08.12, 4.09.01, 4.09.06, 4.10.07, 4.12.01, 4.12.02, 4.12.06, 4.13.04, 4.13.09, 4.14.02, 4.14.07, 4.14.08, 4.15.03, 4.15.08, 4.17.04, 4.19.05, 4.20.04, 4.23.01, 4.23.04, 4.24.01, 4.25.01, 4.26.03, 4.27.01, 4.27.02, 4.28.04, 4.30.01, 4.30.02, 4.31.05, 4.32.04, 4.33.02, 4.34.01, 4.34.07, 4.35.03, 4.35.07, 4.39.01, 4.41.01, 4.42.03, 4.43.02, 4.44.01, 4.48.02, 4.49.02, 4.54.01, 4.56.01, 4.61.01, 4.62.01, 4.63.01, 4.64.01, 4.65.01, 4.68.01, 4.69.01, 4.70.01

3 Results: Wavefunction Methods

Table S5: Errors of Hartree-Fock and MP2 calculations in the D1200 data set. Root mean square error, mean unsigned error and mean signed error, all in kcal/mol.

Method	RMSE	MUE	MSE
HF/aQZ	4.444	3.787	3.787
HF/a5Z	4.442	3.786	3.786
MP2/aTZ	0.670	0.432	-0.400
MP2/aQZ	0.798	0.538	-0.522
MP2/a5Z	0.847	0.579	-0.567
MP2/CBS(aTQZ)	0.894	0.619	-0.610
MP2/CBS(aQ5Z)	0.897	0.621	-0.612
SCS-MP2/CBS(aQ5Z)	0.462	0.379	0.358
SCS-MI-MP2/CBS(aQ5Z)	0.319	0.242	0.127

Table S6: RMSE of Hartree-Fock and MP2 calculations in the D1200 data set and the groups it comprises. Values in kcal/mol.

Method	All	HBCNO	PS	Halogens	NobleGases
HF/aQZ	4.444	4.214	5.628	5.152	1.575
HF/a5Z	4.442	4.213	5.626	5.150	1.574
MP2/aTZ	0.670	0.527	0.970	0.738	0.180
MP2/aQZ	0.798	0.595	1.120	0.934	0.249
MP2/a5Z	0.847	0.621	1.177	1.007	0.276
MP2/CBS(aTQZ)	0.894	0.647	1.228	1.080	0.304
MP2/CBS(aQ5Z)	0.897	0.648	1.234	1.082	0.304
SCS-MP2/CBS(aQ5Z)	0.462	0.591	0.506	0.459	0.165
SCS-MI-MP2/CBS(aQ5Z)	0.319	0.395	0.379	0.305	0.110

Table S7: Average benchmark interaction energies E_{int} and SAPT0 dispersion and exchange-repulsion terms in selected subsets of the D1200 data set. First we consider interaction between hydrocarbons divided into saturated and unsaturated (labeled π). Next, we select all complexes of hydrocarbons with molecules involving an element from the series of halogens and noble gases. (all values in kcal/mol)

Subset	E_{int}	$E_{disp.}^{SAPT0}$	$E_{exch.}^{SAPT0}$
Hydrocarbons			
sat. – sat.	-2.502	-3.911	3.811
sat. – π	-2.390	-3.710	3.752
π – π	-3.013	-4.814	4.953
Interactions involving specific element			
F	-2.239	-2.703	3.042
Cl	-2.580	-3.406	3.871
Br	-2.689	-3.728	4.383
I	-3.543	-5.781	6.513
Ne	-0.331	-0.220	0.472
Ar	-0.799	-0.827	1.091
Kr	-1.139	-1.389	1.703
Xe	-1.340	-2.288	2.163

4 Results: DFT in the D1200 Data Set

Figure S1: Systematic error (MSE in kcal/mol) of DFT calculations in the D1200 data set and its groups. The DFT functionals are sorted by ascending RMSE in this data set. The labels of the functionals are coloured as follows: double-hybrids red, range-separated hybrids blue, hybrids black, GGA and meta-GGA green.

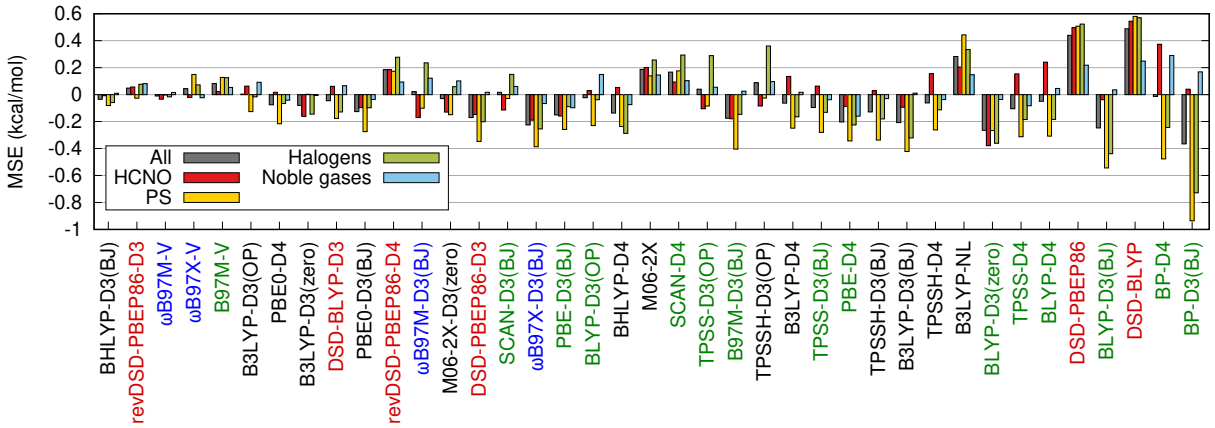

Table S8: Error (RMSE in kcal/mol) of DFT calculations in the D1200 data set and its groups. The DFT functionals are sorted by ascending RMSE in this data set. These data are plotted in the paper in Fig. 2.

Functional	All	HBCNO	PS	Halogens	Noble gases
BHLYP-D3(BJ)	0.135	0.120	0.171	0.161	0.052
revDSD-PBEP86-D3	0.149	0.140	0.167	0.177	0.099
ωB97M-V	0.156	0.166	0.172	0.176	0.089
ωB97X-V	0.189	0.129	0.255	0.228	0.097
B97M-V	0.212	0.187	0.275	0.241	0.101
B3LYP-D3(OP)	0.224	0.162	0.278	0.289	0.112
PBE0-D4	0.244	0.198	0.347	0.265	0.085
B3LYP-D3(zero)	0.245	0.243	0.304	0.281	0.086
DSD-BLYP-D3	0.256	0.182	0.340	0.322	0.095
PBE0-D3(BJ)	0.257	0.201	0.372	0.282	0.077
revDSD-PBEP86-D4	0.259	0.250	0.266	0.341	0.127
ωB97M-D3(BJ)	0.278	0.247	0.246	0.379	0.201
M06-2X-D3(zero)	0.283	0.301	0.299	0.284	0.244
DSD-PBEP86-D3	0.290	0.223	0.427	0.316	0.067
SCAN-D3(BJ)	0.300	0.262	0.303	0.400	0.188
ωB97X-D3(BJ)	0.314	0.245	0.441	0.364	0.090
PBE-D3(BJ)	0.318	0.281	0.407	0.373	0.142
BLYP-D3(OP)	0.323	0.209	0.422	0.404	0.172
BHLYP-D4	0.327	0.198	0.373	0.468	0.162
M06-2X	0.327	0.336	0.312	0.377	0.270
SCAN-D4	0.335	0.244	0.342	0.472	0.212
TPSS-D3(OP)	0.340	0.231	0.309	0.503	0.234
B97M-D3(BJ)	0.343	0.272	0.502	0.372	0.073
TPSSH-D3(OP)	0.356	0.207	0.283	0.557	0.261
B3LYP-D4	0.358	0.311	0.458	0.433	0.131
TPSS-D3(BJ)	0.360	0.245	0.492	0.452	0.093
PBE-D4	0.364	0.249	0.493	0.437	0.180
TPSSH-D3(BJ)	0.364	0.230	0.517	0.448	0.088
B3LYP-D3(BJ)	0.380	0.187	0.529	0.501	0.095
TPSSH-D4	0.380	0.330	0.513	0.435	0.110
B3LYP-NL	0.384	0.267	0.514	0.453	0.218
BLYP-D3(zero)	0.417	0.448	0.468	0.501	0.129
TPSS-D4	0.441	0.345	0.588	0.537	0.141
BLYP-D4	0.508	0.458	0.638	0.610	0.188
DSD-PBEP86	0.512	0.565	0.565	0.581	0.250
BLYP-D3(BJ)	0.541	0.243	0.734	0.736	0.154
DSD-BLYP	0.563	0.615	0.637	0.623	0.288
BP-D4	0.743	0.609	0.909	0.914	0.403
BP-D3(BJ)	0.979	0.439	1.270	1.359	0.397

Table S9: Signed (MSE in kcal/mol) of DFT calculations in the D1200 data set and its groups. The DFT functionals are sorted by ascending RMSE in this data set. These data are plotted here in Fig. S1.

Functional	All	HBCNO	PS	Halogens	Noble gases
BHLYP-D3(BJ)	-0.035	-0.009	-0.081	-0.059	0.010
revDSD-PBEP86-D3	0.048	0.057	-0.027	0.077	0.081
ωB97M-V	-0.010	-0.034	-0.002	-0.017	0.015
ωB97X-V	0.044	-0.021	0.149	0.072	-0.023
B97M-V	0.082	0.023	0.127	0.125	0.053
B3LYP-D3(OP)	0.003	0.063	-0.126	-0.017	0.091
PBE0-D4	-0.076	0.017	-0.216	-0.066	-0.042
B3LYP-D3(zero)	-0.080	-0.162	0.002	-0.145	-0.006
DSD-BLYP-D3	-0.045	0.061	-0.177	-0.129	0.066
PBE0-D3(BJ)	-0.126	-0.097	-0.275	-0.098	-0.036
revDSD-PBEP86-D4	0.184	0.185	0.173	0.277	0.093
ωB97M-D3(BJ)	0.022	-0.170	-0.101	0.236	0.122
M06-2X-D3(zero)	-0.030	-0.130	-0.149	0.058	0.102
DSD-PBEP86-D3	-0.170	-0.148	-0.347	-0.201	0.017
SCAN-D3(BJ)	0.017	-0.114	-0.029	0.150	0.060
ωB97X-D3(BJ)	-0.225	-0.189	-0.387	-0.255	-0.066
PBE-D3(BJ)	-0.151	-0.159	-0.258	-0.089	-0.098
BLYP-D3(OP)	-0.023	0.030	-0.230	-0.038	0.149
BHLYP-D4	-0.137	0.053	-0.236	-0.288	-0.075
M06-2X	0.187	0.201	0.139	0.257	0.146
SCAN-D4	0.167	0.093	0.176	0.293	0.103
TPSS-D3(OP)	0.040	-0.105	-0.085	0.289	0.055
B97M-D3(BJ)	-0.176	-0.179	-0.404	-0.146	0.025
TPSSH-D3(OP)	0.088	-0.085	-0.026	0.360	0.096
B3LYP-D4	-0.064	0.135	-0.249	-0.165	0.017
TPSS-D3(BJ)	-0.096	0.063	-0.280	-0.132	-0.038
PBE-D4	-0.203	-0.088	-0.344	-0.225	-0.160
TPSSH-D3(BJ)	-0.129	0.029	-0.337	-0.180	-0.030
B3LYP-D3(BJ)	-0.208	-0.094	-0.422	-0.322	0.011
TPSSH-D4	-0.062	0.155	-0.262	-0.113	-0.037
B3LYP-NL	0.282	0.204	0.443	0.334	0.147
BLYP-D3(zero)	-0.265	-0.379	-0.267	-0.361	-0.036
TPSS-D4	-0.105	0.154	-0.312	-0.185	-0.083
BLYP-D4	-0.050	0.240	-0.308	-0.184	0.046
DSD-PBEP86	0.439	0.496	0.507	0.523	0.218
BLYP-D3(BJ)	-0.247	-0.038	-0.544	-0.438	0.035
DSD-BLYP	0.488	0.544	0.580	0.570	0.248
BP-D4	-0.014	0.373	-0.477	-0.244	0.290
BP-D3(BJ)	-0.366	0.040	-0.936	-0.728	0.169

5 Results: DFT in the D442×10 Data Set

Table S10: Error (RMSE in kcal/mol) of DFT calculations in the D1200 data set, its subset D442 and in the D442×10 set of dissociation curves.

Functional	D1200	D442	D442×10
Double-hybrid functionals			
revDSD-PBEP86-D4	0.259	0.248	0.307
revDSD-PBEP86-D3	0.149	0.148	0.199
DSD-BLYP-D3	0.256	0.257	0.339
DSD-PBEP86-D3	0.29	0.287	0.367
DSD-PBEP86	0.512	0.488	0.477
DSD-BLYP	0.563	0.533	0.503
Range-separated hybrid functionals			
ωB97M-V	0.156	0.143	0.194
ωB97X-V	0.189	0.181	0.444
ωB97M-D3(BJ)	0.278	0.258	0.39
ωB97X-D3(BJ)	0.314	0.301	0.371
Hybrid functionals			
B3LYP-D3(BJ)	0.38	0.369	0.368
B3LYP-D3(zero)	0.245	0.244	0.718
B3LYP-D3(OP)	0.224	0.222	0.387
B3LYP-D4	0.358	0.354	0.413
B3LYP-NL	0.384	0.362	0.534
BHLYP-D3(BJ)	0.135	0.131	0.347
BHLYP-D4	0.327	0.317	0.391
PBE0-D3(BJ)	0.257	0.253	0.369
PBE0-D4	0.244	0.24	0.36
M06-2X-D3(zero)	0.283	0.266	0.371
M06-2X	0.327	0.316	0.394
TPSSH-D3(BJ)	0.364	0.362	0.537
TPSSH-D3(OP)	0.356	0.336	0.632
TPSSH-D4	0.38	0.376	0.549
Meta-GGA functionals			
TPSS-D3(BJ)	0.36	0.358	0.569
TPSS-D3(OP)	0.34	0.327	0.631
TPSS-D4	0.441	0.434	0.638
B97M-V	0.212	0.203	0.467
B97M-D3(BJ)	0.343	0.333	0.419
SCAN-D3(BJ)	0.3	0.282	0.374
SCAN-D4	0.335	0.312	0.403
GGA functionals			
BLYP-D3(BJ)	0.541	0.526	0.518
BLYP-D3(zero)	0.417	0.408	0.82
BLYP-D3(OP)	0.323	0.322	0.412
BLYP-D4	0.508	0.5	0.527
PBE-D3(BJ)	0.318	0.318	0.472
PBE-D4	0.364	0.36	0.488
BP-D3(BJ)	0.979	0.947	1.039
BP-D4	0.743	0.732	0.794
Average	0.353	0.342	0.475

Table S11: Error (RMSE in kcal/mol) of DFT calculations in the D442×10 data set and its groups. The DFT functionals are sorted by ascending RMSE in this data set.

Functional	All	HBCNO	PS	Halogens	Noble gases
ωB97M-V	0.194	0.174	0.251	0.230	0.121
revDSD-PBEP86-D3	0.199	0.181	0.257	0.239	0.118
revDSD-PBEP86-D4	0.307	0.282	0.303	0.451	0.183
DSD-BLYP-D3	0.339	0.219	0.509	0.430	0.109
BHLYP-D3(BJ)	0.347	0.337	0.416	0.427	0.213
PBE0-D4	0.360	0.296	0.507	0.414	0.199
DSD-PBEP86-D3	0.367	0.248	0.602	0.395	0.090
B3LYP-D3(BJ)	0.368	0.277	0.473	0.461	0.251
PBE0-D3(BJ)	0.369	0.296	0.511	0.441	0.203
M06-2X-D3(zero)	0.371	0.454	0.426	0.398	0.199
ωB97X-D3(BJ)	0.371	0.225	0.552	0.477	0.140
SCAN-D3(BJ)	0.374	0.323	0.423	0.502	0.250
B3LYP-D3(OP)	0.387	0.394	0.397	0.495	0.274
ωB97M-D3(BJ)	0.390	0.264	0.339	0.609	0.307
BHLYP-D4	0.391	0.409	0.467	0.478	0.210
M06-2X	0.394	0.411	0.486	0.466	0.212
SCAN-D4	0.403	0.291	0.430	0.594	0.278
BLYP-D3(OP)	0.412	0.302	0.532	0.484	0.318
B3LYP-D4	0.413	0.416	0.454	0.505	0.292
B97M-D3(BJ)	0.419	0.331	0.674	0.434	0.101
ωB97X-V	0.444	0.333	0.633	0.550	0.203
B97M-V	0.467	0.282	0.624	0.636	0.263
PBE-D3(BJ)	0.472	0.409	0.614	0.593	0.250
DSD-PBEP86	0.477	0.557	0.515	0.587	0.241
PBE-D4	0.488	0.377	0.727	0.560	0.216
DSD-BLYP	0.503	0.593	0.563	0.587	0.266
BLYP-D3(BJ)	0.518	0.371	0.697	0.644	0.329
BLYP-D4	0.527	0.512	0.687	0.583	0.326
B3LYP-NL	0.534	0.408	0.711	0.666	0.322
TPSSH-D3(BJ)	0.537	0.394	0.808	0.591	0.282
TPSSH-D4	0.549	0.422	0.865	0.545	0.262
TPSS-D3(BJ)	0.569	0.417	0.860	0.630	0.284
TPSS-D3(OP)	0.631	0.399	0.455	0.985	0.579
TPSSH-D3(OP)	0.632	0.361	0.434	1.012	0.581
TPSS-D4	0.638	0.450	1.032	0.655	0.251
B3LYP-D3(zero)	0.718	0.442	1.027	0.899	0.403
BP-D4	0.794	0.613	1.148	0.866	0.487
BLYP-D3(zero)	0.820	0.499	1.099	1.069	0.529
BP-D3(BJ)	1.039	0.520	1.597	1.312	0.420

Table S12: Systematic error (MSE in kcal/mol) of DFT calculations in the D442×10 data set and its groups. The DFT functionals are sorted by ascending RMSE in this data set.

Functional	All	HBCNO	PS	Halogens	Noble gases
ωB97M-V	0.020	-0.014	0.043	0.031	0.021
revDSD-PBEP86-D3	0.052	0.080	-0.051	0.096	0.078
revDSD-PBEP86-D4	0.185	0.197	0.160	0.323	0.102
DSD-BLYP-D3	-0.042	0.073	-0.216	-0.120	0.052
BHLYP-D3(BJ)	0.095	0.106	0.096	0.134	0.061
PBE0-D4	-0.030	0.047	-0.244	0.045	0.021
DSD-PBEP86-D3	-0.150	-0.120	-0.371	-0.180	0.010
B3LYP-D3(BJ)	-0.030	0.025	-0.241	-0.053	0.099
PBE0-D3(BJ)	-0.069	-0.049	-0.288	0.006	0.028
M06-2X-D3(zero)	-0.063	-0.211	-0.074	0.007	0.010
ωB97X-D3(BJ)	-0.154	-0.090	-0.339	-0.186	-0.046
SCAN-D3(BJ)	0.011	-0.115	-0.067	0.159	0.063
B3LYP-D3(OP)	0.148	0.197	0.050	0.199	0.150
ωB97M-D3(BJ)	0.071	-0.132	-0.031	0.309	0.139
BHLYP-D4	0.052	0.183	0.005	0.015	0.013
M06-2X	0.125	0.088	0.185	0.206	0.053
SCAN-D4	0.125	0.057	0.094	0.274	0.097
BLYP-D3(OP)	0.065	0.079	-0.174	0.118	0.194
B3LYP-D4	0.095	0.224	-0.081	0.123	0.109
B97M-D3(BJ)	-0.162	-0.171	-0.400	-0.138	0.002
ωB97X-V	0.149	0.094	0.295	0.216	0.037
B97M-V	0.162	0.069	0.274	0.257	0.086
PBE-D3(BJ)	-0.091	-0.094	-0.293	0.014	-0.010
DSD-PBEP86	0.343	0.429	0.372	0.455	0.182
PBE-D4	-0.134	-0.034	-0.377	-0.085	-0.063
DSD-BLYP	0.369	0.459	0.420	0.468	0.198
BLYP-D3(BJ)	-0.042	0.082	-0.358	-0.117	0.149
BLYP-D4	0.084	0.283	-0.201	0.094	0.137
B3LYP-NL	0.289	0.230	0.464	0.365	0.155
TPSSH-D3(BJ)	-0.105	0.013	-0.459	-0.101	0.064
TPSSH-D4	-0.071	0.104	-0.424	-0.050	0.044
TPSS-D3(BJ)	-0.098	0.028	-0.448	-0.084	0.057
TPSS-D3(OP)	0.157	-0.001	-0.029	0.466	0.204
TPSSH-D3(OP)	0.194	0.015	0.032	0.521	0.226
TPSS-D4	-0.124	0.088	-0.509	-0.133	0.005
B3LYP-D3(zero)	0.166	0.024	0.370	0.202	0.100
BP-D4	0.037	0.310	-0.505	-0.074	0.305
BLYP-D3(zero)	0.139	-0.050	0.294	0.173	0.143
BP-D3(BJ)	-0.264	0.011	-0.928	-0.526	0.195

6 Results: DFT Across NCIA data sets

Figure S2: Error of DFT calculations involving a specific element across the D1200, HB375, HB300SPX, SH250 and R739 data sets. RMSE in kcal/mol averaged over the respective subsets of the datasets containing the element. The DFT functionals are listed in groups with decreasing computational complexity, from double-hybrid methods to pure GGA functionals. The data used for construction of this plot are available in Table S15.

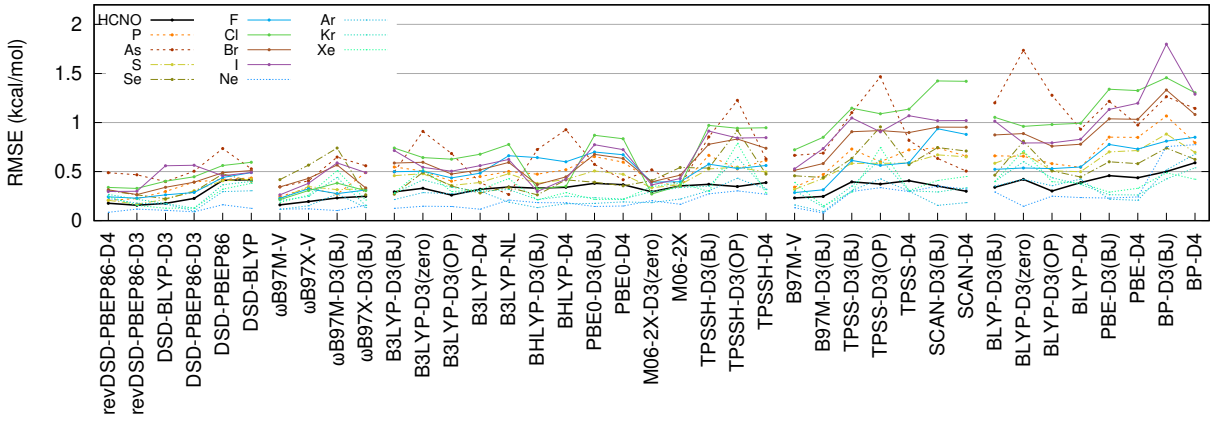

Table S13: Error (RMSE in kcal/mol) of DFT calculations in the NCIA data sets (equilibrium geometries, or closest-contact geometries in the R739 set). These results are plotted in Fig. 3 in the paper.

Functional	D1200	HB375	HB300SPX	SH250	R739
Double-hybrid functionals					
revDSD-PBEP86-D4	0.259	0.137	0.185	0.338	0.261
revDSD-PBEP86-D3	0.149	0.124	0.256	0.397	0.211
DSD-BLYP-D3	0.256	0.129	0.38	0.569	0.212
DSD-PBEP86-D3	0.29	0.241	0.492	0.607	0.212
DSD-PBEP86	0.512	0.367	0.377	0.57	0.346
DSD-BLYP	0.563	0.358	0.418	0.525	0.373
Range-separated hybrid functionals					
ωB97M-V	0.156	0.159	0.239	0.431	0.229
ωB97X-V	0.189	0.138	0.298	0.522	0.399
ωB97M-D3(BJ)	0.278	0.198	0.273	0.751	0.417
ωB97X-D3(BJ)	0.314	0.18	0.46	0.423	0.311
Hybrid functionals					
B3LYP-D3(BJ)	0.38	0.292	0.79	0.751	0.393
B3LYP-D3(zero)	0.245	0.38	0.597	0.679	0.61
B3LYP-D3(OP)	0.224	0.171	0.58	0.657	0.409
B3LYP-D4	0.358	0.252	0.575	0.677	0.39
B3LYP-NL	0.384	0.421	0.666	0.792	0.447
BHLYP-D3(BJ)	0.135	0.445	0.617	0.448	0.395
BHLYP-D4	0.327	0.425	0.544	0.568	0.36
PBE0-D3(BJ)	0.257	0.423	1.005	0.929	0.522
PBE0-D4	0.244	0.406	0.919	0.897	0.499
M06-2X-D3(zero)	0.283	0.308	0.378	0.467	0.315
M06-2X	0.327	0.338	0.379	0.495	0.366
TPSSH-D3(BJ)	0.364	0.312	0.897	1.191	0.528
TPSSH-D3(OP)	0.356	0.292	0.659	1.107	0.725
TPSSH-D4	0.38	0.309	0.811	1.158	0.487
Meta-GGA functionals					
TPSS-D3(BJ)	0.36	0.345	1.012	1.487	0.558
TPSS-D3(OP)	0.34	0.32	0.765	1.346	0.727
TPSS-D4	0.441	0.334	0.959	1.485	0.518
B97M-V	0.212	0.258	0.345	0.821	0.451
B97M-D3(BJ)	0.343	0.247	0.62	1.168	0.225
SCAN-D3(BJ)	0.3	0.474	1.179	1.669	0.582
SCAN-D4	0.335	0.376	1.097	1.614	0.581
GGA functionals					
BLYP-D3(BJ)	0.541	0.242	0.866	1.283	0.517
BLYP-D3(zero)	0.417	0.389	0.656	1.221	0.794
BLYP-D3(OP)	0.323	0.244	0.812	1.227	0.476
BLYP-D4	0.508	0.239	0.658	1.206	0.475
PBE-D3(BJ)	0.318	0.486	1.308	1.646	0.652
PBE-D4	0.364	0.482	1.27	1.661	0.647
BP-D3(BJ)	0.979	0.413	1.342	1.849	0.717
BP-D4	0.743	0.445	1.006	1.617	0.705
Hartree–Fock					
HF-D3(BJ)	0.265	0.632	0.685	1.342	0.693

Table S14: Average and maximum error (RMSE in kcal/mol) of DFT methods in the D1200, HB375, HB300SPX, SH250 and R739 data sets. The table is sorted by the average error.

Functional	Avg. err.	Max. err.
revDSD-PBEP86-D3	0.227	0.397
revDSD-PBEP86-D4	0.236	0.338
ωB97M-V	0.243	0.431
DSD-BLYP-D3	0.309	0.569
ωB97X-V	0.309	0.522
ωB97X-D3(BJ)	0.338	0.46
M06-2X-D3(zero)	0.350	0.467
DSD-PBEP86-D3	0.368	0.607
M06-2X	0.381	0.495
ωB97M-D3(BJ)	0.383	0.751
BHLYP-D3(BJ)	0.408	0.617
B3LYP-D3(OP)	0.408	0.657
B97M-V	0.417	0.821
DSD-PBEP86	0.434	0.57
BHLYP-D4	0.445	0.568
DSD-BLYP	0.447	0.563
B3LYP-D4	0.450	0.677
B3LYP-D3(zero)	0.502	0.679
B97M-D3(BJ)	0.521	1.168
B3LYP-D3(BJ)	0.521	0.79
B3LYP-NL	0.542	0.792
PBE0-D4	0.593	0.919
BLYP-D3(OP)	0.616	1.227
BLYP-D4	0.617	1.206
PBE0-D3(BJ)	0.627	1.005
TPSSH-D3(OP)	0.628	1.107
TPSSH-D4	0.629	1.158
TPSSH-D3(BJ)	0.658	1.191
BLYP-D3(BJ)	0.690	1.283
BLYP-D3(zero)	0.695	1.221
TPSS-D3(OP)	0.700	1.346
HF-D3(BJ)	0.723	1.342
TPSS-D4	0.747	1.485
TPSS-D3(BJ)	0.752	1.487
SCAN-D4	0.801	1.614
SCAN-D3(BJ)	0.841	1.669
PBE-D3(BJ)	0.882	1.646
PBE-D4	0.885	1.661
BP-D4	0.903	1.617
BP-D3(BJ)	1.060	1.849

Table S15: Error of DFT calculations involving a specific element across the D1200, HB375, HB300SPX, SH250 and R739 data sets. RMSE in kcal/mol averaged over the respective subsets of the datasets containing the element. These results are plotted in Fig. S2.

Functional	HCNO	S	P	F	Cl	Br	I	As	Se	Ne	Ar	Kr	Xe
Double-hybrid functionals													
revDSD-PBEP86-D4	0.177	0.207	0.207	0.242	0.338	0.314	0.297	0.488	0.306	0.085	0.226	0.268	0.247
revDSD-PBEP86-D3	0.156	0.168	0.225	0.228	0.328	0.265	0.299	0.467	0.263	0.118	0.177	0.217	0.155
DSD-BLYP-D3	0.176	0.216	0.297	0.262	0.401	0.340	0.559	0.398	0.224	0.105	0.169	0.175	0.126
DSD-PBEP86-D3	0.226	0.305	0.393	0.287	0.447	0.391	0.564	0.504	0.293	0.090	0.097	0.125	0.129
DSD-PBEP86	0.410	0.407	0.420	0.439	0.562	0.488	0.457	0.734	0.476	0.162	0.295	0.365	0.320
DSD-BLYP	0.417	0.421	0.434	0.490	0.594	0.510	0.491	0.528	0.401	0.125	0.304	0.400	0.380
Range-separated hybrid functionals													
ωB97M-V	0.160	0.224	0.245	0.227	0.213	0.343	0.263	0.345	0.418	0.118	0.113	0.205	0.185
ωB97X-V	0.194	0.347	0.332	0.325	0.307	0.431	0.380	0.405	0.566	0.120	0.151	0.246	0.259
ωB97M-D3(BJ)	0.231	0.316	0.269	0.275	0.382	0.567	0.588	0.647	0.740	0.102	0.273	0.441	0.510
ωB97X-D3(BJ)	0.247	0.325	0.490	0.312	0.305	0.333	0.490	0.559	0.262	0.157	0.133	0.153	0.210
Hybrid functionals													
B3LYP-D3(BJ)	0.293	0.458	0.547	0.499	0.740	0.587	0.714	0.497	0.269	0.125	0.215	0.278	0.281
B3LYP-D3(zero)	0.329	0.479	0.506	0.504	0.642	0.594	0.543	0.909	0.492	0.146	0.285	0.415	0.525
B3LYP-D3(OP)	0.260	0.358	0.401	0.435	0.626	0.474	0.504	0.684	0.352	0.144	0.275	0.340	0.295
B3LYP-D4	0.319	0.389	0.449	0.483	0.676	0.517	0.561	0.386	0.282	0.116	0.300	0.299	0.316
B3LYP-NL	0.342	0.478	0.501	0.661	0.776	0.593	0.623	0.266	0.339	0.210	0.188	0.317	0.428
BHLYP-D3(BJ)	0.331	0.379	0.473	0.642	0.324	0.367	0.297	0.725	0.263	0.183	0.132	0.214	0.211
BHLYP-D4	0.340	0.420	0.519	0.599	0.349	0.446	0.423	0.927	0.415	0.182	0.172	0.249	0.285
PBE0-D3(BJ)	0.380	0.506	0.653	0.699	0.870	0.675	0.774	0.573	0.391	0.144	0.176	0.234	0.216
PBE0-D4	0.366	0.472	0.596	0.671	0.834	0.634	0.723	0.417	0.355	0.149	0.190	0.220	0.214
M06-2X-D3(zero)	0.294	0.335	0.402	0.392	0.271	0.395	0.364	0.518	0.408	0.203	0.183	0.295	0.347
M06-2X	0.353	0.366	0.349	0.396	0.356	0.463	0.429	0.354	0.542	0.167	0.220	0.337	0.384
TPSSH-D3(BJ)	0.370	0.526	0.665	0.575	0.970	0.777	0.912	0.853	0.535	0.273	0.304	0.351	0.296
TPSSH-D3(OP)	0.347	0.545	0.527	0.532	0.941	0.832	0.842	1.226	0.919	0.303	0.432	0.644	0.788
TPSSH-D4	0.388	0.495	0.614	0.562	0.947	0.737	0.846	0.629	0.474	0.269	0.320	0.316	0.282
Meta-GGA functionals													
B97M-V	0.231	0.318	0.341	0.285	0.722	0.512	0.527	0.665	0.457	0.131	0.159	0.307	0.390
B97M-D3(BJ)	0.246	0.430	0.469	0.314	0.850	0.582	0.735	0.688	0.439	0.078	0.094	0.138	0.150
TPSS-D3(BJ)	0.395	0.583	0.730	0.613	1.146	0.906	1.046	1.100	0.634	0.297	0.305	0.351	0.292
TPSS-D3(OP)	0.372	0.584	0.599	0.563	1.089	0.918	0.905	1.467	0.956	0.336	0.425	0.626	0.743
TPSS-D4	0.407	0.586	0.721	0.589	1.136	0.897	1.069	0.818	0.571	0.295	0.300	0.304	0.306
SCAN-D3(BJ)	0.352	0.670	0.741	0.937	1.424	0.953	1.018	0.634	0.745	0.352	0.155	0.292	0.409
SCAN-D4	0.299	0.650	0.658	0.879	1.420	0.952	1.020	0.505	0.709	0.322	0.183	0.334	0.449
GGA functionals													
BLYP-D3(BJ)	0.340	0.585	0.660	0.523	1.053	0.874	1.013	1.200	0.463	0.291	0.352	0.399	0.398
BLYP-D3(zero)	0.420	0.691	0.656	0.536	0.961	0.887	0.792	1.735	0.809	0.147	0.422	0.599	0.713
BLYP-D3(OP)	0.302	0.507	0.580	0.530	0.980	0.758	0.793	1.277	0.508	0.249	0.356	0.437	0.396
BLYP-D4	0.388	0.499	0.544	0.546	0.992	0.779	0.829	0.931	0.444	0.235	0.423	0.367	0.384
PBE-D3(BJ)	0.458	0.701	0.851	0.776	1.339	1.038	1.134	1.216	0.600	0.230	0.218	0.264	0.290
PBE-D4	0.437	0.715	0.848	0.730	1.325	1.033	1.197	0.974	0.581	0.241	0.207	0.261	0.329
BP-D3(BJ)	0.500	0.882	1.067	0.810	1.457	1.332	1.798	1.263	0.740	0.755	0.516	0.425	0.487
BP-D4	0.590	0.696	0.794	0.849	1.304	1.082	1.289	1.144	0.624	0.774	0.678	0.538	0.421

Table S16: Relative error of DFT calculations involving a specific element across the D1200, HB375, HB300SPX, SH250 and R739 data sets. A relative error (RMSE divided by the magnitude of interaction energies) averaged over the respective subsets of the datasets containing the element. These data are plotted in Fig. 4 in the paper.

Functional	HCNO	S	P	F	Cl	Br	I	As	Se	Ne	Ar	Kr	Xe
Double-hybrid functionals													
revDSD-PBEP86-D4	6.5 %	5.9 %	5.5 %	10.2 %	11.7 %	10.8 %	9.6 %	4.8 %	5.0 %	7.9 %	18.5 %	18.4 %	15.2 %
revDSD-PBEP86-D3	5.8 %	4.7 %	6.0 %	8.9 %	10.5 %	7.9 %	8.5 %	4.6 %	4.3 %	13.8 %	14.3 %	15.1 %	9.7 %
DSD-BLYP-D3	6.6 %	6.1 %	8.4 %	9.4 %	12.0 %	8.9 %	16.1 %	3.9 %	3.7 %	15.5 %	14.0 %	12.1 %	8.0 %
DSD-PBEP86-D3	7.8 %	8.3 %	10.5 %	9.9 %	13.0 %	9.9 %	16.2 %	4.9 %	4.8 %	10.4 %	7.7 %	8.7 %	8.2 %
DSD-PBEP86	14.5 %	11.4 %	11.2 %	18.3 %	19.0 %	15.6 %	13.9 %	7.2 %	7.8 %	27.5 %	26.8 %	27.6 %	21.4 %
DSD-BLYP	14.9 %	12.0 %	12.3 %	19.4 %	20.0 %	16.5 %	15.1 %	5.2 %	6.6 %	24.0 %	28.3 %	30.7 %	25.6 %
Range-separated hybrid functionals													
ωB97M-V	5.6 %	6.5 %	5.9 %	8.1 %	7.2 %	9.4 %	8.0 %	3.4 %	6.8 %	21.3 %	8.7 %	13.8 %	11.3 %
ωB97X-V	7.5 %	10.7 %	9.1 %	11.2 %	10.8 %	12.8 %	12.0 %	4.0 %	9.2 %	29.0 %	9.8 %	15.3 %	15.0 %
ωB97M-D3(BJ)	8.4 %	9.3 %	7.0 %	10.6 %	13.3 %	16.7 %	18.0 %	6.3 %	12.1 %	10.6 %	20.4 %	30.3 %	31.6 %
ωB97X-D3(BJ)	9.3 %	9.3 %	13.1 %	11.1 %	9.6 %	9.6 %	15.2 %	5.5 %	4.3 %	17.6 %	9.0 %	10.7 %	13.1 %
Hybrid functionals													
B3LYP-D3(BJ)	10.5 %	12.7 %	15.4 %	16.4 %	22.0 %	16.1 %	20.8 %	4.9 %	4.4 %	13.7 %	14.4 %	17.4 %	17.0 %
B3LYP-D3(zero)	11.2 %	14.6 %	13.3 %	15.8 %	20.6 %	18.2 %	18.2 %	8.9 %	8.0 %	21.2 %	18.4 %	25.6 %	28.8 %
B3LYP-D3(OP)	10.1 %	10.1 %	11.1 %	15.5 %	19.2 %	13.3 %	14.8 %	6.7 %	5.7 %	14.8 %	19.9 %	22.1 %	17.4 %
B3LYP-D4	11.8 %	10.9 %	13.0 %	17.4 %	20.7 %	15.1 %	16.5 %	3.8 %	4.6 %	10.2 %	22.7 %	19.1 %	19.3 %
B3LYP-NL	11.4 %	13.3 %	14.3 %	20.7 %	23.6 %	17.1 %	18.8 %	2.6 %	5.5 %	20.5 %	16.1 %	23.2 %	28.0 %
BHLYP-D3(BJ)	10.9 %	10.3 %	11.2 %	19.5 %	10.0 %	10.4 %	9.2 %	7.1 %	4.3 %	15.3 %	9.1 %	13.4 %	12.0 %
BHLYP-D4	11.4 %	11.1 %	12.4 %	19.1 %	11.4 %	13.4 %	12.8 %	9.1 %	6.8 %	17.5 %	12.2 %	17.4 %	19.5 %
PBE0-D3(BJ)	13.2 %	13.8 %	17.9 %	23.7 %	25.2 %	17.0 %	22.3 %	5.6 %	6.4 %	20.6 %	11.4 %	14.8 %	12.5 %
PBE0-D4	12.7 %	13.0 %	16.7 %	23.1 %	24.2 %	16.0 %	20.7 %	4.1 %	5.8 %	22.5 %	12.7 %	14.0 %	12.6 %
M06-2X-D3(zero)	10.2 %	9.2 %	9.8 %	13.6 %	9.2 %	11.7 %	11.4 %	5.1 %	6.7 %	31.8 %	18.0 %	22.9 %	23.8 %
M06-2X	12.5 %	10.6 %	9.1 %	14.3 %	12.3 %	14.0 %	13.4 %	3.5 %	8.9 %	22.1 %	22.0 %	26.4 %	26.4 %
TPSSH-D3(BJ)	13.8 %	14.5 %	17.9 %	20.8 %	28.8 %	19.8 %	26.0 %	8.3 %	8.7 %	24.3 %	19.0 %	22.0 %	17.2 %
TPSSH-D3(OP)	12.9 %	15.8 %	13.6 %	19.5 %	30.0 %	25.3 %	26.4 %	12.0 %	15.0 %	33.5 %	27.3 %	42.2 %	48.1 %
TPSSH-D4	14.4 %	13.7 %	16.9 %	20.8 %	28.1 %	18.7 %	23.8 %	6.1 %	7.7 %	27.7 %	21.6 %	19.9 %	16.9 %
Meta-GGA functionals													
B97M-V	7.9 %	9.4 %	8.6 %	9.4 %	22.5 %	15.7 %	17.2 %	6.5 %	7.5 %	13.9 %	11.9 %	20.1 %	22.6 %
B97M-D3(BJ)	8.6 %	11.4 %	12.0 %	11.9 %	24.3 %	13.8 %	20.1 %	6.7 %	7.2 %	11.0 %	8.0 %	9.6 %	9.5 %
TPSS-D3(BJ)	14.5 %	15.8 %	19.3 %	22.2 %	33.7 %	22.2 %	29.4 %	10.7 %	10.4 %	28.7 %	19.1 %	21.9 %	16.9 %
TPSS-D3(OP)	13.8 %	16.4 %	15.3 %	20.7 %	33.9 %	26.4 %	27.7 %	14.3 %	15.6 %	39.1 %	26.2 %	40.4 %	44.8 %
TPSS-D4	15.1 %	16.1 %	19.7 %	21.7 %	33.3 %	21.9 %	30.0 %	8.0 %	9.3 %	33.4 %	19.8 %	19.6 %	19.3 %
SCAN-D3(BJ)	11.3 %	17.4 %	18.4 %	30.9 %	40.5 %	23.5 %	28.8 %	6.2 %	12.2 %	46.0 %	10.5 %	20.0 %	25.9 %
SCAN-D4	9.8 %	17.2 %	16.9 %	28.6 %	41.0 %	24.3 %	29.2 %	4.9 %	11.6 %	40.1 %	13.6 %	23.5 %	28.6 %
GGA functionals													
BLYP-D3(BJ)	13.0 %	16.1 %	18.5 %	20.0 %	31.9 %	23.3 %	29.1 %	11.7 %	7.6 %	32.9 %	23.5 %	25.0 %	24.4 %
BLYP-D3(zero)	15.0 %	20.5 %	16.2 %	19.8 %	31.1 %	26.2 %	26.0 %	17.0 %	13.2 %	18.1 %	25.6 %	37.5 %	39.8 %
BLYP-D3(OP)	11.3 %	13.5 %	15.2 %	20.0 %	29.1 %	19.4 %	22.8 %	12.5 %	8.3 %	30.2 %	27.0 %	29.0 %	23.1 %
BLYP-D4	14.9 %	13.9 %	15.8 %	22.3 %	30.2 %	21.0 %	23.7 %	9.1 %	7.3 %	22.2 %	32.5 %	23.4 %	24.0 %
PBE-D3(BJ)	16.1 %	19.0 %	23.1 %	26.8 %	38.7 %	24.9 %	32.0 %	11.9 %	9.8 %	40.8 %	15.5 %	17.4 %	17.3 %
PBE-D4	15.2 %	19.7 %	23.7 %	25.3 %	38.2 %	24.7 %	33.8 %	9.5 %	9.5 %	44.7 %	14.5 %	18.4 %	20.8 %
BP-D3(BJ)	18.5 %	24.5 %	29.8 %	32.6 %	43.9 %	35.2 %	52.3 %	12.3 %	12.1 %	110.1 %	40.9 %	29.5 %	33.3 %
BP-D4	22.0 %	19.6 %	22.7 %	35.4 %	39.8 %	28.4 %	36.8 %	11.2 %	10.2 %	111.5 %	55.7 %	37.2 %	26.7 %

7 Results: DFT in Smaller Subsets of NCIA data sets

Table S17: Error (RMSE in kcal/mol) of DFT calculations in reduced-size subsets of the D1200, HB375, HB300SPX, SH250 and R739 data sets. Subsets with 50 entries obtained using a clustering analysis described in the paper.

Functional	HB375	HB300SPX	D1200	SH250	REP739	Avg.
revDSD-PBEP86-D4	0.149	0.180	0.205	0.381	0.261	0.235
revDSD-PBEP86-D3	0.151	0.210	0.144	0.490	0.227	0.244
DSD-BLYP-D3	0.138	0.319	0.207	0.716	0.223	0.321
DSD-PBEP86-D3	0.258	0.421	0.218	0.739	0.226	0.372
DSD-PBEP86	0.268	0.362	0.431	0.639	0.326	0.405
DSD-BLYP	0.249	0.398	0.476	0.627	0.352	0.420
wB97M-V	0.172	0.222	0.117	0.425	0.198	0.227
wB97X-V	0.148	0.246	0.192	0.513	0.340	0.288
wB97M-D3(BJ)	0.198	0.272	0.251	0.752	0.416	0.378
wB97X-D3(BJ)	0.160	0.369	0.258	0.467	0.352	0.321
B3LYP-D3(BJ)	0.337	0.660	0.299	0.978	0.468	0.548
B3LYP-D3(zero)	0.424	0.525	0.233	0.725	0.627	0.507
B3LYP-D3(OP)	0.194	0.461	0.195	0.813	0.470	0.427
B3LYP-D4	0.308	0.502	0.293	0.873	0.447	0.485
B3LYP-NL	0.551	0.639	0.352	1.066	0.550	0.632
BHLYP-D3(BJ)	0.581	0.573	0.102	0.435	0.445	0.427
BHLYP-D4	0.567	0.521	0.259	0.506	0.393	0.449
PBE0-D3(BJ)	0.526	0.831	0.223	1.192	0.606	0.676
PBE0-D4	0.519	0.776	0.195	1.163	0.582	0.647
M06-2X-D3(zero)	0.336	0.377	0.199	0.488	0.273	0.335
M06-2X	0.341	0.375	0.283	0.513	0.301	0.363
TPSSH-D3(BJ)	0.382	0.752	0.296	1.488	0.623	0.708
TPSSH-D3(OP)	0.368	0.553	0.339	1.273	0.810	0.669
TPSSH-D4	0.380	0.718	0.298	1.469	0.583	0.690
TPSS-D3(BJ)	0.426	0.904	0.297	1.858	0.674	0.832
TPSS-D3(OP)	0.410	0.681	0.334	1.592	0.838	0.771
TPSS-D4	0.414	0.903	0.348	1.883	0.626	0.835
B97M-V	0.238	0.519	0.179	1.048	0.416	0.480
B97M-D3(BJ)	0.240	0.726	0.296	1.530	0.267	0.612
SCAN-D3(BJ)	0.607	1.177	0.202	2.247	0.712	0.989
SCAN-D4	0.513	1.113	0.248	2.189	0.702	0.953
BLYP-D3(BJ)	0.281	0.758	0.441	1.595	0.608	0.737
BLYP-D3(zero)	0.361	0.624	0.356	1.309	0.803	0.691
BLYP-D3(OP)	0.289	0.731	0.307	1.522	0.590	0.688
BLYP-D4	0.240	0.656	0.436	1.530	0.550	0.682
PBE-D3(BJ)	0.596	1.183	0.311	2.099	0.747	0.987
PBE-D4	0.594	1.181	0.328	2.136	0.724	0.993
BP-D3(BJ)	0.480	1.153	0.806	2.293	0.862	1.119
BP-D4	0.475	0.900	0.659	2.032	0.867	0.987

Table S18: Error (RMSE in kcal/mol) of DFT calculations in reduced-size subsets of the D1200, HB375, HB300SPX, SH250 and R739 data sets. Subsets with 20 entries obtained using a clustering analysis described in the paper.

Functional	HB375	HB300SPX	D1200	SH250	REP739	Avg.
revDSD-PBEP86-D4	0.101	0.162	0.204	0.465	0.267	0.240
revDSD-PBEP86-D3	0.099	0.173	0.139	0.589	0.203	0.241
DSD-BLYP-D3	0.101	0.285	0.157	0.811	0.232	0.317
DSD-PBEP86-D3	0.230	0.371	0.217	0.851	0.277	0.389
DSD-PBEP86	0.215	0.331	0.438	0.748	0.348	0.416
DSD-BLYP	0.216	0.372	0.478	0.716	0.417	0.440
wB97M-V	0.138	0.215	0.095	0.456	0.211	0.223
wB97X-V	0.110	0.211	0.135	0.545	0.350	0.270
wB97M-D3(BJ)	0.175	0.263	0.252	0.821	0.408	0.384
wB97X-D3(BJ)	0.134	0.359	0.271	0.506	0.352	0.324
B3LYP-D3(BJ)	0.328	0.674	0.276	0.933	0.467	0.536
B3LYP-D3(zero)	0.436	0.489	0.107	0.717	0.672	0.484
B3LYP-D3(OP)	0.166	0.447	0.210	0.783	0.364	0.394
B3LYP-D4	0.277	0.497	0.240	0.801	0.421	0.447
B3LYP-NL	0.531	0.596	0.360	1.032	0.638	0.631
BHLYP-D3(BJ)	0.535	0.575	0.082	0.432	0.501	0.425
BHLYP-D4	0.489	0.548	0.211	0.558	0.421	0.445
PBE0-D3(BJ)	0.512	0.720	0.265	1.208	0.760	0.693
PBE0-D4	0.497	0.657	0.230	1.159	0.721	0.653
M06-2X-D3(zero)	0.256	0.366	0.194	0.455	0.313	0.317
M06-2X	0.235	0.370	0.238	0.487	0.326	0.331
TPSSH-D3(BJ)	0.329	0.697	0.352	1.534	0.533	0.689
TPSSH-D3(OP)	0.326	0.480	0.318	1.293	0.726	0.629
TPSSH-D4	0.344	0.644	0.322	1.492	0.498	0.660
TPSS-D3(BJ)	0.349	0.860	0.360	1.925	0.578	0.814
TPSS-D3(OP)	0.343	0.629	0.321	1.634	0.738	0.733
TPSS-D4	0.370	0.849	0.364	1.926	0.539	0.810
B97M-V	0.185	0.542	0.136	0.999	0.473	0.467
B97M-D3(BJ)	0.202	0.770	0.304	1.509	0.306	0.618
SCAN-D3(BJ)	0.586	1.000	0.146	2.413	0.935	1.016
SCAN-D4	0.484	0.937	0.211	2.343	0.916	0.978
BLYP-D3(BJ)	0.179	0.815	0.386	1.601	0.515	0.699
BLYP-D3(zero)	0.350	0.586	0.202	1.354	0.847	0.668
BLYP-D3(OP)	0.201	0.759	0.308	1.539	0.519	0.665
BLYP-D4	0.179	0.673	0.354	1.492	0.492	0.638
PBE-D3(BJ)	0.590	1.099	0.288	2.142	0.875	0.999
PBE-D4	0.591	1.102	0.297	2.159	0.841	0.998
BP-D3(BJ)	0.368	1.189	0.774	2.396	0.788	1.103
BP-D4	0.384	0.887	0.631	2.093	0.733	0.946

8 Results: DFT-D3 Damping Function Parametrization

Table S19: Errors of B3LYP-D3 with different versions of the D3 correction in the D442×10 0and HB300SPX×10 data sets, and in the XH–Br and XH–I groups of the latter.

		RMSE, kcal/mol			
Version	Fitted to	D442×10	HB300SPX×10	XH–Br	XH–I
D3(BJ)	original parameters	0.37	0.73	0.40	0.68
	D442×10	0.29	0.83	0.50	0.75
	D442×10 + H-bonds	0.32	0.71	0.39	0.65
D3(BJ')	D442×10	0.26	0.90	0.65	0.90
	D442×10 + H-bonds	0.31	0.76	0.52	0.81
D3(OP)	original parameters	0.39	0.53	0.27	0.46
	D442×10	0.29	0.72	0.40	0.64
	D442×10 + H-bonds	0.31	0.57	0.28	0.47
D3(BJ, scaled C_6)	D442×10	0.19	0.31	0.60	0.78
D3(BJ, scaled R_0)	D442×10	0.19	0.31	0.61	0.76

Table S20: D3 damping parameters, original and optimized in this work. Value of $a2$ in a.u., the remaining parameters are dimensionless.

Version	Fitted to	a_1	a_2	s_8	β
D3(BJ)	original parameters	0.3981	4.4211	1.9889	—
	D442×10	0.3910	3.8820	1.1220	—
	D442×10 + H-bonds	0.3792	4.2430	1.4151	—
D3(BJ')	D442×10	0.7790	1.8940	1.0960	—
	D442×10 + H-bonds	0.6997	2.5659	1.3117	—
D3(OP)	original parameters	0.3000	4.2500	0.7831	10.0000
	D442×10	0.3483	3.8177	0.7631	8.0564
	D442×10 + H-bonds	0.3025	3.9302	0.5050	12.3419
D3(BJ, scaled C_6)	D442×10	0.3870	3.8380	1.1240	—
D3(BJ, scaled R_0)	D442×10	0.4150	3.5290	1.0840	—

Table S21: Elementwise scaling factors applied to either dispersion coefficients or damping radii in the D3(BJ) dispersion correction, obtained by parametrization on the D442×10 data set. The corresponding global parameters are listed in Table S20.

Element	$s(C_6)$	$s(R_0)$
H	0.854	1.302
B	1.051	1.034
C	1.058	1.000
N	0.956	1.116
O	1.211	0.869
P	0.820	1.175
S	0.924	1.090
F	1.287	0.840
Cl	1.083	0.967
Br	1.067	0.954
I	0.956	1.015
He	1.049	0.995
Ne	1.054	0.993
Ar	1.203	0.877
Kr	1.259	0.825
Xe	1.221	0.853

9 Results: SQM methods

Table S22: Error of semiempirical QM methods in a subset of the D1200 data sets to which they are applicable. RMSE in kcal/mol.

Method	All	HCNO	PS	Halogens
PM6	2.457	1.814	2.880	2.508
PM6-D3	1.396	0.446	1.495	1.784
PM6-D3H4	1.397	0.457	1.495	1.784
PM6-D3H4X	1.400	0.458	1.490	1.793
PM7	2.379	0.646	2.702	2.937
DFTB3	2.412	2.458	2.870	1.837
DFTB3-D3	1.137	0.509	0.697	1.712
DFTB3-D3H4	1.194	0.526	0.732	1.800
DFTB3-D3H5	1.177	0.497	0.712	1.783
DFTB3-D3H5X	1.326	0.497	0.712	2.055
GFN2-XTB	0.762	0.491	0.901	0.810

Table S23: Systematic error of semiempirical QM methods in a subset of the D1200 data sets to which they are applicable. MSE in kcal/mol.

Method	All	HCNO	PS	Halogens
PM6	1.900	1.607	2.474	1.617
PM6-D3	0.349	0.036	0.804	0.195
PM6-D3H4	0.346	0.026	0.804	0.195
PM6-D3H4X	0.374	0.013	0.791	0.294
PM7	-0.447	0.053	0.076	-1.363
DFTB3	2.015	2.218	2.689	1.212
DFTB3-D3	-0.260	0.136	0.045	-0.882
DFTB3-D3H4	-0.254	0.158	0.214	-1.042
DFTB3-D3H5	-0.258	0.169	0.191	-1.042
DFTB3-D3H5X	-0.049	0.169	0.191	-0.459
GFN2-XTB	-0.055	0.292	-0.258	-0.164