

Supporting Information

A comparative study of transferable theoretical aspherical pseudoatom data bank and classical force field in predicting the coulombic interaction in molecular dimers.

Prashant Kumar¹, Sławomir Bojarowski¹, Katarzyna N. Jarzembska¹, Sławomir Domagała¹, Kenno Vanommeslaeghe², Alexander D. MacKerell Jr.² and Paulina M . Dominiak^{1*}

¹*Faculty of chemistry, University of Warsaw, Pasteura 1, 02-093 Warsaw, Poland*

²*Department of Pharmaceutical Sciences, University of Maryland School of Pharmacy,
20 Penn Street HSF II, Baltimore, Maryland 21201, United States*

* Electronic address: pdomin@chem.uw.edu.pl

Number of Pages: 13

Electrostatic interaction using SPDFG

This method is based on a set of orthogonal (Rys) polynomials, which yields a simple general formula for integrals over basis functions, X , of arbitrarily high angular momentum:

$$\langle X_i(1)X_j(1)|r_{12}^{-1}|X_k(2)X_l(2)\rangle = \sum_{\alpha=1}^N I_x(U_\alpha)I_y(U_\alpha)I_z^*(U_\alpha)W_\alpha \quad (1)$$

In which U_α and W_α are the roots and weights of the N^{th} order Rys polynomial and I_x , I_y , and I_z^* are simple two dimensional integrals, evaluated using efficient and compact recurrence formulas(1).

To calculate reference energy particularly S66 dataset, this was calculated using the standard aug-cc-pVTZ basis set. This calculation was performed by Hobza and co-workers at Prague laboratory.

Table I: Statistics from energies computed from SPDFG, compared with DFT-SAPT(E_{POL}) for S66 dataset.

Statistical Descriptors	SPDFG
R^2	0.99
Slope	1.0
RMSE kcal mol ⁻¹	0.5
MAE kcal mol ⁻¹	0.4
ME kcal mol ⁻¹	0.2
% Error	4

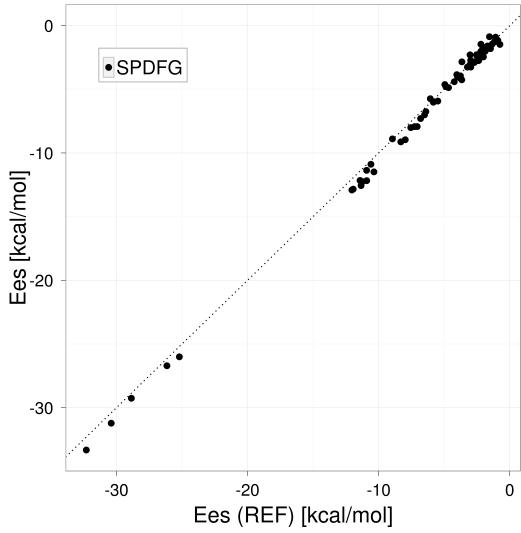


Figure 1: Comparison of Electrostatic Interaction Energies Computed from SPDFG to the DFT-SAPT, for the S66 Dataset.

DFT-SAPT

$$E^{SAPT} = E_{elst}^{(1)} + E_{exch-r}^{(1)} + E_{ind}^{(2)} + E_{exch-ind}^{(2)} + E_{disp}^{(2)} + E_{exch-disp}^{(2)} + \delta(HF) \quad (2)$$

Where denoted terms are, E_{elst} , electrostatic; E_{exch-r} , Pauli exchange repulsion; E_{ind} , induction; E_{ind} , dispersion; $E_{exch-ind}$ and $E_{exch-disp}$ exchange counterparts of induction and dispersion and higher order induction and electrostatic are summed in $\delta(HF)$.

Dipole Moments

Table V: Dipole enhancements $\Delta\mu(\%)$ and discrepancy of quantities defined as $\Delta\mu/\mu$ for the S66 and JSCH-2005 dataset. The statistics between the different methods are given $\Delta\mu(\%) = 100 \times (\mu - \mu_{REF})/\mu_{REF}$; $RMS(\mu - \mu_{REF})$ is the Root Mean Square Error between μ and μ_{REF} . The discrepancy factor is defined here as: $\Delta\mu/\mu = \sqrt{\left[\frac{\sum_1^N (\mu - \mu_{REF})^2}{N} \right] / [RMS_{(\mu)} \cdot RMS_{(\mu REF)}]}$

Statistical descriptors	REF	UBDB	AM1-BCC	RESP	CHARMM
JSCH 2005 Data set					
$\Delta\mu(\%)$		12	5	10	31
$RMS(\mu)$	6.80	8.49	9.70	10.84	11.90
$RMS(\mu - \mu_{REF})$		0.80	0.51	0.54	1.19
$R^2(\mu, \mu_{REF})$		0.88	0.92	0.98	0.76
$\Delta\mu/\mu$		10.54%	6.38%	6.32%	13.26%
S66 Data set					
$\Delta\mu(\%)$		-1	35	15	17
$RMS(\mu)$	2.23	2.21	2.83	2.48	2.55
$RMS(\mu - \mu_{REF})$		0.13	0.87	0.31	0.52
$R^2(\mu, \mu_{REF})$		0.99	0.83	0.99	0.97
$\Delta\mu/\mu$		5.88%	32.26%	13.44%	22.08%

Table VI: Statistics from computed dipole moment magnitudes from RESP-SCRF point charge model for single molecules from the S66 dataset.

Statistical Descriptors	RESP-SCRF
R^2	0.99
Slope	0.71
RMSE Debye	0.86
MAE Debye	0.58
ME Debye	0.56

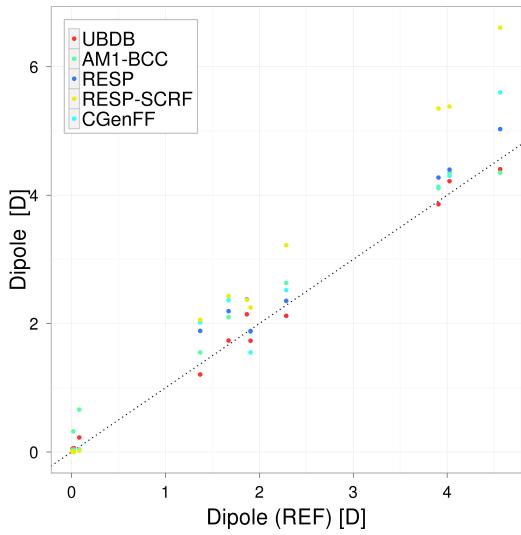


Figure 2: Isolated molecules dipole moments computed from tested methods are compared with those computed directly from the electronic wavefunction from B3LYP/aug-cc-pVTZ calculations (REF) for S66 dataset.

Molecular geometries

Table VII: Description of type of interaction present in molecular geometries of JSCH-2005 dataset.

Complex	Abbreviations	Geometry	Monomer name
Hydrogen bonded DNA base pairs			
1012_GCWC1	G...C WC	Planar	cytosine-guanine
1013_mGmCWC	mG...mC WC	Planar	methylcytosine-methylguanine
1014_ATWC	A...T WC	Planar	adenine-thymine
1015_mAmA	mA...mT H	Planar	methyladenine-methythymine
1016_8oxoGCWC8o	G...C WC pl	Planar	8-oxoguanine-cytosine
1017_ICWC	I...C WC pl	Planar	cytosine-inosine
1018_GUwobble	G...U wobble	Planar	guanine-uracil
1019_CC	C...CH+	Planar	H3-cytosine+-cytosine
1020_UUcalcutta	U...U Calcutta pl	Planar	uracil-uracil
1021_UUpL	U...U pl	Planar	uracil-uracil
1022_6thioGCWC	6tG...C WC	Planar	cytosine-6-thioguanine

Table VII (*continued*)

1059_GA10315TGis	GA10/3.15T//Gis	Interstrand	thymine-guanine
1060_AG08319TGis	AG08/3.19T//Gis	Interstrand	guanine-thymine
1061_AG08319ACis	AG08/3.19A//Cis	Interstrand	adenine-cytosine
1062_TG0319AGis	TG03.19A//Gis	Interstrand	adenine-guanine
1063_TG0319TCis	TG03.19T//Cis	Interstrand	thymine-cytosine
1064_GT10315TCis	GT10/3.15T//Cis	Interstrand	thymine-cytosine
1065_GT10315AGis	GT10/3.15A//Gis	Interstrand	adenine-guanine
1066_AT10326TTis	AT10/3.26T//Tis	Interstrand	thymine-thymine
1067_AT10326AAis	AT10/3.26A//Ais	Interstrand	adenine-adenine
1068_TA08316AAis	TA08/3.16A//Ais	Interstrand	adenine-adenine
1069_TA08316TTis	TA08/3.16T//Tis	Interstrand	thymine-thymine
1070_AA0324ATis	AA0/3.24A//Tis	Interstrand	adenine-thymine
1071_AA0324TAis	AA0/3.24T//Ais	Interstrand	adenine-thymine
1072_AA	A...A IS	Interstrand	methyladenine-methyladenine
1073_TT	T...T IS	Interstrand	thymine-thymine
1074_GGinterstrand	G...G IS	Interstrand	methylguanine-methylguanine
1075_CC	C...C IS	Interstrand	methylcytosine-methylcytosine
1076_AG	A...G IS	Interstrand	methyladenine-methylguanine
1077_TC	T...C IS	Interstrand	methylcytosine-thymine
1078_CA	C...A IS	Interstrand	cytosine-adenine
1079_GG9	G...G IS	Interstrand	guanine-guanine
1080_GG10	G...G IS	Interstrand	guanine-guanine
1081_CC11	C...C IS	Interstrand	cytosine-cytosine
Stacked base pairs			
1082_GCS	G...C S	Stacked	guanine-cytosine
1083_mGmCS	mG...mC S	Stacked	methylguanine-methylcytosine
1084_ATS1	A...T S	Stacked	adenine-thymine
1085_mA mTS	mA...mT S	Stacked	methyladenine-methylthymine
1086_CC1	CC1	Stacked	cytosine-cytosine
1087_CC2	CC2	Stacked	cytosine-cytosine
1088_CC3	CC3	Stacked	cytosine-cytosine
1089_CC4	CC4	Stacked	cytosine-cytosine
1090_CC5	CC5	Stacked	cytosine-cytosine
1091_CC6	CC6	Stacked	cytosine-cytosine
1092_CC7	CC7	Stacked	cytosine-cytosine
1093_CC8	CC8	Stacked	cytosine-cytosine
1094_CC9	CC9	Stacked	cytosine-cytosine

Table VII (*continued*)

1095_CC10	CC10	Stacked	cytosine-cytosine
1096_CC11	CC11	Stacked	cytosine-cytosine
1097_CC12	CC12	Stacked	cytosine-cytosine
1098_CC13	CC13	Stacked	cytosine-cytosine
1099_CC14	CC14	Stacked	cytosine-cytosine
1100_AAS	AAS	Stacked	adenine-adenine
1101_GGst	GGst	Stacked	guanine-guanine
1102_ACst	ACst	Stacked	adenine-cytosine
1103_GAst	GAst	Stacked	guanine-adenine
1104_CCst	CCst	Stacked	cytosine-cytosine
1105_AUst	AUst	Stacked	adenine-uracil
1106_GCst	GCst	Stacked	guanine-cytosine
1107_CUst	CUst	Stacked	cytosine-uracil
1108_UUst	UUst	Stacked	uracil-uracil
1109_GUst	GUst	Stacked	guanine-uracil
1110_GG0336GGs036	GG0/3.36 GGs036	Stacked	cytosine-cytosine
1111_GG0336CCs036	GG0/3.36 CCs036	Stacked	guanine-guanine
1112_AA20305AAs2005	AA20/3.05 AAs2005	Stacked	adenine-adenine
1113_AA20305TTs2005	AA20/3.05 TTs2005	Stacked	thymine-thymine
1114_GC0325GCs	GC0/3.25C//Cis	Stacked	cytosine-guanine
1115(CG)0319GCs	CG0/3.19G//Cs	Stacked	cytosine-guanine
1116_GA10315AGs	GA10/3.15A//Gs	Stacked	adenine-guanine
1117_GA10315TCs	GA10/3.15T//Cs	Stacked	thymine-cytosine
1118_AG08319AGs	AG08/3.19A//Gs	Stacked	adenine-guanine
1119_AG08319TCs	AG08/3.19T//Cs	Stacked	thymine-cytosine
1120_TG0319TGs	TG03.19T//Gs	Stacked	thymine-guanine
1121_TG0319ACs	TG03.19A//Cs	Stacked	adenine-cytosine
1122_GT10315TGs	GT10/3.15T//Gs	Stacked	thymine-guanine
1123_GT10315ACs	GT10/3.15A//Cs	Stacked	adenine-cytosine
1124_AT10326ATs	AT10/3.26A//Ts	Stacked	adenine-thymine
1125_TA08316ATs	TA08/3.16A//Ts	Stacked	adenine-thymine
1126_AA0324AAs	AA0/3.24A//As	Stacked	adenine-adenine
1127_AA0324TTs	AA0/3.24T//Ts	Stacked	thymine-thymine
1128_ATS2	A...T S	Stacked	methyladenine-thymine
1129_GCS1	G...C S	Stacked	methylcytosine-methylguanine
1130ACS	A...C S	Stacked	methyladenine-methylcytosine
1131_TGS	T...G S	Stacked	methylthymine-methylguanine
1132_GCS2	G...C S	Stacked	cytosine-guanine

Table VII (*continued*)

1133_AGS	A...G S	Stacked	adenine-guanine
1134_CGS	C...G S	Stacked	guanine-cytosine
1135_GCS3	G...C S	Stacked	guanine-cytosine
Amino acid pairs			
1136_F30K46	F30-K46	Aa	amino acids pairs from Rubredoxine
1137_F30L33	F30-L33	Aa	amino acids pairs from Rubredoxine
1138_F30Y13	F30-Y13	Aa	amino acids pairs from Rubredoxine
1139_F30F49	F30-F49	Aa	amino acids pairs from Rubredoxine
1140_F30Y4	F30-Y4	Aa	amino acids pairs from Rubredoxine
1141_F49C39	F49-C39	Aa	amino acids pairs from Rubredoxine
1142_F49C6	F49-C6	Aa	amino acids pairs from Rubredoxine
1143_F49K46	F49-K46	Aa	amino acids pairs from Rubredoxine
1144_F49V5	F49-V5	Aa	amino acids pairs from Rubredoxine
1145_F49W37	F49-W37	Aa	amino acids pairs from Rubredoxine
1146_F49Y4	F49-Y4	Aa	amino acids pairs from Rubredoxine
1147_F49PB(Y4V5)	F49-PB (Y4V5)	Aa	amino acids pairs from Rubredoxine
1148_F49PB(V5C6)	F49-PB (V5C6)	Aa	amino acids pairs from Rubredoxine
1149_E47K61IU5	E47-K6 (1IU5)	Aa	ion pairs (Glu-Lys salt bridges)
1150_E49K61BQ9	E49-K6 (1BQ9)	Aa	ion pairs (Glu-Lys salt bridges)
1151_E54K21SMM	E54-K2 (1SMM)	Aa	ion pairs (Glu-Lys salt bridges)
1152_E50K301BRF	E50-K30 (1BRF)	Aa	ion pairs (Glu-Lys salt bridges)
1153_E50K521BRF	E50-K52 (1BRF)	Aa	ion pairs (Glu-Lys salt bridges)
1154_E49K61BRF	E49-K6 (1BRF)	Aa	ion pairs (Glu-Lys salt bridges)

*The abbreviation used in first coulumn are same as in the original reference. For the amino acid pairs the last eight complexes also contains the PDB code of the protein from where that were oobtained.

Table VIII: Description of type of interaction present in molecular geometries of S66 dataset

Complex	Type of Interaction
Hydrogen bonded	
01_Water-Water.xyz	Hbond
02_Water-MeOH.xyz	Hbond
03_Water-MeNH2.xyz	Hbond
04_Water-Peptide.xyz	Hbond

Table VIII (*continued*)

05_MeOH-MeOH.xyz	Hbond
06_MeOH-MeNH2.xyz	Hbond
07_MeOH-Peptide.xyz	Hbond
08_MeOH-Water.xyz	Hbond
09_MeNH2-MeOH.xyz	Hbond
10_MeNH2-MeNH2.xyz	Hbond
11_MeNH2-Peptide.xyz	Hbond
12_MeNH2-Water.xyz	Hbond
13_Peptide-MeOH.xyz	Hbond
14_Peptide-MeNH2.xyz	Hbond
15_Peptide-Peptide.xyz	Hbond
16_Peptide-Water.xyz	Hbond
17_Uracil-Uracil.xyz	Hbond
18_Water-Pyridine.xyz	Hbond
19_MeOH-Pyridine.xyz	Hbond
20_AcOH-AcOH.xyz	Hbond
21_AcNH2-AcNH2.xyz	Hbond
22_AcOH-Uracil.xyz	Hbond
23_AcNH2-Uracil.xyz	Hbond
Dispersion Dominated	
24_Benzene-Benzene-pi.xyz	Dispersion
25_Pyridine-Pyridine-pi.xyz	Dispersion
26_Uracil-Uracil-pi.xyz	Dispersion
27_Benzene-Pyridine-pi.xyz	Dispersion
28_Benzene-Uracil-pi.xyz	Dispersion
29_Pyridine-Uracil-pi.xyz	Dispersion
30_Benzene-Ethene.xyz	Dispersion
31_Uracil-Ethene.xyz	Dispersion
32_Uracil-Ethyne.xyz	Dispersion
33_Pyridine-Ethene.xyz	Dispersion
34_Pentane-Pentane.xyz	Dispersion
35_Neopentane-Pentane.xyz	Dispersion
36_Neopentane-Neopentane.xyz	Dispersion
37_Cyclopentane-Neopentane.xyz	Dispersion
38_Cyclopentane-Cyclopentane.xyz	Dispersion
39_Benzene-Cyclopentane.xyz	Dispersion
40_Benzene-Neopentane.xyz	Dispersion
41_Uracil-Pentane.xyz	Dispersion

Table VIII (*continued*)

42_Uracil-Cyclopentane.xyz	Dispersion
43_Uracil-Neopentane.xyz	Dispersion
44_Ethene-Pentane.xyz	Dispersion
45_Ethyne-Pentane.xyz	Dispersion
46_Peptide-Pentane.xyz	Dispersion
Mixed type of interaction	
47_Benzene-Benzene.xyz	Mixed
48_Pyridine-Pyridine.xyz	Mixed
49_Benzene-Pyridine.xyz	Mixed
50_Benzene-Ethyne-pi.xyz	Mixed
51_Ethyne-Ethyne.xyz	Mixed
52_Benzene-AcOH-pi.xyz	Mixed
53_Benzene-AcNH2-pi.xyz	Mixed
54_Benzene-Water-pi.xyz	Mixed
55_Benzene-MeOH-pi.xyz	Mixed
56_Benzene-MeNH2-pi.xyz	Mixed
57_Benzene-Peptide-pi.xyz	Mixed
58_Pyridine-Pyridine-N.xyz	Mixed
59_Ethyne-Water-O.xyz	Mixed
60_Ethyne-AcOH-pi.xyz	Mixed
61_Pentane-AcOH.xyz	Mixed
62_Pentane-AcNH2.xyz	Mixed
63_Benzene-AcOH.xyz	Mixed
64_Peptide-Ethene.xyz	Mixed
65_Pyridine-Ethyne.xyz	Mixed
66_MeNH2-Pyridine.xyz	Mixed

References

- [1] M. Dupuis, J. Rys, and H. F. King. Evaluation of molecular integrals over gaussian basis functions. *J. Chem. Phys.*, 65:111–16., 1976.