Supporting Information: Assessing the accuracy of across-the-scale methods for predicting carbohydrate conformational energies

Mateusz Marianski,* Adriana Supady,* Teresa Ingram,* Markus Schneider,* and Carsten Baldauf*

Fritz-Haber-Institut der Max-Planck-Gesellschaft, Faradayweg 4-6, D-14195 Berlin, Germany

E-mail: marianski@fhi-berlin.mpg.de; supady@fhi-berlin.mpg.de; ingram@fhi-berlin.mpg.de; markus.schneider@fhi-berlin.mpg.de; baldauf@fhi-berlin.mpg.de

*To whom correspondence should be addressed

Settings for the conformational searches

The genetic algorithm (GA) runs were performed with the PBE functional and MBD dispersion correction via Fafoom's interface to FHI-aims. The settings used for 18 GA runs for α- and β-glucose are summarized in Table 1. Calculations yielded 1,210 structures for each isomer. In order to increase the sampling of less favourable ring puckers, *i.e.* skew boats and boats, in runs 11 to 18 chair, half-chair and envelope puckers were removed from the list of available types. For the open-chain glucose we performed 15 GA runs which yielded 986 structures. All GA settings remained unchanged.

Table 1: GA parameters for α- and β-glucose.

	Parameter	Value
Molecule	distance_cutoff_1	1.2 Å
	distance_cutoff_2	2.15 Å
	rmsd_cutoff_uniq	0.25 Å
	chiral	True
Run settings	max_iter	30 (runs 1-17) 80 (run 18)
	iter_limit_conv	20 (runs 1-17) 70 (run 18)
	energy_diff_conv	0.001 eV
GA settings	popsize	10
	energy_var	0.001 eV
	selection	roulette wheel
	fitness_sum_limit	1.2
	prob_for_crossing	0.95
	prob_for_mut_pyranosering	0.6
	prob_for_mut_torsion	0.8
	max_mutations_pyranosering	1
	max_mutations_torsions	3

For α-maltose 40 GA runs resulted in 3,750 structures. The GA parameters are identical to these used for α- and β-glucose. Few exceptions (increased iteration limits and number of allowed mutation per generation) are listed in Table 2.

Table 2: Selected GA parameters for α-maltose.

	Parameter	Value
	max_iter	30 (runs 1-30) 80 (runs 31-40)
	iter_limit_conv	20 (runs 1-30) 70 (runs 31-40)
	max_mutations_torsions	5

From the resulting pools of structures duplicates were removed based on the geometrical similarity (RMSD cut-off=0.05 Å) of the conformers. This procedure reduced the number of the conformers to 429 unique conformers for α-glucose, 479 for β-glucose and 570 for the open-chain glucose. A similar procedure yielded 2,092 unique α-maltose conformations.

Mean absolute error and maximal error

Table 3: Mean absolute errors and maximum errors (in kcal/mol) against DLPNO-CCSD(T)/CBS(3,4) relative energies for 205 conformations of α, β and open-chain glucose.

	Mean Absolute Error (MAE)				Maximum Error (ME)			
Energy function		+MBD	+vdW^{TS}	+D3		+MBD	+vdW^{TS}	+D3
CHARMM36	2.08	-	-	-	5.88	-	-	-
GLYCAM06	2.58	-	-	-	8.14	-	-	-
PM3	1.91	-	-	-	6.02	-	-	-
PM6	4.20	-	4.80[a]	4.70	12.91	-	13.74[a]	13.17
PM7	2.93	-	-	-	10.38	-	-	-
AM1	2.93	-	-	-	15.10	-	-	-
DFTB3/3OB	2.01	-	-	1.98	7.01	-	-	7.17
PBE	0.99	0.95	1.08	1.01	3.94	3.57	4.27	3.90
BLYP	2.19	-	1.20	2.03	8.58	-	4.68	8.00
M06L	1.64	-	-	1.64	6.85	-	-	6.85
M11L	1.15	-	-	-	3.57	-	-	-
SCAN	0.44	-	-	-	2.01	-	-	-
PBE0	0.44	0.65	0.60	0.55	1.71	2.35	2.26	2.10
B3LYP	1.31	-	1.25	1.06	4.59	-	4.89	4.04
M06	0.69	-	-	0.67	2.51	-	-	2.51
M06-2X	0.46	-	-	0.46	1.68	-	-	1.64
M06-HF	1.55	-	-	1.55	6.25	-	-	6.23
M08-SO	0.55	-	-	-	1.91	-	-	-
M08-HX	0.42	-	-	-	1.89	-	-	-
M11	0.65	-	-	-	2.28	-	-	-
XYG3	0.65	-	-	-	2.26	-	-	-
HF	1.74	-	-	-	6.60	-	-	-

a) PM6 augmented with D3H4 correction

Table 4: Mean absolute errors and maximum errors (in kcal/mol) against DLPNO-CCSD(T)/CBS(3,4) relative energies for 156 conformations of α and β-glucose.

Energy function	Mean Absolute Error (MAE)				Maximum Error (ME)			
		+MBD	+vdW^{TS}	+D3		+MBD	+vdW^{TS}	+D3
CHARMM36	2.08	-	-	-	5.88	-	-	-
GLYCAM06	2.58	-	-	-	8.14	-	-	-
PM3	2.01	-	-	-	5.58	-	-	-
PM6	3.11	-	4.06[a]	4.15	12.15	-	14.41[a]	15.04
PM7	3.02	-	-	-	10.88	-	-	-
AM1	2.31	-	-	-	7.73	-	-	-
DFTB3/3OB	1.48	-	-	1.64	4.52	-	-	5.56
PBE	0.51	0.71	0.78	0.69	1.41	2.26	2.08	2.19
BLYP	0.55	-	0.95	0.67	1.73	-	3.80	2.05
M06L	0.46	-	-	0.46	1.43	-	-	1.36
M11L	0.76	-	-	-	2.61	-	-	-
SCAN	0.39	-	-	-	2.01	-	-	-
PBE0	0.37	0.39	0.44	0.37	1.36	1.27	1.27	1.15
B3LYP	0.55	-	0.60	0.44	1.50	-	1.84	1.34
M06	0.28	-	-	0.28	1.04	-	-	1.04
M06-2X	0.28	-	-	0.28	0.90	-	-	0.90
M06-HF	0.32	-	-	0.32	1.13	-	-	1.13
M08-SO	0.51	-	-	-	1.43	-	-	-
M08-HX	0.42	-	-	-	1.80	-	-	-
M11	0.51	-	-	-	1.52	-	-	-
XYG3	0.25	-	-	-	0.71	-	-	-
HF	1.84	-	-	-	5.99	-	-	-

a) PM6 augmented with D3H4 correction

Table 5: Mean absolute errors and maximum errors (in kcal/mol) against DLPNO-CCSD(T)/CBS(3,4) relative energies for 223 conformations of α-maltose.

Energy function	Mean Absolute Error (MAE)	+MBD	+vdW^{TS}	+D3	Maximum Error (ME)	+MBD	+vdW^{TS}	+D3
CHARMM36	2.84	-	-	-	11.55	-	-	-
GLYCAM06	2.77	-	-	-	15.61	-	-	-
PM3	2.26	-	-	-	9.66	-	-	-
PM6	3.30	-	4.29[a]	4.54	16.70	-	20.52[a]	23.73
PM7	3.51	-	-	-	17.04	-	-	-
AM1	2.31	-	-	-	11.18	-	-	-
DFTB3/3OB	1.78	-	-	2.05	6.11	-	-	6.30
PBE	0.71	0.81	0.90	0.74	3.25	3.37	3.74	2.91
BLYP	0.92	-	1.18	0.83	4.10	-	4.31	3.18
M06L	0.58	-	-	0.60	2.21	-	-	2.33
M11L	0.85	-	-	-	3.14	-	-	-
SCAN	0.46	-	-	-	1.61	-	-	-
PBE0	0.62	0.51	0.60	0.46	2.70	1.75	1.98	1.43
B3LYP	0.88	-	0.71	0.55	3.18	-	3.04	1.94
M06	0.46	-	-	0.48	2.05	-	-	1.94
M06-2X	0.39	-	-	0.39	1.59	-	-	1.61
M06-HF	0.48	-	-	0.48	1.55	-	-	1.52
M08-SO	0.55	-	-	-	2.10	-	-	-
M08-HX	0.51	-	-	-	2.21	-	-	-
M11	0.60	-	-	-	2.10	-	-	-
XYG3	0.37	-	-	-	1.18	-	-	-
HF	1.99	-	-	-	9.09	-	-	-

a) PM6 augmented with D3H4 correction

Correlation plots

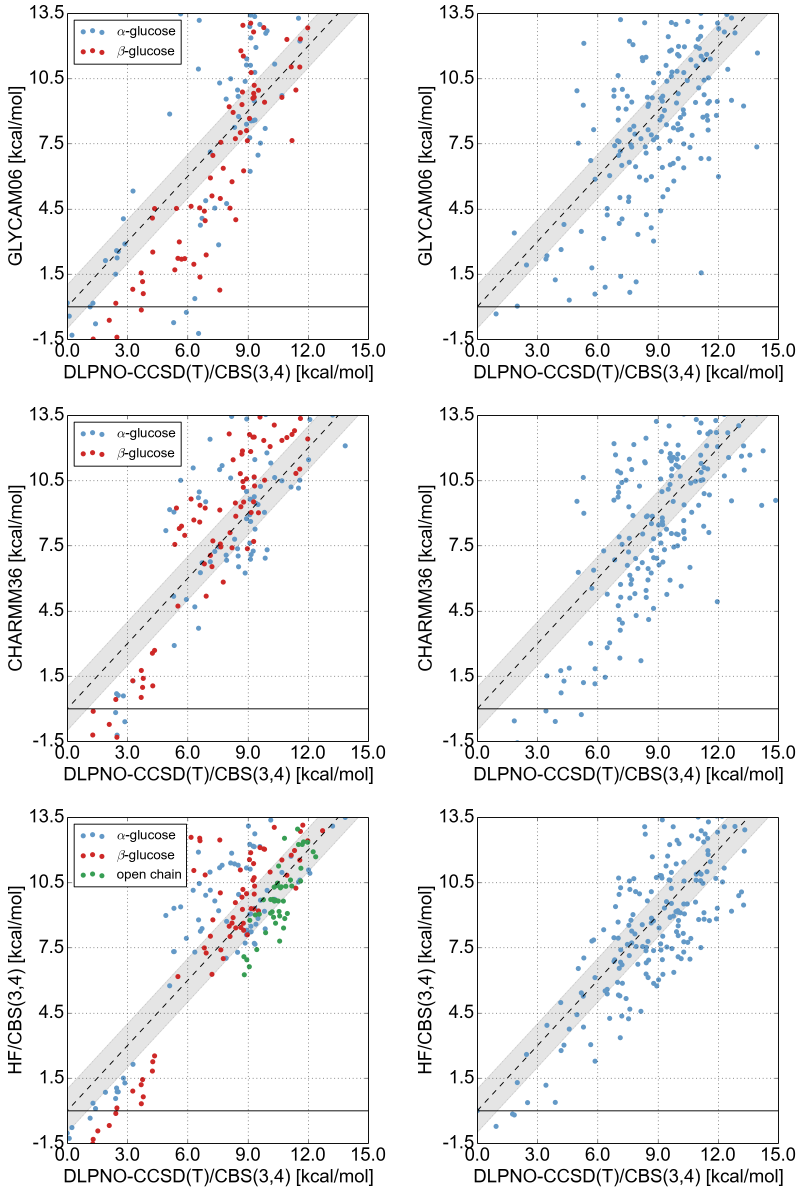

Figure 1: Force Fields and Hartree-Fock correlation plots for glucose (left column) and α-maltose (right column).

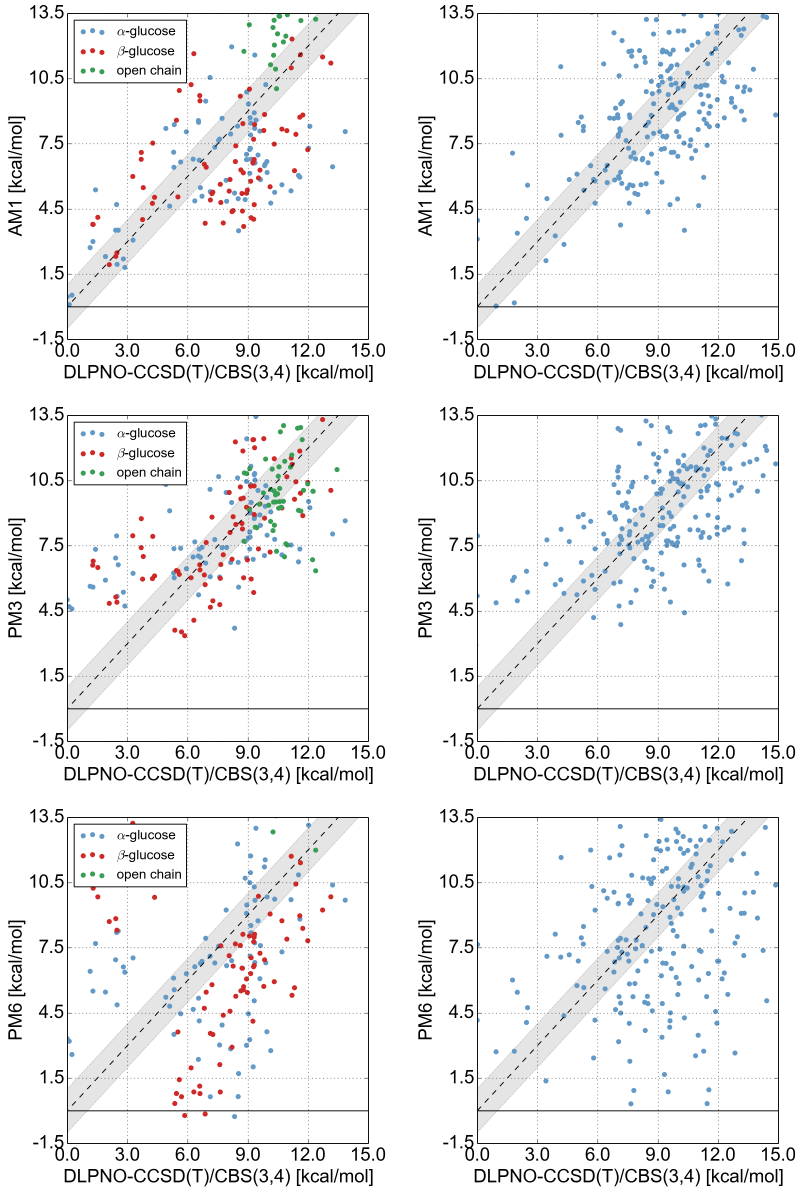

Figure 2: Correlation plots of SQM methods with older parametrization (AM1, PM3 and PM6) for glucose (left column) and α-maltose (right column).

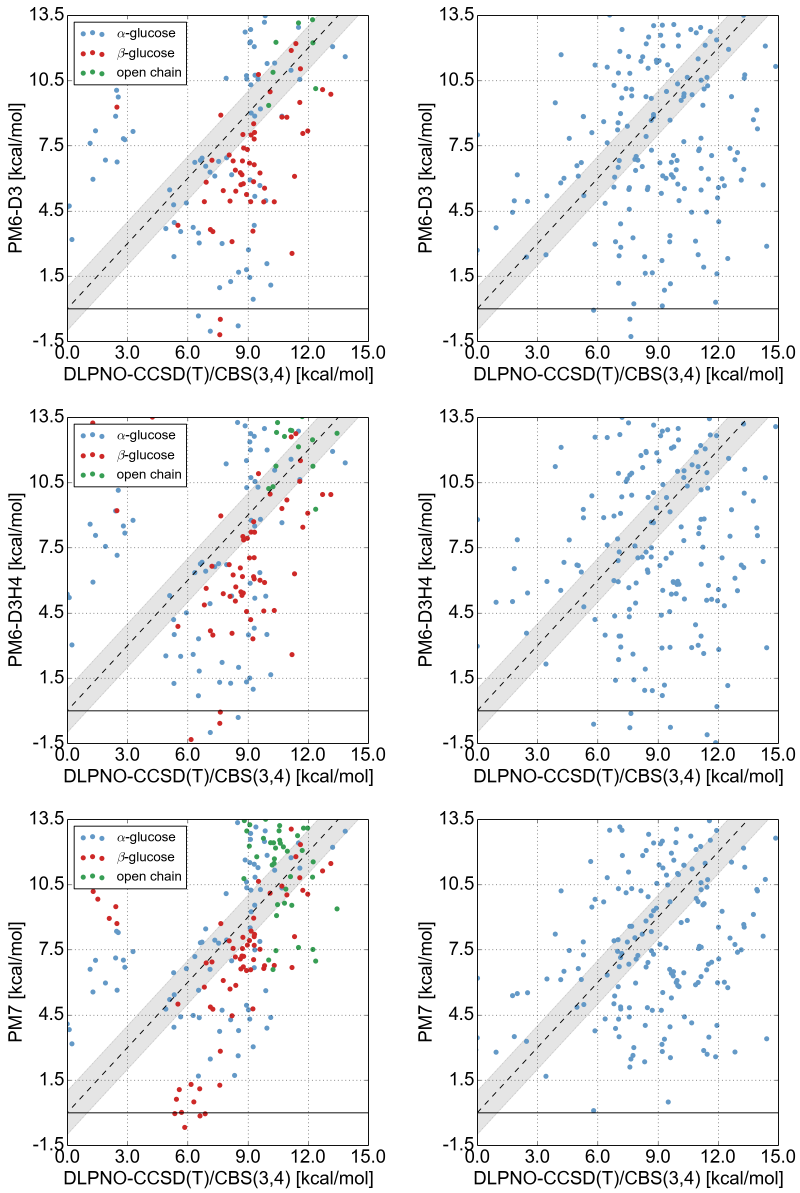

Figure 3: Correlation plots of SQM methods with newer parametrization (PM6-D3, PM6-D3H4 and PM7) for glucose (left column) and α-maltose (right column).

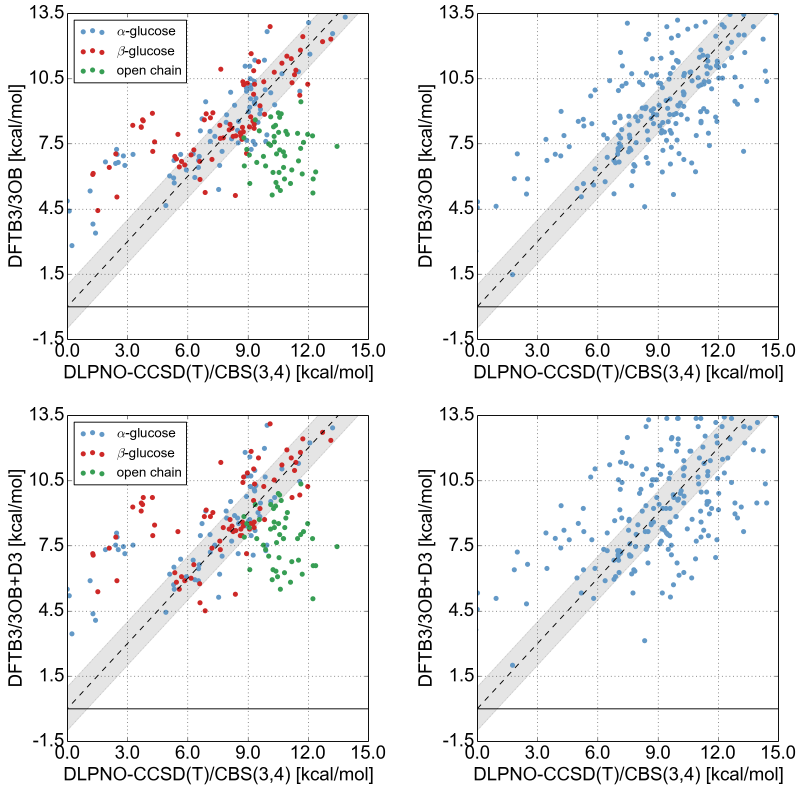

Figure 4: Correlation plots of DFTB3/3OB with and without dispersion correction. Glucose is shown in the left column and α-maltose in the right.

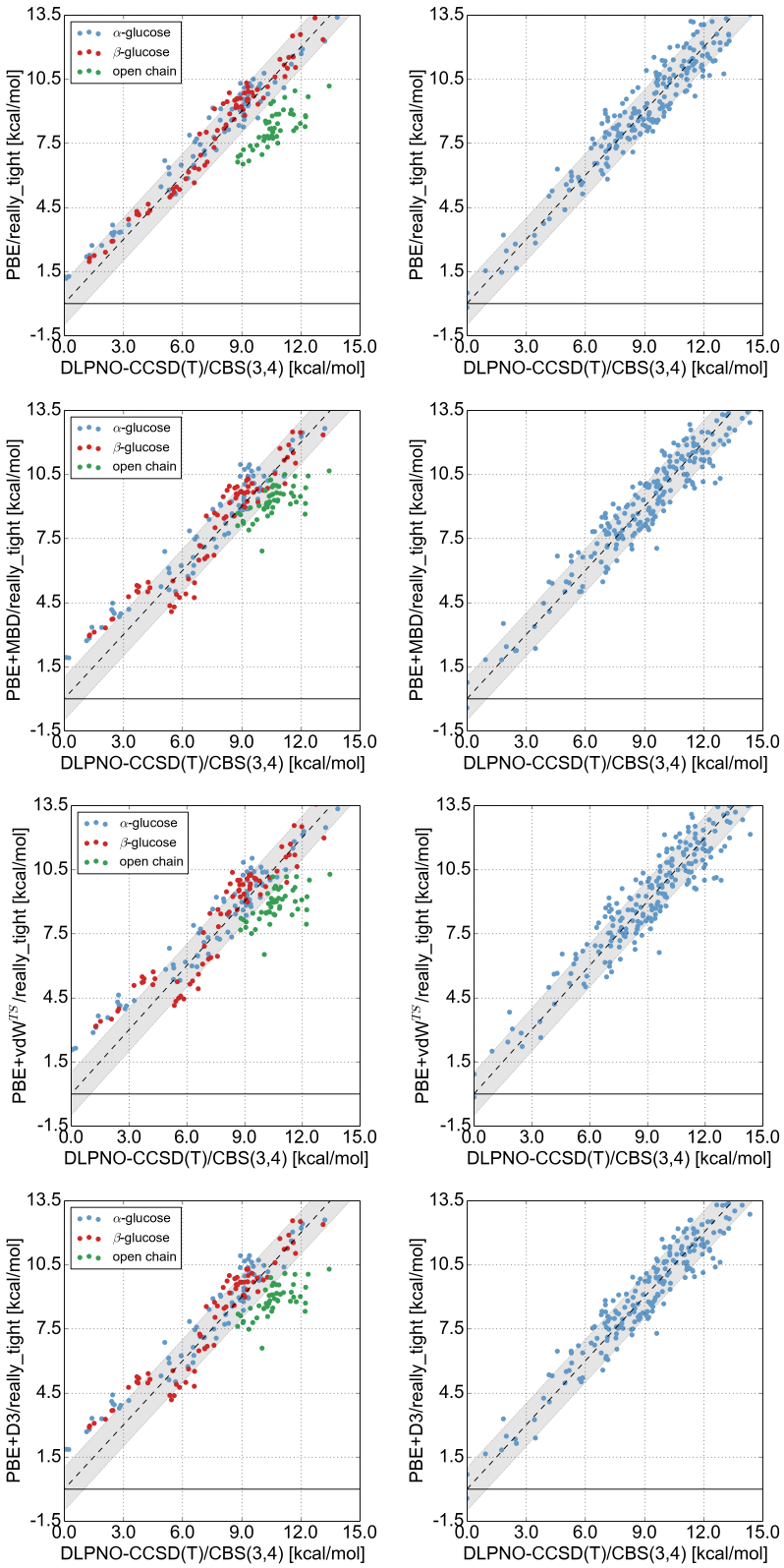

Figure 5: Correlation plots of plain PBE GGA and augmented with different dispersion corrections. Glucose is shown in the left column and α-maltose in the right.

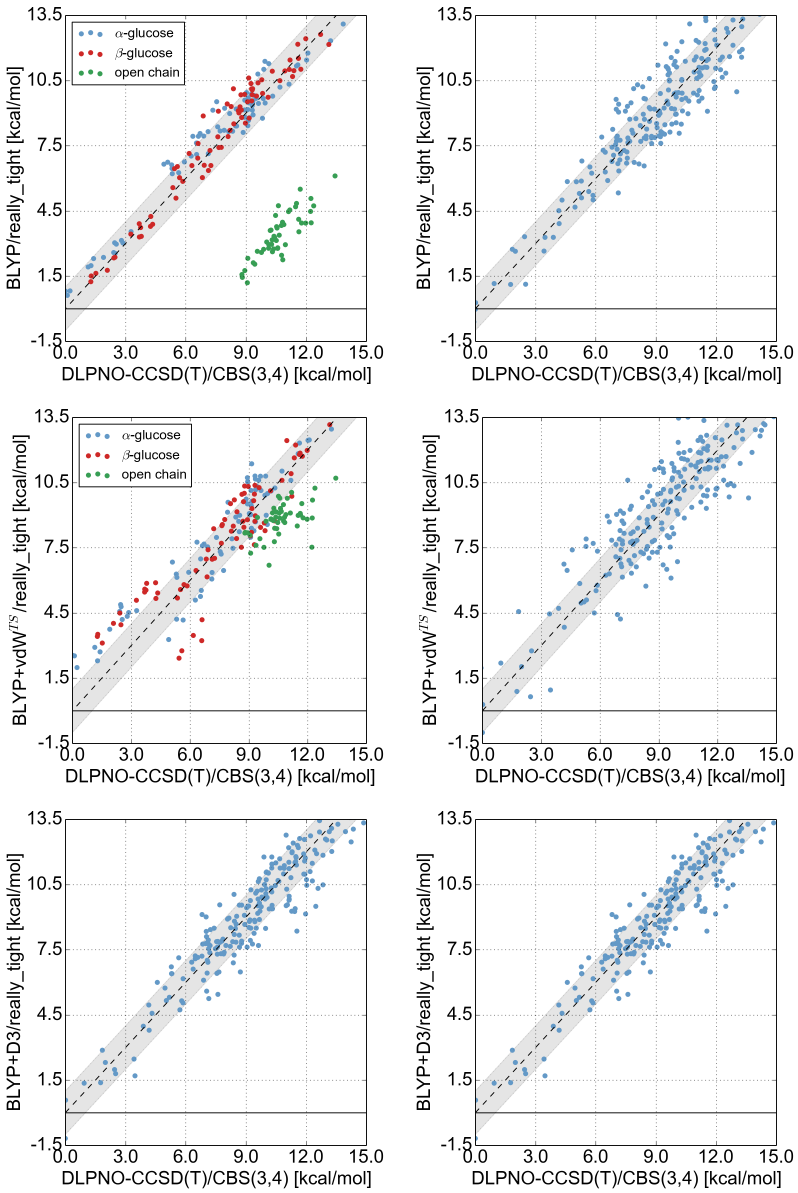

Figure 6: Correlation plots of plain BLYP GGA and augmented with different dispersion corrections. Glucose is shown in the left column and α-maltose in the right.

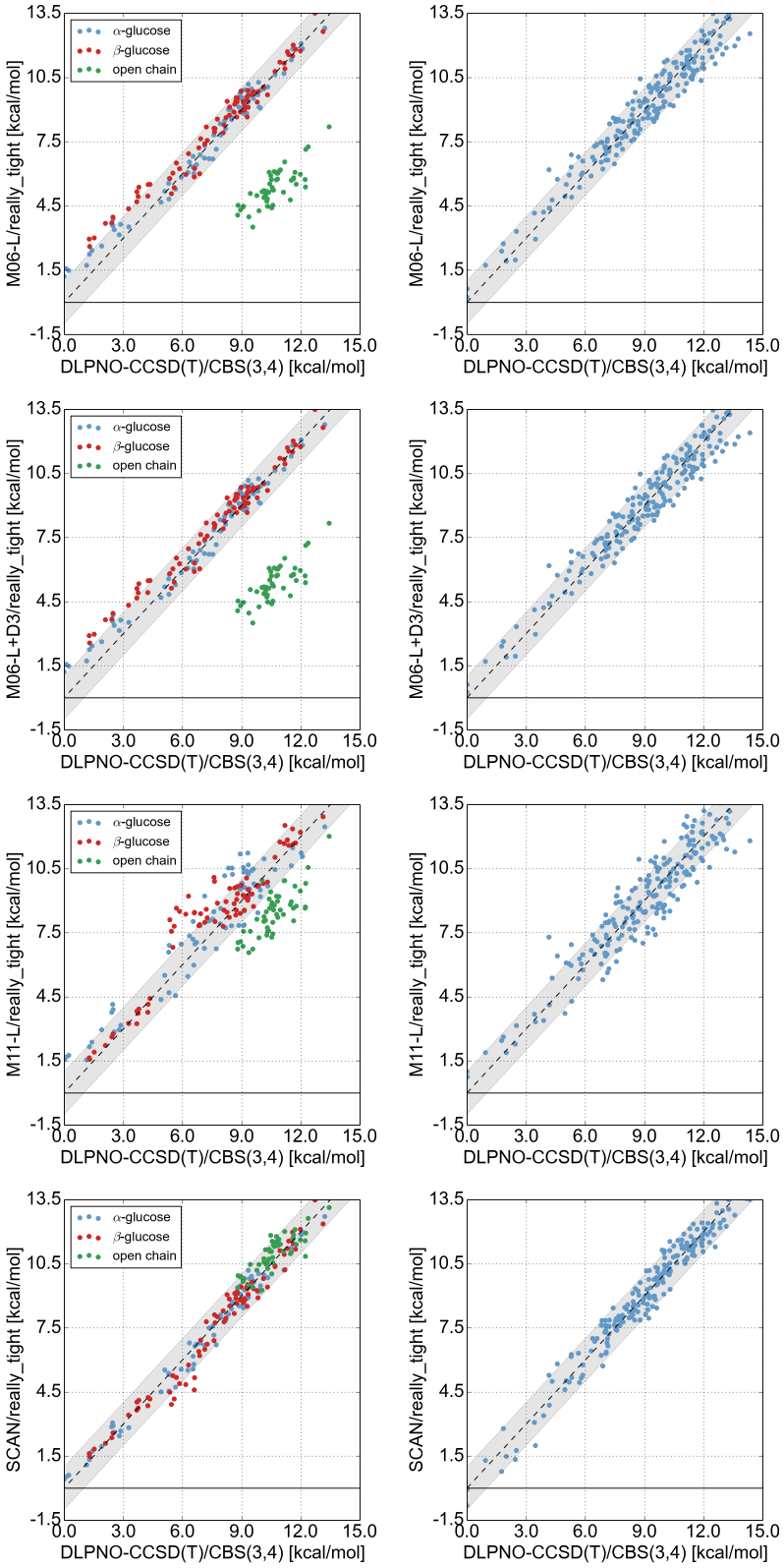

Figure 7: Correlation plots for meta-GGA functionals (M06L, M11L and SCAN) and dispersion-corrected M06L+D3 meta-GGA. Glucose is shown in the left column and α-maltose in the right.

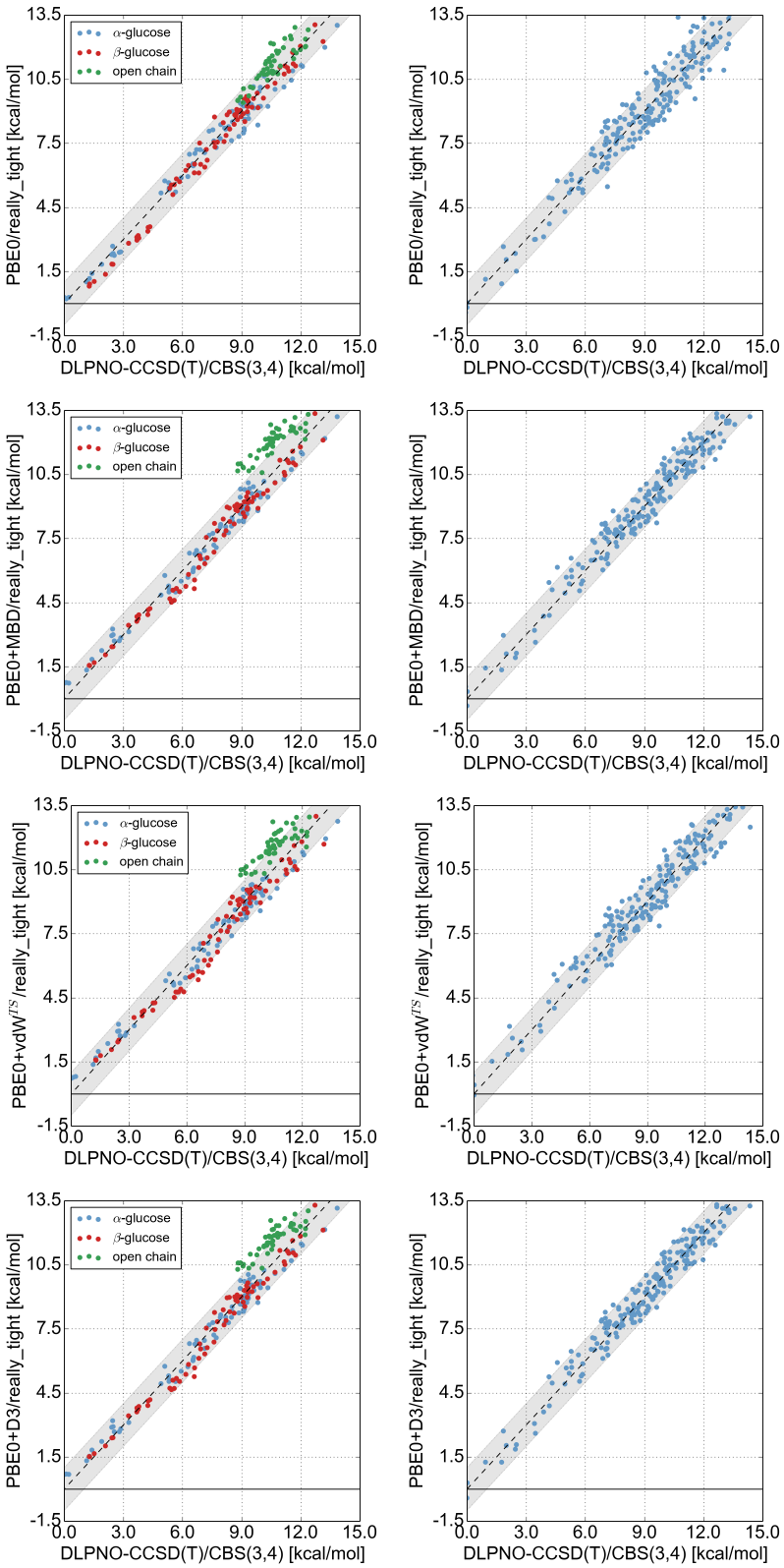

Figure 8: Correlation plots of plain PBE0 hybrid and augmented with different dispersion corrections. Glucose is shown in the left column and α-maltose in the right.

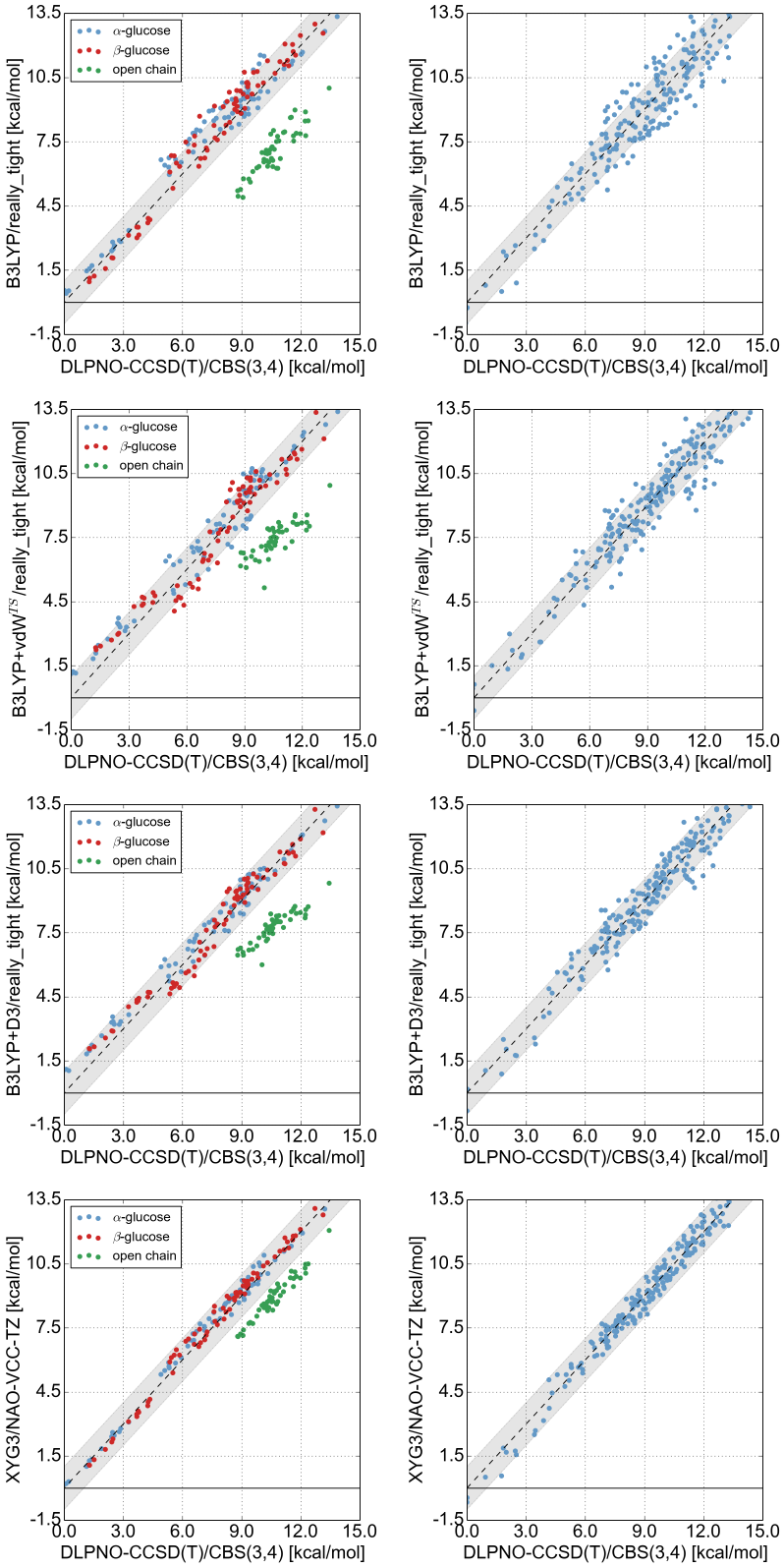

Figure 9: Correlation plots of plain B3LYP hybrid, its two dispersion-corrected counterparts (B3LYP+vdW^{TS}, B3LYP+D3) and XYG3 double-hybrid functional. Glucose is shown in the left column and α-maltose in the right.

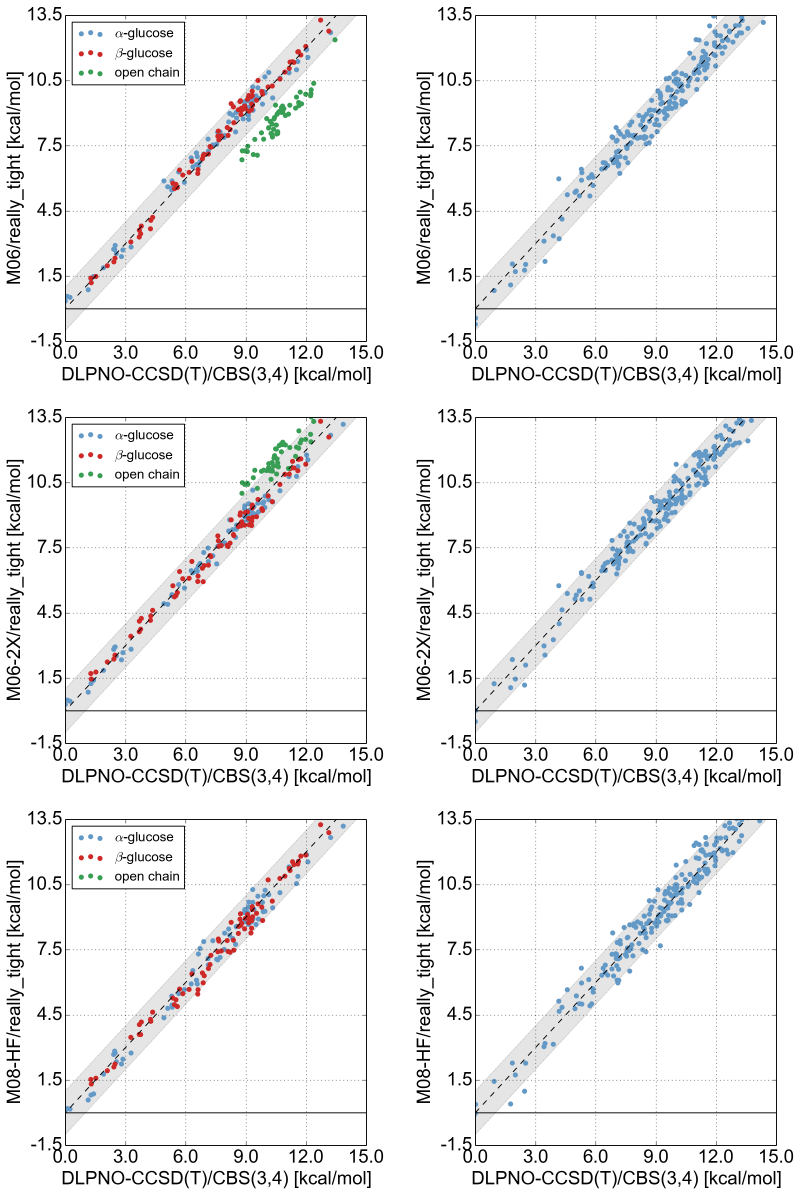

Figure 10: Correlation plots for hybrid M06-family of functionals. Glucose is shown in the left column and α-maltose in the right.

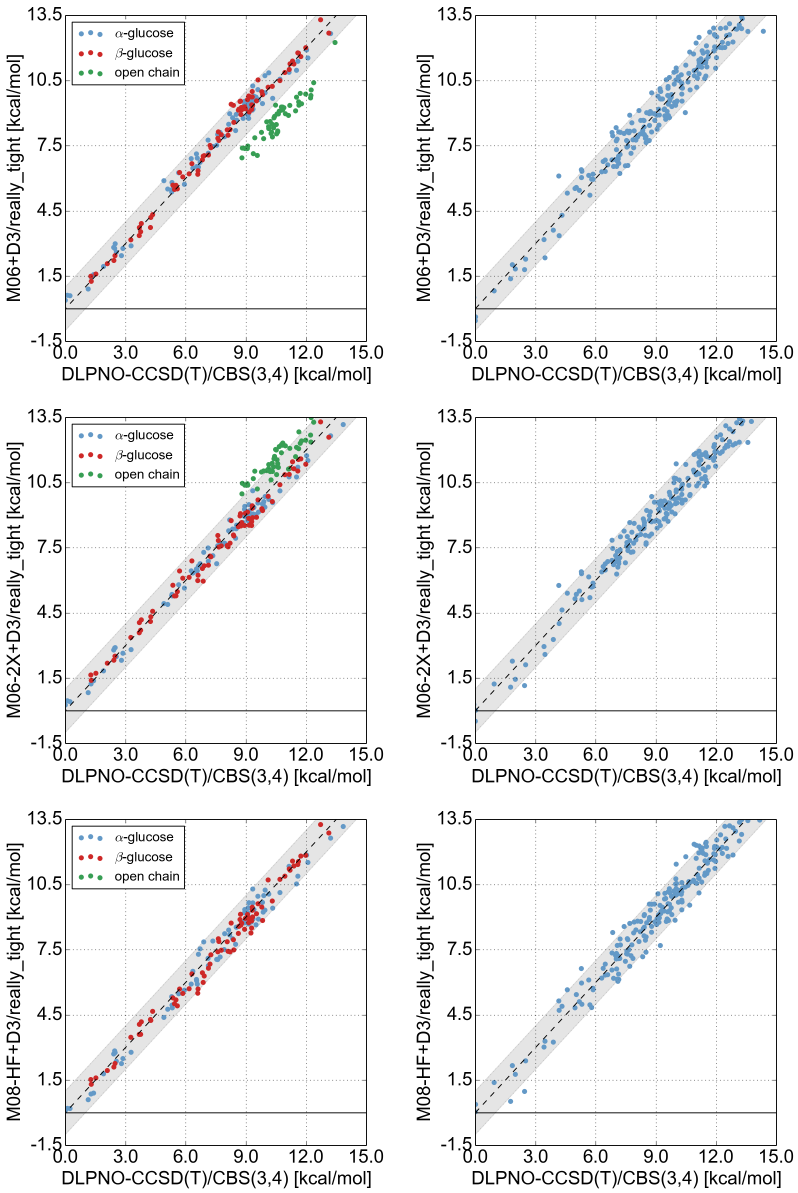

Figure 11: Correlation plots for hybrid M06-family of functionals augmented with D3 correction scheme.Glucose is shown in the left column and α-maltose in the right.

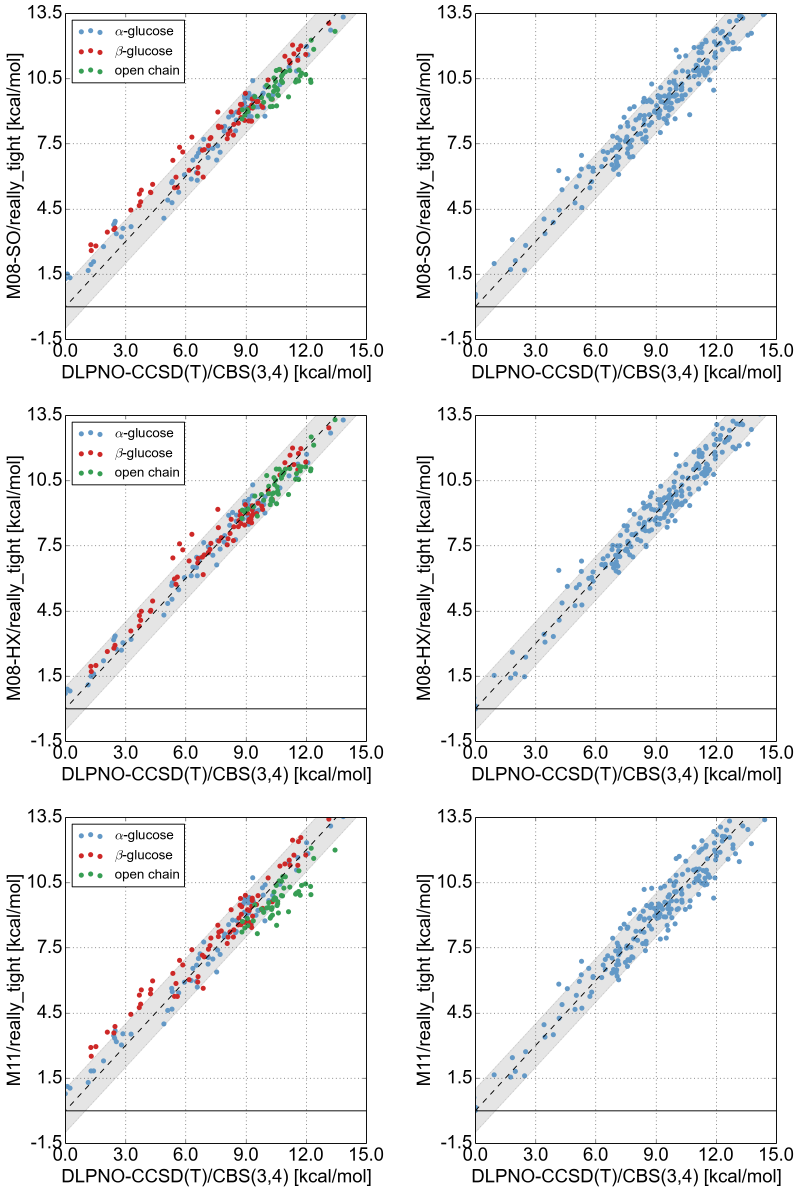

Figure 12: Correlation plots for hybrid M08-SO, M08-HX and M11 functionals. Glucose is shown in the left column and α-maltose in the right.

Fitting the PBE0 mixing parameter α

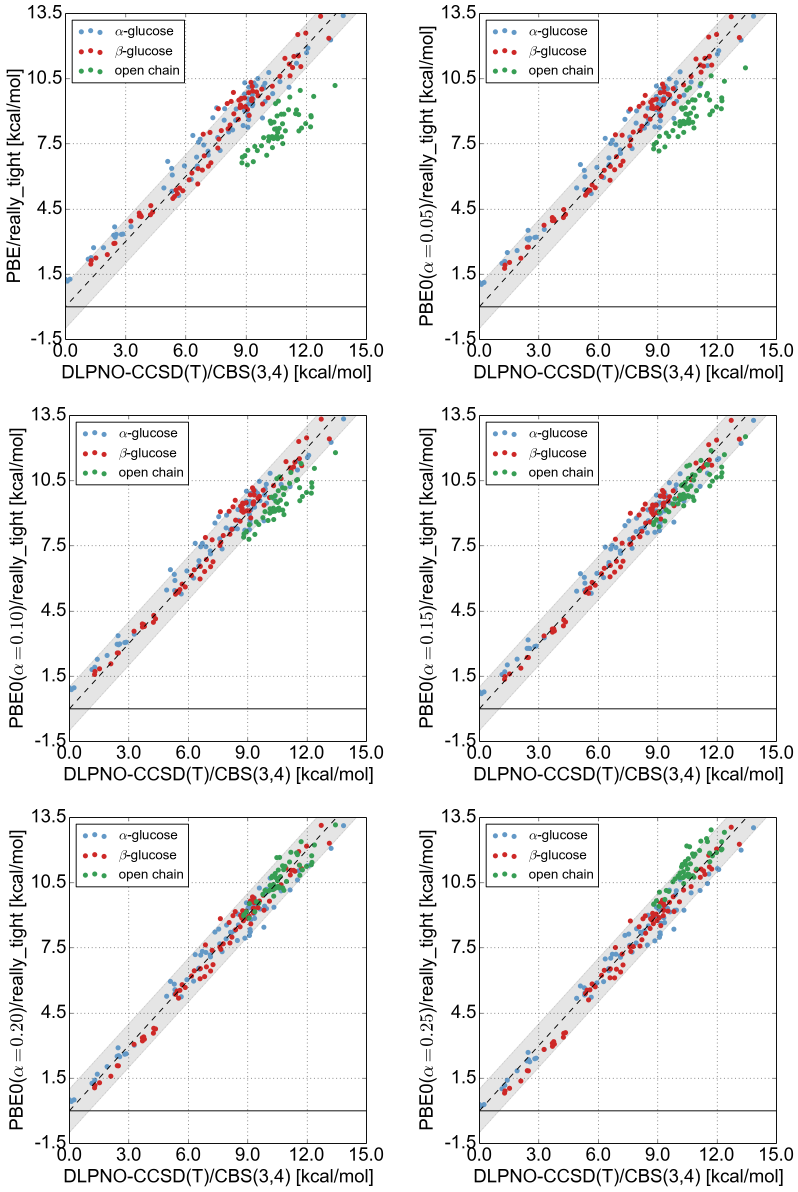

Figure 13: Correlation plots for different values of mixing parameter α. α=0 corresponds to the GGA functional PBE and α=0.25 to the original formulation of the PBE0.

Table 6: Mean absolute errors (MAE) and maximal errors (ME) (in kcal/mol) for different values of mixing parameter α. α=0 corresponds to GGA PBE and α=0.25 to the original formulation of the PBE0. Smallest MAE is emphasized in bold.

α	MAE	ME
0.00 (PBE)	0.99	3.94
0.05	0.83	2.91
0.10	0.60	2.24
0.15	0.44	1.64
0.20	**0.37**	**1.66**
0.25 (PBE0)	0.44	2.17