Electronic Supplementary Material (ESI) for Environmental Science: Processes & Impacts.

Appendix A. Supplementary data for "3D-QSAR predictions for bovine serum albumin-water partitioning coefficients of organic anions using quantum mechanically based descriptors"

Lukas Linden[a], *Kai-Uwe Goss* [a,b], *Satoshi Endo*[a,c]*

[a] Helmholtz Centre for Environmental Research UFZ, Permoserstr. 15, D-04318 Leipzig, Germany

[b] University of Halle-Wittenberg, Institute of Chemistry, Kurt Mothes Str. 2, D-06120 Halle, Germany

[c] Osaka City University, Urban Research Plaza & Graduate School of Engineering, Sugimoto 3-3-138, Sumiyoshi-ku, 558-8585 Osaka, Japan

***corresponding author: phone/fax: ++81-6-6605-2763; e-mail: satoshi.endo@urban.eng.osaka-cu.ac.jp**

S-1 Local sigma profiles

The number of local sigma profiles depends on the default σ-interval (*delsig* parameter in COSMOsar3D) and the range of sigma values of all training set chemicals. We chose a *delsig* value of 0.006 σ [e/Å²] as it was proposed as sufficient by Klamt *et al.* (Klamt, A. *et al.*; *Journal of Chemical Information and Modeling,* 2012, **52** (8), 2157-2164). For the derivation of the LSPs a grid size of 2 Å was chosen because this grid size was also proposed by Klamt *et al.* and it is a common setting for 3D-QSAR models (Verma, J. *et al.*; *Current Topics in Medicinal Chemistry*, 2010, **10** (1), 95-115).

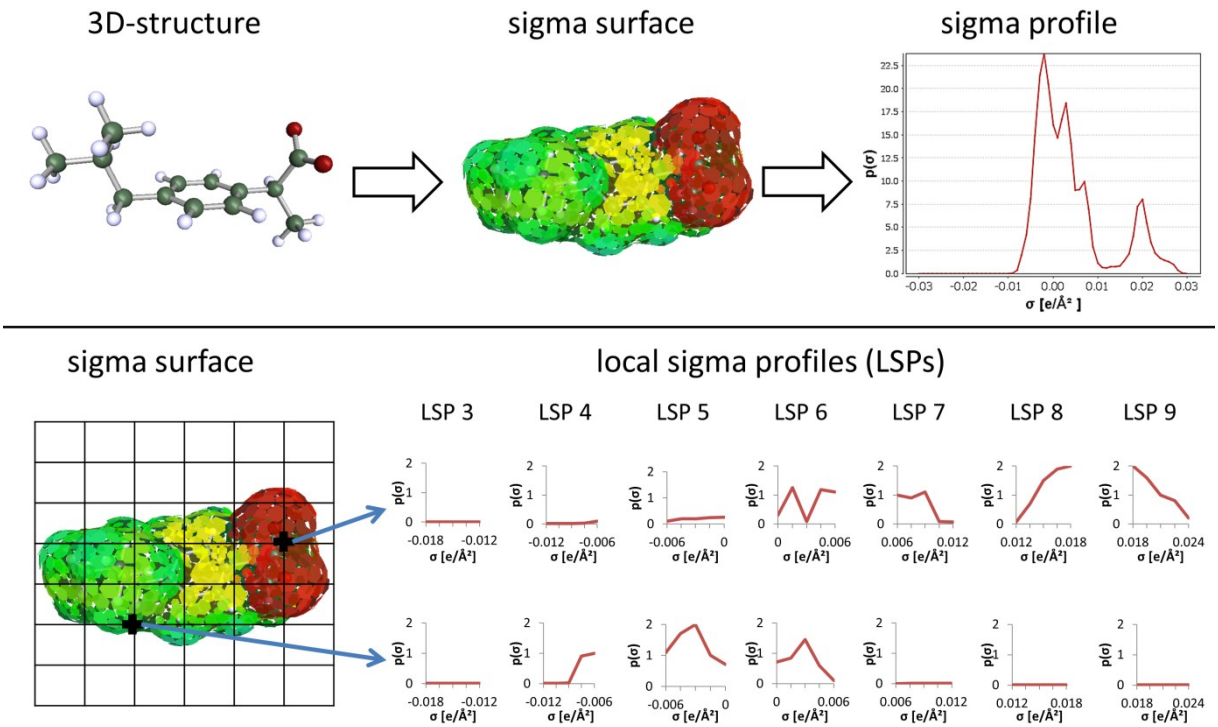


Figure SI 1 Generation of the sigma profile (top) and the local sigma profiles (LSPs) (bottom). The LSPs are shown for two exemplary grid points indicated with a black plus sign. The lattice is moved through the whole molecule in steps of 2 Å and the MIFs are derived at each grid point. The figure is designed following Linden et al. [CD paper model] and omitting LSP 1, LSP 2, and LSP 10 for the sake of clarity.

S-2 Alignment Results

The alignment of the five template chemicals is illustrated in Figures SI2.

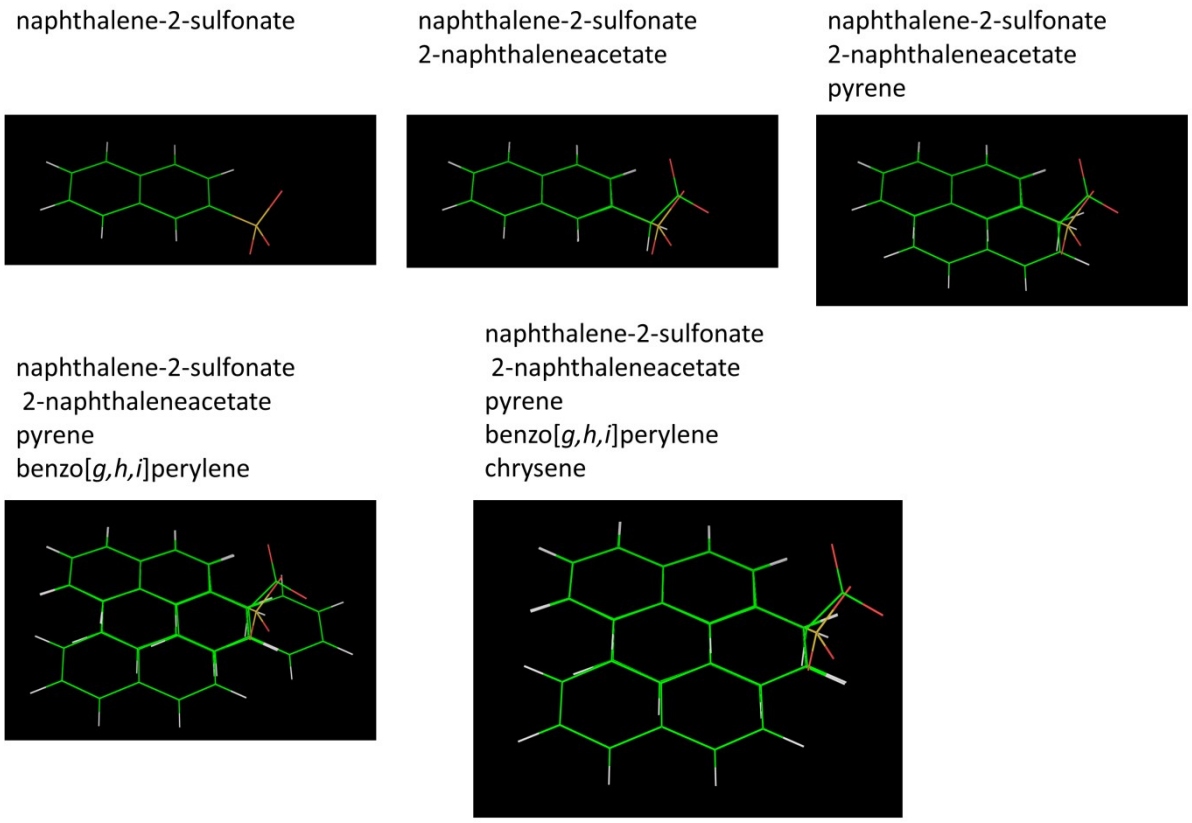


Figure SI 2 Result of the COSMOsim3D alignment for the template chemicals. Note that the order of the chemicals is arbitrary.

Comparison of the alignment of 2,4,6-trimethylbenzene sulfonate and 2,4,6-trimethylbenzoate:

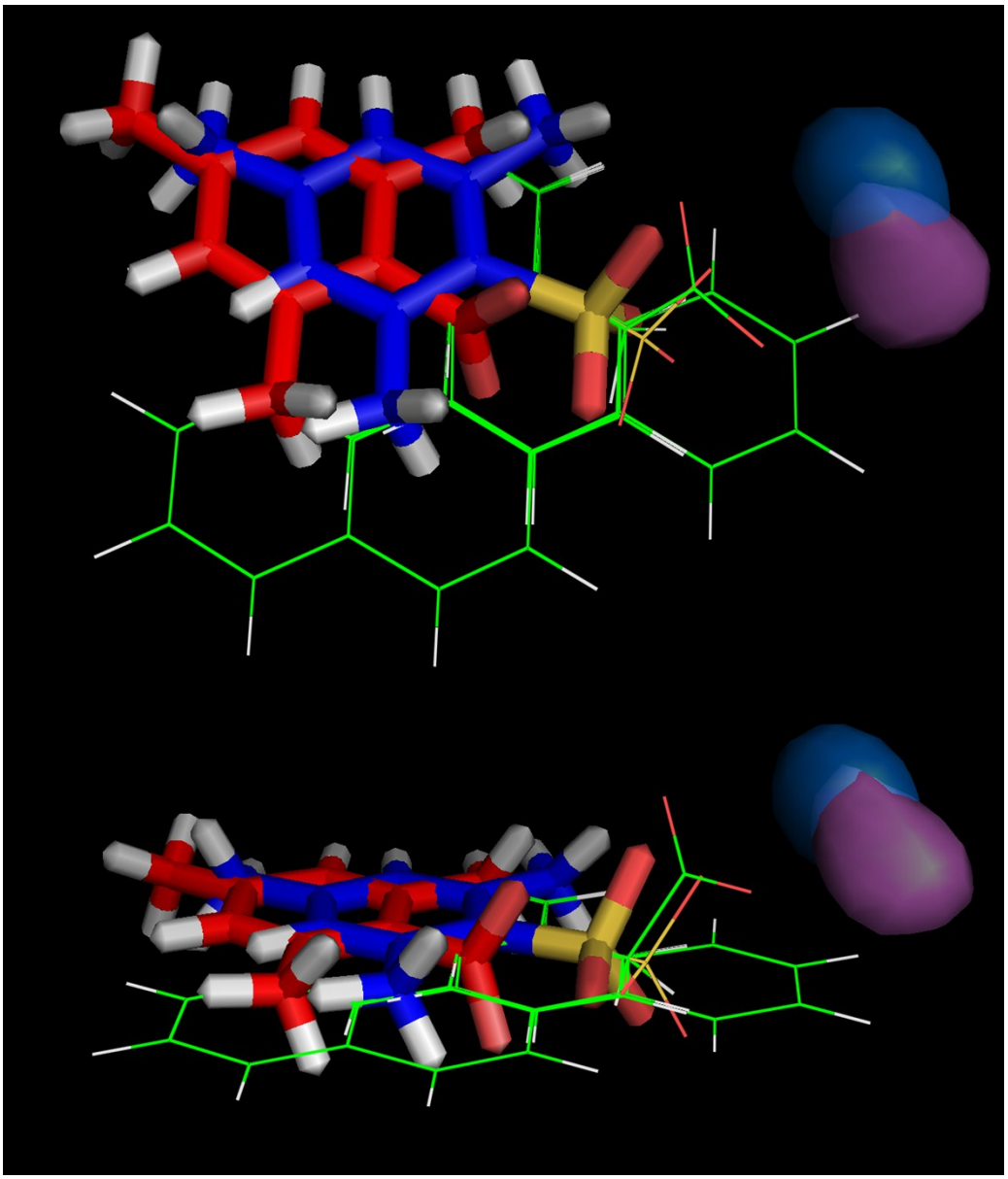


Figure SI 3 Alignment of 2,4,6-trimethylbenzene sulfonate and 2,4,6-trimethylbenzoate. The green lines in the pictures show the template chemicals while the blue sticks show 2,4,6-trimethylbenzene sulfonate and the red sticks show 2,4,6-trimethylbenzoate. The teal area indicates the area where LSP7 is positively correlated to log $K_{BSA/water}$ and the violet area indicates the area where LSP8 is positively correlated to log $K_{BSA/water}$. The alignment figures were generated using Pymol38.

S-3 Results of the random test set predictions

Table SI 1 Statistical results for the random test set predictions

test set	$q^2_{training\ set}$	$RMSE_{test\ set}$	$R^2_{test\ set}$
random1	0.57	0.52	0.69
random2	0.60	0.59	0.60
random3	0.62	0.74	0.39
random4	0.60	0.72	0.40
random5	0.54	0.58	0.60

55

56 **Table SI 2 Chemicals, experimental and predicted log $K_{BSA/water}$ values of the random test sets**

test set	name	experimental log $K_{BSA/water}$	predicted log $K_{BSA/water}$
random1	1,2-dichlorobenzene	3.03	3.17
	1-heptyne	2.49	2.7
	1-hexanol	1.64	2.65
	1-nitrooctane	3.38	3.39
	2,4,5-T (anion)	3.83	3.41
	2,4,6-trimethylbenzoic acid (anion)	2.26	1.95
	2,6-dichlorobenzoic acid (anion)	1.65	2.3
	2-naphthaleneacetic acid (anion)	4.77	4.06
	3-chlorophenol	2.35	1.91
	4-bromobenzoic acid (anion)	3.48	3.68
	4-ethylbenzoic acid (anion)	3.03	2.86
	4-fluorobenzoic acid (anion)	2.84	2.86
	4-fluorophenol	1.57	2.35
	4-iodoaniline	2.95	2.65
	4-methylbenzoic acid (anion)	2.67	3.48
	4-nitroaniline	1.69	2.45
	4-nitroanisole	2.48	2.42
	anisole	2.16	2.01
	benzophenone	2.62	3.18
	bisphenol A	2.88	2.57
	chlorobenzene	2.32	2.58
	chrysene	4.46	4.52
	dibenzofuran	3.79	3.95
	dibenzothiophene	4.16	4.13
	di-n-butyl ether	2.01	3.34
	ethylbenzene	2.7	2.42
	fenoprofen (anion)	3.92	3.68
	ketoprofen (anion)	3.35	3.28
	mefenamic acid (anion)	4.36	3.59
	methoxyflurane	1.77	2.18
	naphthalene	3.56	3.34
	n-nonane	4.45	3.55
random2	1,2,4-trimethylbenzene	3.35	2.33
	1-chlorooctane	3.85	3.31
	1-heptyne	2.49	2.83
	2,4,5-T (anion)	3.83	3.21
	2,4,6-trimethylbenzoic acid (anion)	2.26	1.84
	2-chlorobenzoic acid (anion)	1.84	1.69
	2-cyclohexylbenzoic acid (anion)	3.59	3.45
	2-naphthaleneacetic acid (anion)	4.77	4.41
	2-nitrotoluene	2.12	1.76

	2-nonanone	2.48	2
	2-phenylphenol	2.62	3.2
	3-chlorobenzoic acid (anion)	3.22	2.52
	4-iodophenol	3.41	2.92
	4-methylbenzoic acid (anion)	2.67	3.47
	a,4-dimethylphenylacetic acid (anion)	2.96	3.57
	di-n-butyl ether	2.01	3.7
	di-n-pentyl ether	3.00	3.78
	di-n-propyl phthalate	2.84	3.2
	fenoprofen (anion)	3.92	3.73
	fluoranthene	4.28	3.77
	hexafluorobenzene	1.55	2.47
	isoflurane	1.58	1.8
	ketoprofen (anion)	3.35	2.67
	mefenamic acid (anion)	4.36	4.27
	methoxyflurane	1.77	1.61
	methylpentafluorobenzene	2.32	2.67
	metolachlor	1.74	2.72
	n-propylbenzene	2.95	3.12
	phenanthrene	4.15	4.35
	pyrene	4.76	4.4
	styrene	2.76	2.67
	toluene	2.26	2.19
random3	1,2,4-trichlorobenzene	3.60	3.49
	1-naphthoic acid (anion)	2.81	3.41
	1-nitronaphthalene	3.17	2.85
	2,4,5-T (anion)	3.83	2.98
	2,4-dinitrotoluene	1.73	1.22
	2,6-dichlorobenzoic acid (anion)	1.65	2.07
	2-chloroaniline	1.95	2.69
	2-cyclohexylbenzoic acid (anion)	3.59	3.66
	2-decanone	2.88	3.97
	2-octanone	2.09	2.21
	3,4-dichlorobenzoic acid (anion)	4.06	2.83
	3-chlorobenzoic acid (anion)	3.22	3.14
	4-aminobiphenyl	2.55	3.74
	4-chlorobenzyl alcohol	2.10	3.27
	4-iodophenol	3.41	3.73
	4-methylbenzoic acid (anion)	2.67	3.49
	benzene	1.58	2.55
	benzo[*b*]fluoranthene	4.42	4.88
	benzoic acid (anion)	2.16	3.02
	diazepam	2.68	2.11
	flufenamic acid (anion)	4.83	2.93
	halothane	1.62	2.44
	indene	2.92	2.81
	indole	2.25	3.03

	ketoprofen (anion)	3.35	2.49
	mefenamic acid (anion)	4.36	3.61
	n-hexane	3.09	2.92
	n-octane	4.01	3.52
	pyrene	4.76	4.6
	styrene	2.76	2.93
	tetrachloroethene	2.40	2.73
	tri-n-butyl phosphate	2.47	2.69
random4	1-nitrooctane	3.38	2.83
	1-octanol	2.74	2.77
	2,4,5-T (anion)	3.83	2.99
	2,4-DB (anion)	4.12	2.96
	2-chloroaniline	1.95	2.72
	2-cyclohexylbenzoic acid (anion)	3.59	3.3
	2-naphthaleneacetic acid (anion)	4.77	4.64
	2-nitrotoluene	2.12	1.77
	3-chlorobenzoic acid (anion)	3.22	2.3
	3-methoxy-2-naphthoic acid (anion)	2.86	3.88
	4-aminobiphenyl	2.55	3.86
	4-iodophenol	3.41	2.99
	4-methylbenzoic acid (anion)	2.67	3.59
	4-nitroanisole	2.48	2.34
	benzene	1.58	2.54
	chlorobenzene	2.32	2.62
	coumachlor (anion)	3.37	2.74
	cyclohexane	2.01	2.26
	dibenzofuran	3.79	4.10
	di-n-propyl phthalate	2.84	3.29
	estrone	2.69	3.73
	fluoranthene	4.28	3.86
	indene	2.92	2.75
	isoflurane	1.58	2.12
	mefenamic acid (anion)	4.36	4.00
	methylpentafluorobenzene	2.32	2.05
	metolachlor	1.74	2.39
	n-hexane	3.09	2.80
	n-octane	4.01	3.36
	phenoxyacetic acid (anion)	2.57	3.14
	pyrene	4.76	4.49
	sulcotrione (anion)	1.72	3.70
random5	1,2,4-trimethylbenzene	3.35	2.39
	1,4-dibromobenzene	3.97	3.31
	1-naphthoic acid (anion)	2.81	3.01
	1-nonene	4.22	3.20
	2,4,5-T (anion)	3.83	3.20
	2,4-DB (anion)	4.12	3.05
	2-chloroaniline	1.95	2.76

2-chlorobenzoic acid (anion)	1.84	1.67
2-octanone	2.09	2.48
4-aminobiphenyl	2.55	3.58
4-bromobenzoic acid (anion)	3.48	3.87
4-chlorobenzoic acid (anion)	3.21	3.75
4-fluorophenol	1.57	2.34
4-n-octylbenzenesulfonate (anion)	5.03	4.33
anisole	2.16	2.11
bisphenol A	2.88	2.98
coumachlor (anion)	3.37	2.80
coumafuryl (anion)	2.79	2.67
diazepam	2.68	1.94
dibenzothiophene	4.16	4.20
enflurane	1.59	2.28
indole	2.25	2.90
mesitylenesulfonate (anion)	4.23	3.82
metolachlor	1.74	2.44
n-heptane	3.59	3.26
n-hexane	3.09	2.87
n-propylbenzene	2.95	3.35
phenoxyacetic acid (anion)	2.57	3.29
pyrene	4.76	4.44
styrene	2.76	2.82
tetrachloroethene	2.4	2.65
tri-n-butyl phosphate	2.47	2.55

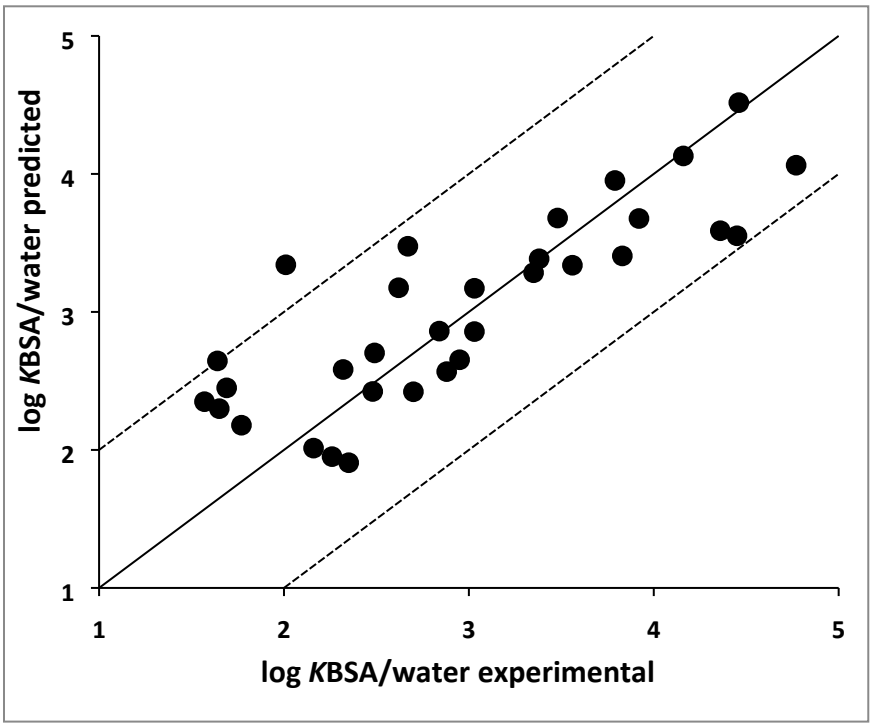


Figure SI 4 Prediction of the test set random1. The solid line indicates the 1:1 line and the dashed lines indicate a deviation of 1 log unit from the 1:1 line.

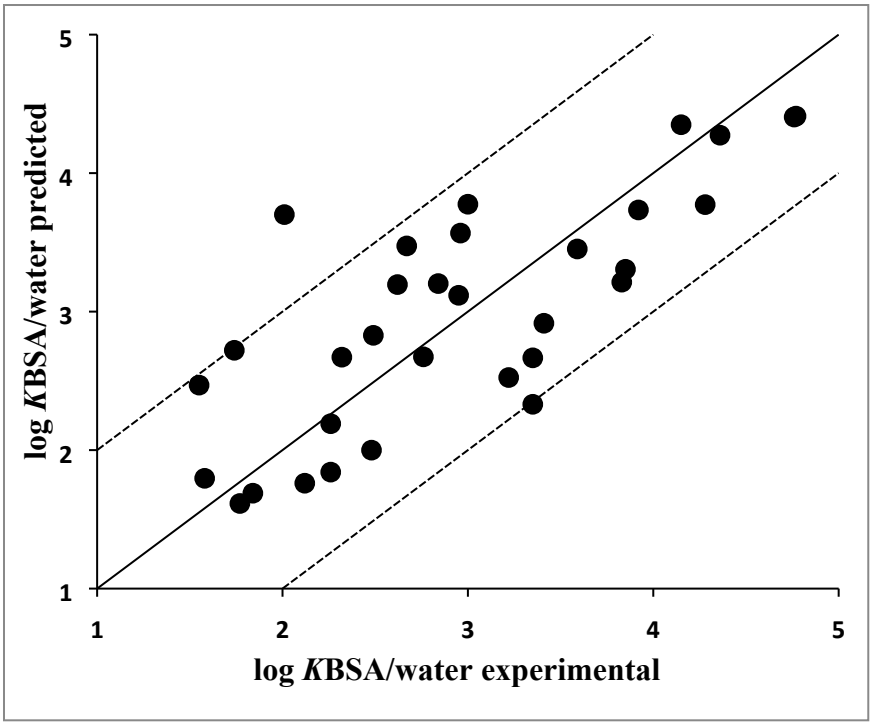


Figure SI 5 Prediction of the test set random2. The solid line indicates the 1:1 line and the dashed lines indicate a deviation of 1 log unit from the 1:1 line

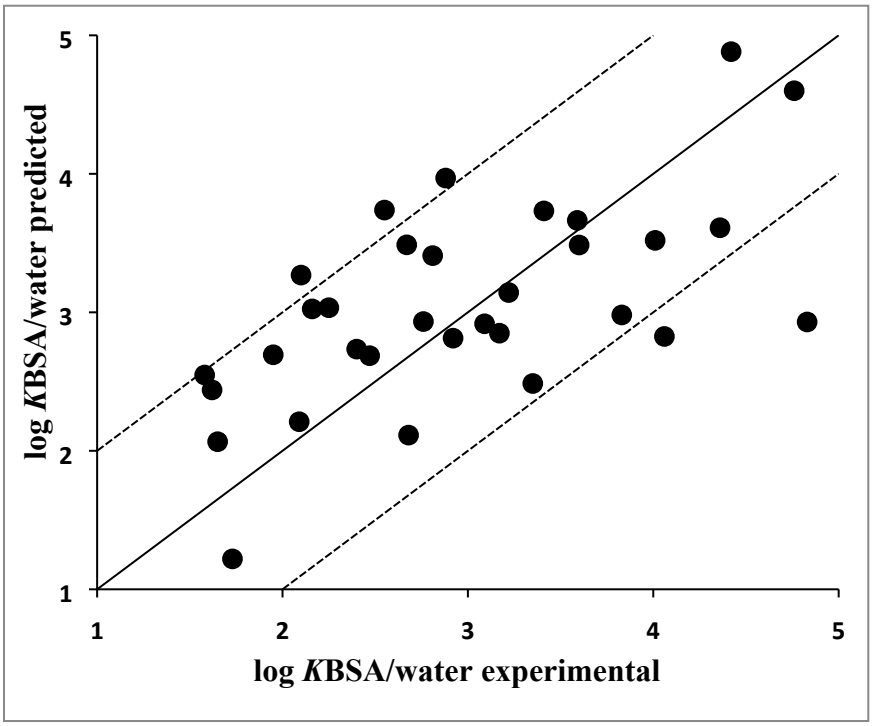


Figure SI 6 Prediction of the test set random3. The solid line indicates the 1:1 line and the dashed lines indicate a deviation of 1 log unit from the 1:1 line

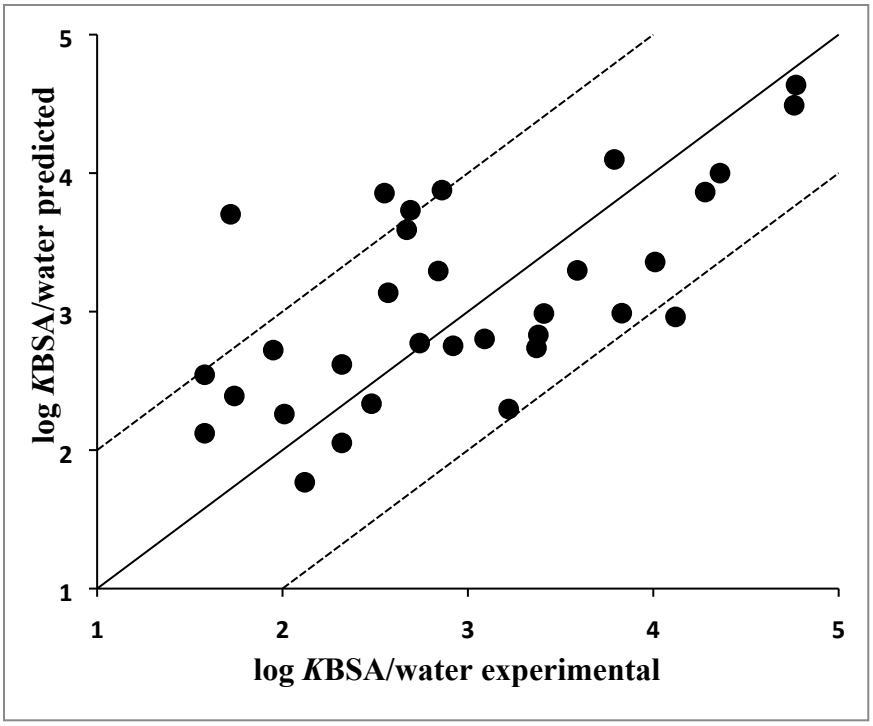


Figure SI 7 Prediction of the test set random4. The solid line indicates the 1:1 line and the dashed lines indicate a deviation of 1 log unit from the 1:1 line

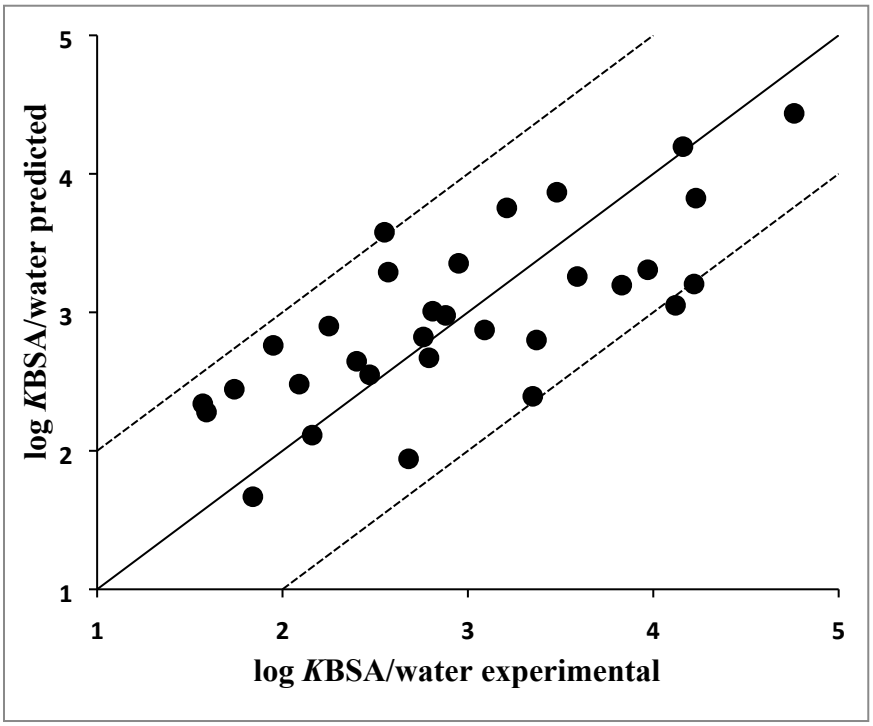


Figure SI 8 Prediction of the test set random5. The solid line indicates the 1:1 line and the dashed lines indicate a deviation of 1 log unit from the 1:1 line

S-4 Conformers of isomer pairs

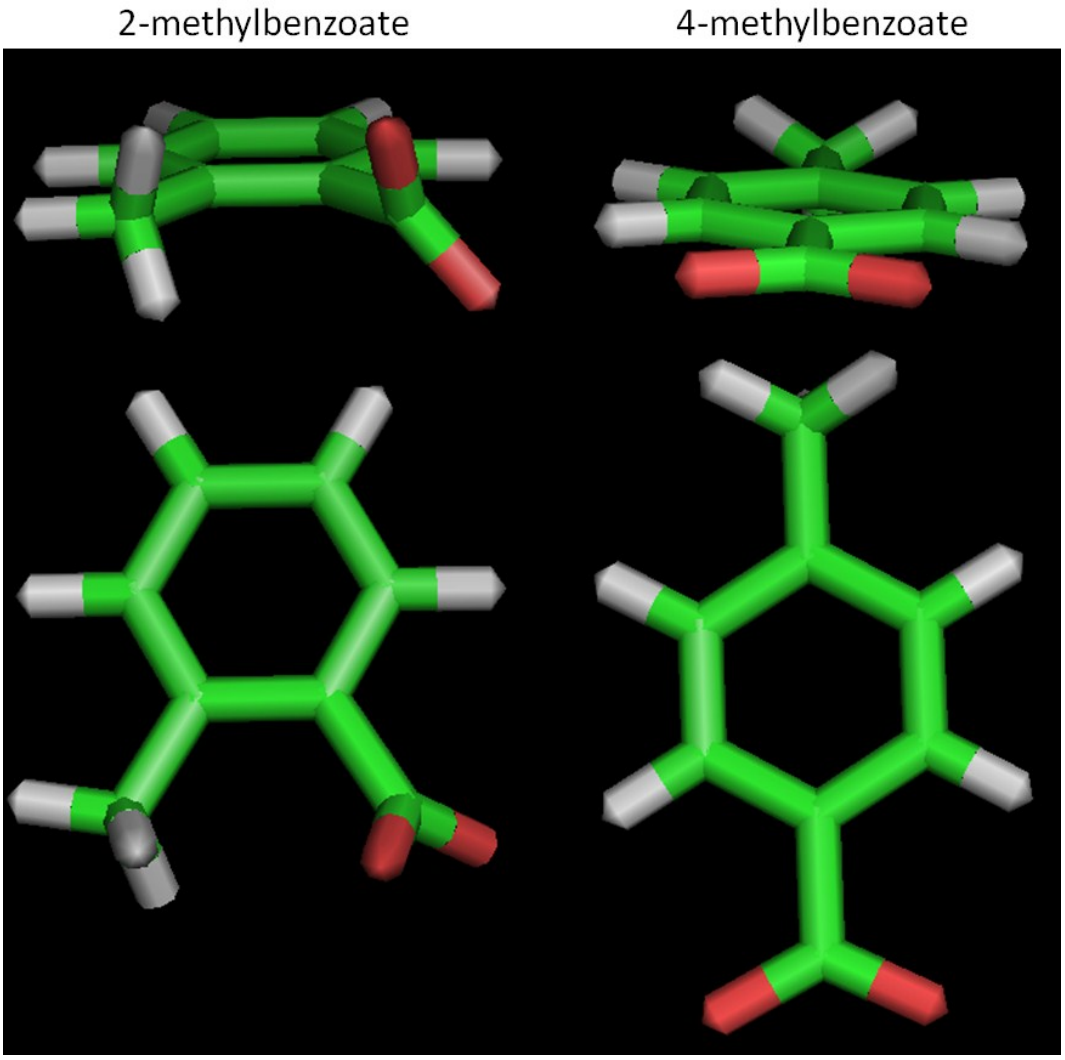


Figure SI 9 Side and top view of the used conformer of 2-methylbenzoate and 4-methylbenzoate

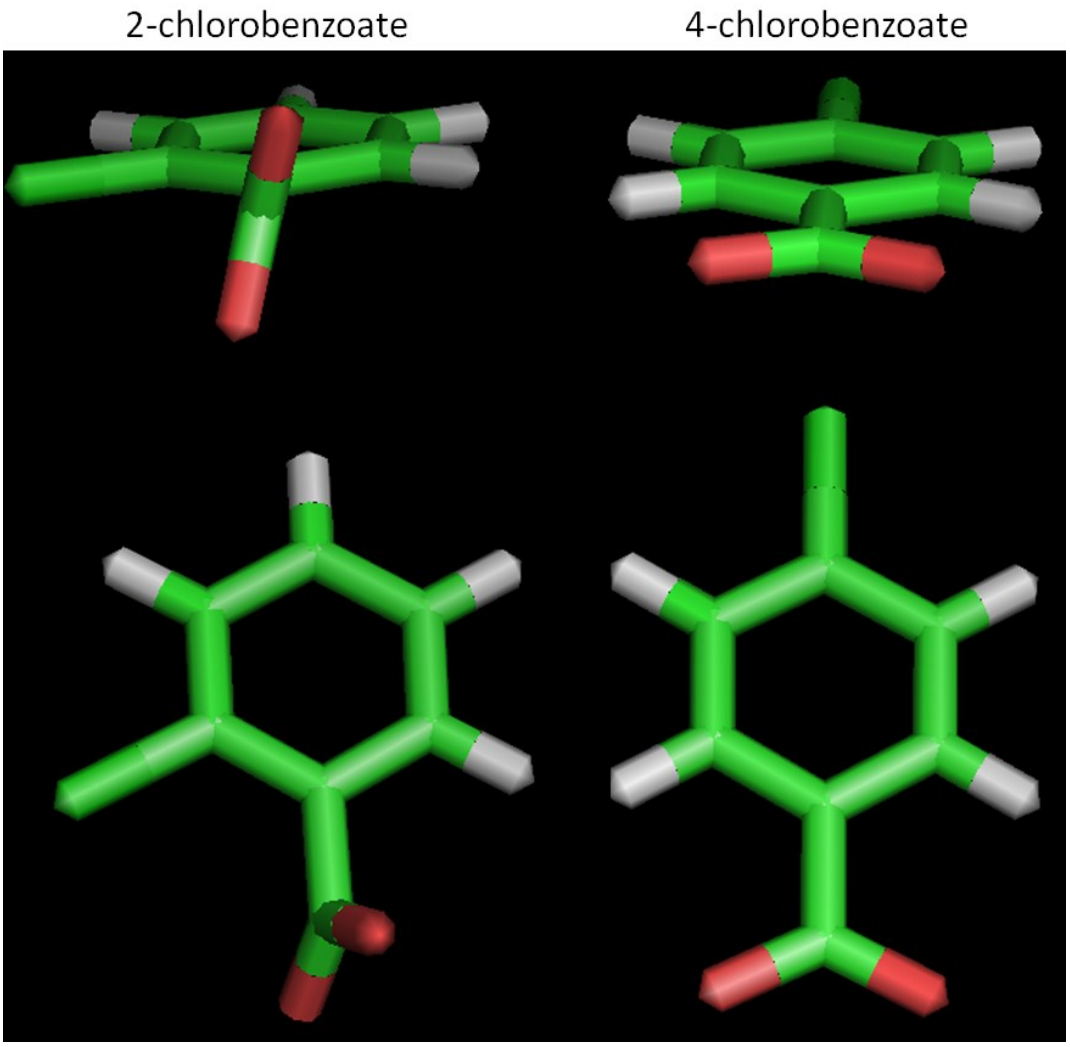
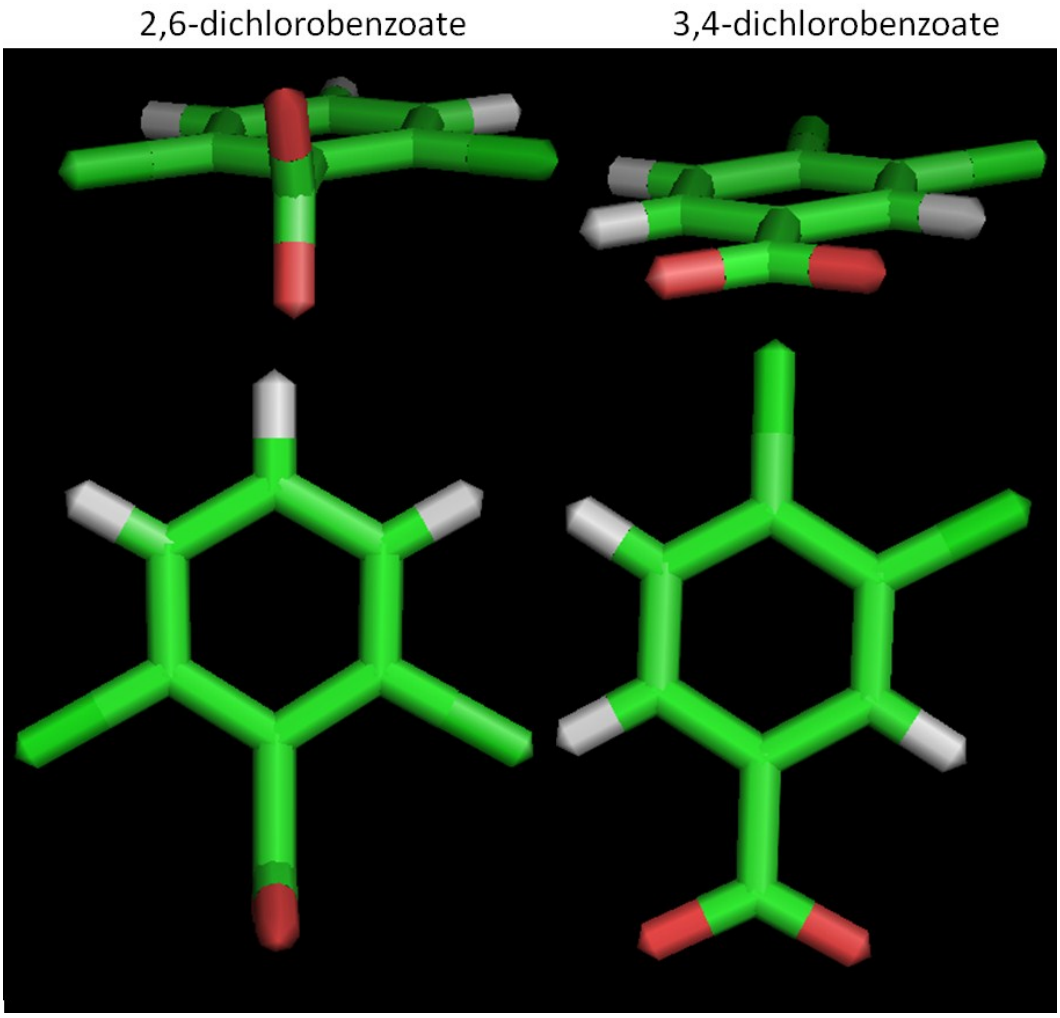


Figure SI 10 Side and top view of the used conformer of 2-chlorobenzoate and 4-chlorobenzoate



78

79 **Figure SI 11 Side and top view of the used conformer of 2,6-dichlorobenzoate and 3,4-dichlorobenzoate**

S-5 Comparison of 2,4,6-trimethylbenzene sulfonate and 2,4,6-trimethylbenzoate

The COSMOfiles of 2,4,6-trimethylbenzene sulfonate and 2,4,6-trimethylbenzoate show the surface polarisation, i.e., the distribution of the partial charges. The benzoic ring of 2,4,6-trimethylbenzene sulfonate (Fig. SI 12) is less polarized than the benzoic ring of 2,4,6-trimethylbenzoate (Fig. SI 13), which is indicated by the greenish patches of 2,4,6-trimethylbenzene sulfonate compared to the yellow patches of 2,4,6-trimethylbenzoate.

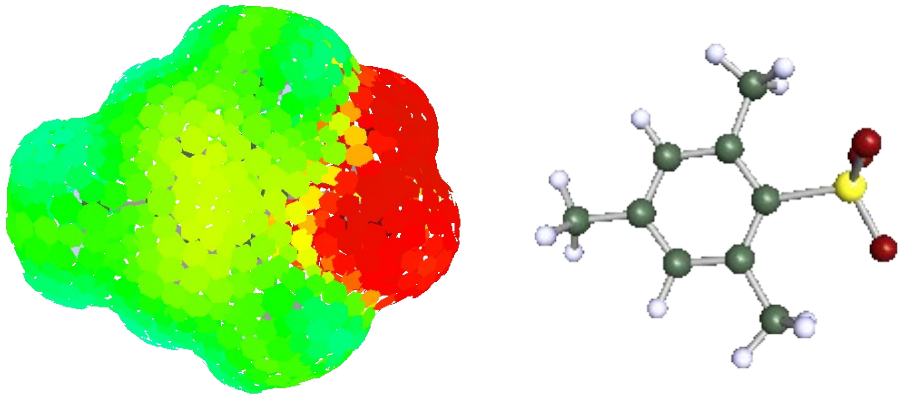


Figure SI 12 COSMOfile of 2,4,6-trimethylbenzene sulfonate. The red colour indicates a negative partial charge, the green colour indicates a neutral surface segment, and yellow patches indicate a partial charge in between.

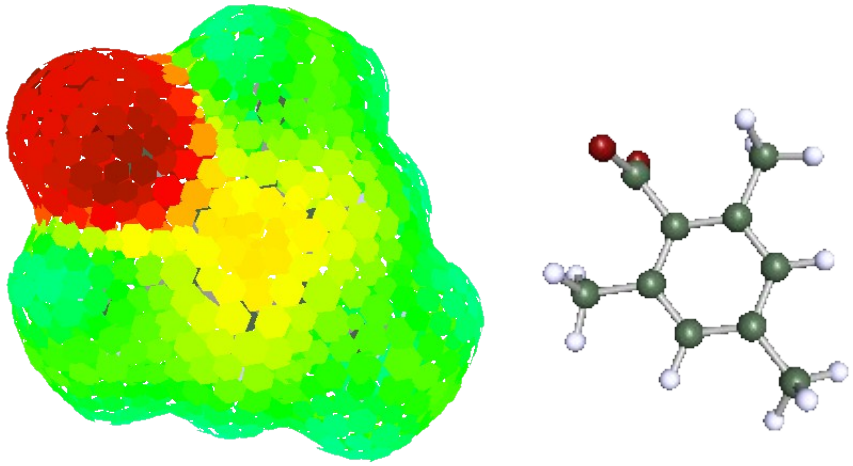


Figure SI 13 COSMOfile of 2,4,6-trimethylbenzoate. The red colour indicates a negative partial charge, the green colour indicates a neutral surface segment, yellow patches indicate a partial charge in between.

S-6 Results of Tanimoto Index

Table SI 3 Statistical results for the comparison of the different Tanimoto mean groups using the five random test sets

Tanimoto mean group		Brown–Forsythe variance analysis	Mann-Whitney U median analysis
reference group	Test group	p-value	p-value
0.60-0.70 (n=30)	0.10-0.20 (n= 10)	0.09	0.35
0.60-0.70 (n=30)	0.20-0.30 (n= 24)	0.84	0.35
0.60-0.70 (n=30)	0.30-0.40 (n= 42)	0.28	0.04
0.60-0.70 (n=30)	0.40-0.50 (n= 26)	0.24	0.12
0.60-0.70 (n=30)	0.50-0.60 (n= 24)	0.67	0.70
0.60-0.70 (n=30)	0.70-0.80 (n= 4)	0.20	0.21

For the analysis of the domain of applicability, the standard test set selection procedure was also modified to represent all chemicals in the test set once: the numbers one to four were randomly assigned to the four chemicals of each bin and then all chemicals with an identical number formed a test set while the rest of the chemicals formed the respective training set. This is necessary to get the highest possible number of individual test cases and thus an extended statistical evaluation of the domain of applicability.

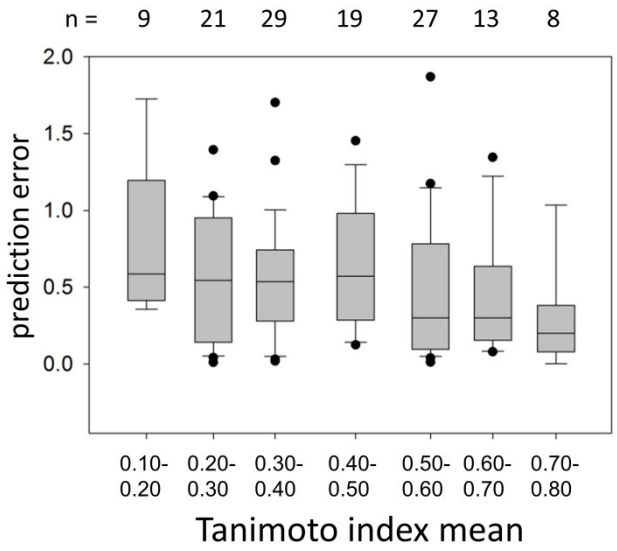


Table SI 4 Statistical results for the comparison of the different Tanimoto mean groups using the four constructed test sets

Tanimoto mean group	Brown–Forsythe variance analysis	Mann-Whitney U median analysis
reference group	p-value	p-value

0.70-0.80 (n= 8)	0.10-0.20 (n= 9)	0.36	0.01
0.70-0.80 (n= 8)	0.20-0.30 (n= 21)	0.03	0.01
0.70-0.80 (n= 8)	0.30-0.40 (n= 29)	0.51	0.03
0.70-0.80 (n= 8)	0.40-0.50 (n= 19)	0.17	0.02
0.70-0.80 (n= 8)	0.50-0.60 (n= 27)	0.21	0.28
0.70-0.80 (n= 8)	0.60-0.70 (n=13)	0.64	0.21

S-7 COSMOconf template

COSMOconf was used to create more conformers than in common applications by increasing the total number of possible conformers, reducing the energetic distance between conformers, and loosening the clustering steps.

```
<?xml version="1.0" encoding="UTF-8" standalone="no"?>
<job>
  <clean_up>1</clean_up>
  <info>BP-TZVPD-FINE-COSMO+GAS_new</info>
  <molecule_set_in>input_set.xml</molecule_set_in>
  <molecule_set_out extractable="join">output_set.xml</molecule_set_out>
  <job_schedule>
    <step>
      <number>1</number>
      <info>conf. creation using balloon</info>
      <method>BALLOON_CONF_GEN</method>
      <status>ready</status>
    </step>
    <step>
      <number>2</number>
      <info>consistency check (using UNIQUECODE)</info>
      <method>REDUCE_TO_UNIQUECODE</method>
      <status>ready</status>
      <reference>input_set.xml</reference>
```

```
130     </step>
131     <step>
132       <number>3</number>
133       <info>sort w.r.t. energy</info>
134       <method>SORT_BY_E</method>
135       <status>ready</status>
136     </step>
137     <step>
138       <number>4</number>
139       <info>geodis clustering (thr1=1)</info>
140       <method>CLUSTER_GEODIS</method>
141       <status>ready</status>
142       <geodis_threshold1>1</geodis_threshold1>
143       <geodis_threshold2>1</geodis_threshold2>
144     </step>
145     <step>
146       <number>5</number>
147       <info>reduce to 150 structures</info>
148       <method>REDUCE_BY_E_MAX</method>
149       <status>ready</status>
150       <energy_window>2000</energy_window>
151       <n_max>150</n_max>
152     </step>
153     <step>
154       <number>6</number>
155       <info>BP/SV(P) cosmo single points</info>
```

```
156       <method>BP-SV_P-COSMO-SP</method>
157       <status>ready</status>
158     </step>
159     <step>
160       <number>7</number>
161       <info>reduce to 50 structures or 25 kcal/mol energy window</info>
162       <method>REDUCE_BY_E_MAX</method>
163       <status>ready</status>
164       <energy_window>25</energy_window>
165       <n_max>50</n_max>
166     </step>
167     <step>
168       <number>8</number>
169       <info>BP/SV(P) opt. with loose settings</info>
170       <method>BP-SV_P-COSMO-LOOSE</method>
171       <status>ready</status>
172     </step>
173     <step>
174       <number>9</number>
175       <info>sort w.r.t. energy</info>
176       <method>SORT_BY_E</method>
177       <status>ready</status>
178     </step>
179     <step>
180       <number>10</number>
181       <info>standard geodis clustering (thr1=1)</info>
```

```
182       <method>CLUSTER_GEODIS</method>
183       <status>ready</status>
184       <geodis_threshold1>1</geodis_threshold1>
185       <geodis_threshold2>1</geodis_threshold2>
186     </step>
187     <step>
188       <number>11</number>
189       <info>reduce to 25.0 kcal/mol window or max. 50 structures</info>
190       <method>REDUCE_BY_E_MAX</method>
191       <status>ready</status>
192       <energy_window>25</energy_window>
193       <n_max>12</n_max>
194     </step>
195     <step>
196       <number>12</number>
197       <info>BP/TZVP COSMO optimization</info>
198       <method>BP-TZVP-COSMO</method>
199       <status>ready</status>
200     </step>
201     <step>
202       <number>13</number>
203       <info>sort w.r.t. energy</info>
204       <method>SORT_BY_E</method>
205       <status>ready</status>
206     </step>
207     <step>
```

```
208         <number>14</number>
209         <info>standard geodis clustering (thr=1)</info>
210         <method>CLUSTER_GEODIS</method>
211         <status>ready</status>
212         <geodis_threshold1>1</geodis_threshold1>
213         <geodis_threshold2>1</geodis_threshold2>
214       </step>
215       <step>
216         <number>15</number>
217         <info>reduce to 50 conformers or 25 kcal/mol energy window</info>
218         <method>REDUCE_BY_E_MAX</method>
219         <status>ready</status>
220         <energy_window>25</energy_window>
221         <n_max>50</n_max>
222       </step>
223       <step>
224         <number>16</number>
225         <info>BP-TZVP-COSMO conformer(s) info</info>
226         <method>PRINT_CONF_INFO</method>
227         <status>ready</status>
228       </step>
229       <step>
230         <number>17</number>
231         <molecule_set_out directory="BP-TZVP" extractable="separate">BP-TZVP-
232 COSMO.xml</molecule_set_out>
233         <info/>
234         <method>COPY_COSMO_FILE</method>
```

```
235         <status>ready</status>
236         <path>Results_of_BP-TZVP-COSMO</path>
237       </step>
238       <step>
239         <number>18</number>
240         <info>BP-TZVPD-FINE COSMO single point</info>
241         <method>BP-TZVPD-FINE-COSMO-SP</method>
242         <status>ready</status>
243       </step>
244       <step>
245         <number>19</number>
246         <info>sort w.r.t. energy</info>
247         <method>SORT_BY_E</method>
248         <status>ready</status>
249       </step>
250       <step>
251         <number>20</number>
252         <info>BP-TZVPD-FINE-COSMO conformer(s) info</info>
253         <method>PRINT_CONF_INFO</method>
254         <status>ready</status>
255       </step>
256       <step>
257         <number>21</number>
258         <molecule_set_out extractable="join">BP-TZVPD-FINE-
259 COSMO.xml</molecule_set_out>
260         <info/>
261         <method>COPY_COSMO_FILE</method>
```

```
    <status>ready</status>
    <path>Results_of_BP-TZVPD-FINE-COSMO</path>
  </step>
  <step>
    <number>22</number>
    <info>BP/TZVP GAS optimization</info>
    <method>BP-TZVP-GAS</method>
    <status>ready</status>
  </step>
  <step>
    <number>23</number>
    <info>sort w.r.t. energy</info>
    <method>SORT_BY_E</method>
    <status>ready</status>
  </step>
  <step>
    <number>24</number>
    <info>standard geodis clustering (thr=1)</info>
    <method>CLUSTER_GEODIS</method>
    <status>ready</status>
    <geodis_threshold1>1</geodis_threshold1>
    <geodis_threshold2>1</geodis_threshold2>
  </step>
  <step>
    <number>25</number>
    <info/>
```

```
288       <method>PRINT_CONF_INFO</method>
289       <status>ready</status>
290     </step>
291     <step>
292       <number>26</number>
293       <molecule_set_out directory="BP-TZVP" extractable="separate">BP-TZVP-
294 GAS.xml</molecule_set_out>
295       <info/>
296       <method>WRITE_ENERGY_FILE</method>
297       <status>ready</status>
298       <path>Results_of_BP-TZVP-COSMO</path>
299       <add_comment>METHOD=b-p;BASIS=def-TZVP;</add_comment>
300     </step>
301     <step>
302       <number>27</number>
303       <info>BP-TZVPD-FINE single point</info>
304       <method>BP-TZVPD-GAS-SP</method>
305       <status>ready</status>
306     </step>
307     <step>
308       <number>28</number>
309       <info>sort w.r.t. energy</info>
310       <method>SORT_BY_E</method>
311       <status>ready</status>
312     </step>
313     <step>
314       <number>29</number>
```

```
315         <info>BP-TZVPD-FINE conformer(s) info</info>
316         <method>PRINT_CONF_INFO</method>
317         <status>ready</status>
318       </step>
319       <step>
320         <number>30</number>
321         <info/>
322         <method>WRITE_ENERGY_FILE</method>
323         <status>ready</status>
324         <path>Results_of_BP-TZVPD-FINE-COSMO</path>
325         <add_comment>METHOD=b-p;BASIS=def2-TZVPD;</add_comment>
326       </step>
327     </job_schedule>
328 </job>
329
```