Support Information for:

An accurate prediction of hydration free energy by combination of molecular integral equations theory with structural descriptors

Ekaterina L. Ratkova,[†] Gennady N. Chuev,[†,‡] Volodymyr P. Sergiievskyi,[†] and Maxim V. Fedorov[†]

[†] The Max Planck Institute for Mathematics in the Sciences, Inselstrasse 22, Leipzig, 04103, Germany

[‡] Institute of Theoretical and Experimental Biophysics, Russian Academy of Science, Pushchino, Moscow Region, 142290, Russia

E-mail: fedorov@mis.mpg.de

Phone: +49 341 9959 756. Fax: +49 341 9959 999

PART I. Composition of the training and test sets

Table S-1. Composition of the training [(a)] and test [(b)] sets. Experimental hydration free energies[1-7] (kcal/mol) and corresponding values calculated with PW, GF methods as well as the SDC model based on these methods (kcal/mol). PLHNC closure, Lorentz-Berthelot mixing rules.

Name	ρV	N_{br}	N_{db}	N_{bz}	N_{OH}	N_{ph}	N_{Hal}	N_{eth}	N_{ald}	N_{ket}	$\Delta\mu_{hyd}^{PW}$	$\Delta\mu_{hyd}^{GF}$	$\Delta\mu_{hyd}^{SDC+PW}$	$\Delta\mu_{hyd}^{SDC+GF}$	$\Delta\mu_{hyd}^{\exp}$	
alkane																
2,2,4-trimethylpentane	4.86	3	0	0	0	0	0	0	0	0	9.79	-0.65	2.07	1.52	2.87	[a]
2,2,5-trimethylhexane	5.39	3	0	0	0	0	0	0	0	0	10.67	-0.94	2.14	1.58	2.79	[a]
2,2-dimethylbutane	4.14	2	0	0	0	0	0	0	0	0	10.22	2.20	2.52	2.53	2.57	[a]
2,4-dimethylpentane	4.65	2	0	0	0	0	0	0	0	0	11.15	1.98	2.69	2.63	2.87	[a]
2-methylbutane	3.86	1	0	0	0	0	0	0	0	0	11.25	4.34	2.90	3.10	2.38	[a]
2-methylhexane	4.88	1	0	0	0	0	0	0	0	0	12.83	3.65	2.94	3.06	2.93	[a]
n-decane	6.57	0	0	0	0	0	0	0	0	0	16.33	3.87	2.83	2.97	3.16	[a]
n-hexane	4.54	0	0	0	0	0	0	0	0	0	13.33	5.23	2.89	3.03	2.50	[a]
n-octane	5.57	0	0	0	0	0	0	0	0	0	14.96	4.67	2.96	3.13	2.89	[a]
n-pentane	4.04	0	0	0	0	0	0	0	0	0	12.47	5.47	2.79	2.96	2.36	[a]
propane	3.00	0	0	0	0	0	0	0	0	0	10.70	5.96	2.57	2.79	1.97	[a]
2,2-dimethylpentane	4.63	2	0	0	0	0	0	0	0	0	11.07	1.73	2.63	2.38	2.88	[b]
2,3,4-trimethylpentane	4.82	3	0	0	0	0	0	0	0	0	9.56	-0.45	1.90	1.70	2.56	[b]
2,3-dimethylpentane	4.57	2	0	0	0	0	0	0	0	0	10.59	1.64	2.24	2.24	2.52	[b]
2-methylpentane	4.36	1	0	0	0	0	0	0	0	0	11.95	3.93	2.84	3.01	2.52	[b]
3-methylhexane	4.81	1	0	0	0	0	0	0	0	0	12.26	3.20	2.48	2.56	2.71	[b]
3-methylpentane	4.29	1	0	0	0	0	0	0	0	0	11.37	3.47	2.36	2.51	2.51	[b]
ethane	2.45	0	0	0	0	0	0	0	0	0	9.56	5.94	2.27	2.42	1.84	[b]
methane	1.79	0	0	0	0	0	0	0	0	0	7.66	5.11	1.37	1.17	1.98	[b]
n-butane	3.47	0	0	0	0	0	0	0	0	0	11.35	5.52	2.53	2.64	2.09	[b]
n-heptane	5.04	0	0	0	0	0	0	0	0	0	14.08	4.91	2.88	3.04	2.63	[b]
n-nonane	6.05	0	0	0	0	0	0	0	0	0	15.56	4.19	2.84	2.95	3.14	[b]
alkene																
3-methylbut-1-ene	3.81	1	1	0	0	0	0	0	0	0	11.39	4.87	2.19	2.37	1.82	[a]
but-1-ene	3.38	0	1	0	0	0	0	0	0	0	11.17	5.78	1.55	1.63	1.37	[a]
ethene	2.22	0	1	0	0	0	0	0	0	0	8.41	5.21	0.55	0.31	1.28	[a]
hept-1-ene	4.93	0	1	0	0	0	0	0	0	0	13.81	5.04	1.86	1.86	1.66	[a]

hex-1-ene	4.41	0	1	0	0	0	0	0	0	0	12.94	5.27	1.77	1.77	1.64	[a]
non-1-ene	5.95	0	1	0	0	0	0	0	0	0	15.43	4.44	1.93	1.91	2.06	[a]
2-methylbut-2-ene	3.84	1	1	0	0	0	0	0	0	0	11.90	5.69	2.66	3.21	1.31	[b]
oct-1-ene	5.45	0	1	0	0	0	0	0	0	0	14.68	4.79	1.94	1.94	2.08	[b]
pent-1-ene	3.90	0	1	0	0	0	0	0	0	0	12.11	5.58	1.71	1.74	1.66	[b]
propene	2.84	0	1	0	0	0	0	0	0	0	10.06	5.81	1.27	1.30	1.29	[b]
trans-hept-2-ene	4.98	0	1	0	0	0	0	0	0	0	14.42	5.66	2.39	2.52	1.67	[b]
alkylbenzene																
ethylbenzene	4.59	1	0	1	0	0	0	0	0	0	10.59	2.35	-0.57	-0.45	-0.73	[a]
n-butylbenzene	5.59	1	0	1	0	0	0	0	0	0	12.19	1.58	-0.47	-0.60	-0.40	[a]
n-hexylbenzene	6.60	1	0	1	0	0	0	0	0	0	13.67	0.87	-0.51	-0.67	-0.04	[a]
n-pentylbenzene	6.09	1	0	1	0	0	0	0	0	0	12.87	1.21	-0.54	-0.65	-0.23	[a]
n-propylbenzene	5.11	1	0	1	0	0	0	0	0	0	11.66	2.23	-0.28	-0.25	-0.53	[a]
toluene	4.14	1	0	1	0	0	0	0	0	0	10.06	2.94	-0.40	-0.16	-0.84	[a]
1,2,3-trimethylbenzene	4.93	3	0	1	0	0	0	0	0	0	8.49	-0.71	-1.04	-0.52	-1.21	[b]
1,2,4-trimethylbenzene	5.06	3	0	1	0	0	0	0	0	0	9.52	0.23	-0.21	0.50	-0.83	[b]
1,3,5-trimethylbenzene	5.1	3	0	1	0	0	0	0	0	0	9.77	0.33	-0.01	0.62	-0.90	[b]
2-ethyltoluene	4.92	2	0	1	0	0	0	0	0	0	9.44	0.30	-1.14	-0.91	-1.04	[b]
4-ethyltoluene	5.14	2	0	1	0	0	0	0	0	0	11.01	1.62	0.10	0.55	-0.95	[b]
isobutylbenzene	5.34	2	0	1	0	0	0	0	0	0	10.21	-0.21	-1.00	-1.16	0.16	[b]
m-xylene	4.64	2	0	1	0	0	0	0	0	0	10.13	1.85	-0.02	0.47	-0.82	[b]
o-xylene	4.53	2	0	1	0	0	0	0	0	0	9.31	1.17	-0.69	-0.29	-0.90	[b]
p-xylene	4.69	2	0	1	0	0	0	0	0	0	10.55	2.27	0.32	0.91	-0.80	[b]
sec-butylbenzene	5.35	2	0	1	0	0	0	0	0	0	10.46	0.02	-0.76	-0.92	-0.45	[b]
tert-butylbenzene	5.15	3	0	1	0	0	0	0	0	0	9.13	-1.38	-0.73	-1.06	-0.44	[b]
alcohols																
2-methylbutan-1-ol	3.85	1	0	0	1	0	0	0	0	0	2.84	-6.87	-4.75	-4.79	-4.42	[a]
3-methylbutan-1-ol	3.88	1	0	0	1	0	0	0	0	0	3.74	-6.03	-3.90	-3.94	-4.42	[a]
4-methylpentan-2-ol	4.29	2	0	0	1	0	0	0	0	0	2.86	-7.92	-4.33	-4.17	-3.74	[a]
butan-1-ol	3.54	0	0	0	1	0	0	0	0	0	3.86	-4.88	-4.33	-4.38	-4.72	[a]
butan-2-ol	3.46	1	0	0	1	0	0	0	0	0	2.62	-6.06	-4.39	-4.23	-4.60	[a]
decan-1-ol	6.60	0	0	0	1	0	0	0	0	0	8.76	-6.62	-4.07	-4.17	-3.64	[a]
ethanol	2.47	0	0	0	1	0	0	0	0	0	1.63	-4.82	-4.95	-5.00	-5.00	[a]
heptan-1-ol	5.09	0	0	0	1	0	0	0	0	0	6.50	-5.56	-4.04	-4.08	-4.23	[a]

2-methylbutan-2-ol	3.78	2	0	0	1	0	0	0	0	0	2.52	-7.15	-3.91	-3.73	-4.43	[b]
2-methylpentan-2-ol	4.28	2	0	0	1	0	0	0	0	0	3.31	-7.50	-3.87	-3.76	-3.93	[b]
2-methylpentan-3-ol	4.26	2	0	0	1	0	0	0	0	0	3.57	-7.13	-3.59	-3.40	-3.89	[b]
2-methylpropan-1-ol	3.43	1	0	0	1	0	0	0	0	0	2.88	-5.72	-4.08	-3.91	-4.51	[b]
hexan-1-ol	4.57	0	0	0	1	0	0	0	0	0	5.69	-5.27	-4.07	-4.12	-4.39	[b]
hexan-3-ol	4.48	1	0	0	1	0	0	0	0	0	4.64	-6.33	-3.91	-3.85	-4.07	[b]
methanol	1.88	0	0	0	1	0	0	0	0	0	-0.07	-5.46	-5.76	-6.02	-5.11	[b]
nonan-1-ol	6.10	0	0	0	1	0	0	0	0	0	8.02	-6.24	-4.05	-4.12	-3.89	[b]
octan-1-ol	5.60	0	0	0	1	0	0	0	0	0	7.26	-5.88	-4.05	-4.08	-4.09	[b]
pentan-1-ol	4.05	0	0	0	1	0	0	0	0	0	4.71	-5.09	-4.25	-4.28	-4.52	[b]
pentan-2-ol	3.97	1	0	0	1	0	0	0	0	0	3.51	-6.29	-4.26	-4.13	-4.39	[b]
pentan-3-ol	3.93	1	0	0	1	0	0	0	0	0	3.23	-6.65	-4.49	-4.52	-4.35	[b]
propan-1-ol	3.02	0	0	0	1	0	0	0	0	0	3.10	-4.45	-4.32	-4.28	-4.83	[b]
propan-2-ol	2.97	1	0	0	1	0	0	0	0	0	2.29	-5.25	-3.97	-3.73	-4.76	[b]
phenols																
3,5-dimethylphenol	4.63	3	0	1	1	1	0	0	0	0	4.16	-6.42	-5.70	-5.45	-6.27	[a]
3-ethylphenol	4.59	2	0	1	1	1	0	0	0	0	4.45	-6.05	-6.42	-6.49	-6.26	[a]
4-ethylphenol	4.61	2	0	1	1	1	0	0	0	0	4.59	-5.91	-6.30	-6.34	-6.14	[a]
p-cresol	4.15	2	0	1	1	1	0	0	0	0	4.12	-5.29	-6.09	-6.01	-6.14	[a]
phenol	3.61	1	0	1	1	1	0	0	0	0	3.53	-4.68	-6.92	-7.13	-6.61	[a]
2,3-dimethylphenol	4.51	3	0	1	1	1	0	0	0	0	4.15	-6.24	-5.54	-5.34	-6.16	[b]
2,4-dimethylphenol	4.64	3	0	1	1	1	0	0	0	0	4.88	-5.62	-4.99	-4.64	-6.01	[b]
2,5-dimethylphenol	4.65	3	0	1	1	1	0	0	0	0	4.06	-6.35	-5.82	-5.37	-5.92	[b]
2,6-dimethylphenol	4.55	3	0	1	1	1	0	0	0	0	4.08	-6.24	-5.66	-5.32	-5.27	[b]
3,4-dimethylphenol	4.55	3	0	1	1	1	0	0	0	0	3.47	-6.99	-6.27	-6.07	-6.51	[b]
4-n-propylphenol	5.12	2	0	1	1	1	0	0	0	0	5.68	-6.04	-5.99	-6.14	-5.91	[b]
4-tert-butylphenol	5.15	4	0	1	1	1	0	0	0	0	3.25	-9.57	-6.33	-6.88	-5.95	[b]
o-cresol	4.12	2	0	1	1	1	0	0	0	0	4.63	-4.71	-5.53	-5.45	-5.88	[b]
chloroalkanes																
1-chlorobutane	3.93	0	0	0	0	0	1	0	0	0	11.32	4.52	-0.31	-0.45	-0.16	[a]
1-chloroheptane	5.48	0	0	0	0	0	1	0	0	0	13.92	3.80	-0.05	-0.19	0.29	[a]
1-chlorohexane	4.97	0	0	0	0	0	1	0	0	0	13.14	4.13	-0.06	-0.18	0.00	[a]
1-chloropentane	4.45	0	0	0	0	0	1	0	0	0	12.20	4.32	-0.21	-0.32	-0.07	[a]
1-chloropropane	3.42	0	0	0	0	0	1	0	0	0	10.48	4.88	-0.37	-0.42	-0.30	[a]

2-chloro-2-methylpropane	3.78	2	0	0	0	0	1	0	0	0	10.83	4.31	1.57	2.02	1.09	[a]
2-chlorobutane	3.85	1	0	0	0	0	1	0	0	0	10.98	4.38	0.54	0.74	0.00	[a]
2-chloropropane	3.37	1	0	0	0	0	1	0	0	0	10.28	4.80	0.58	0.86	-0.25	[a]
chloroethane	2.9	0	0	0	0	0	1	0	0	0	9.29	4.77	-0.78	-0.86	-0.63	[a]
chloromethane	2.35	0	0	0	0	0	1	0	0	0	7.74	4.20	-1.49	-1.78	-0.54	[a]
aldehydes																
acetaldehyde	2.37	0	0	0	0	0	0	0	1	0	4.65	0.15	-3.42	-3.50	-3.51	[a]
butyraldehyde	3.45	0	0	0	0	0	0	0	1	0	6.81	0.12	-2.90	-2.83	-3.18	[a]
formaldehyde	1.81	0	0	0	0	0	0	0	1	0	2.75	-0.88	-4.48	-4.88	-2.76	[a]
isobutyraldehyde	3.40	1	0	0	0	0	0	0	1	0	6.60	-0.02	-1.95	-1.62	-2.86	[a]
pentanal	3.97	0	0	0	0	0	0	0	1	0	7.68	-0.17	-2.81	-2.79	-3.03	[a]
propionaldehyde	2.93	0	0	0	0	0	0	0	1	0	5.71	0.14	-3.21	-3.15	-3.44	[a]
heptanal	5.00	0	0	0	0	0	0	0	1	0	9.42	-0.62	-2.64	-2.59	-2.67	[b]
hexanal	4.49	0	0	0	0	0	0	0	1	0	8.51	-0.41	-2.76	-2.71	-2.81	[b]
nonanal	6.02	0	0	0	0	0	0	0	1	0	10.96	-1.30	-2.62	-2.62	-2.07	[b]
octanal	5.52	0	0	0	0	0	0	0	1	0	10.22	-0.95	-2.61	-2.59	-2.29	[b]
ketones																
3-methylbutan-2-one'	3.79	2	0	0	0	0	0	0	0	1	6.54	-1.04	-3.07	-2.9549	-3.25	[a]
butanone	3.41	1	0	0	0	0	0	0	0	1	6.39	-0.22	-3.72	-2.95	-3.71	[a]
hexan-2-one	4.46	1	0	0	0	0	0	0	0	1	8.43	-0.43	-3.26	-3.76	-3.29	[a]
pentan-2-one	3.93	1	0	0	0	0	0	0	0	1	7.43	-0.26	-3.46	-3.31	-3.52	[a]
pentan-3-one	3.96	1	0	0	0	0	0	0	0	1	7.23	-0.44	-3.71	-3.48	-3.41	[a]
propanone	2.87	1	0	0	0	0	0	0	0	1	5.55	0.06	-3.74	-3.63	-3.81	[a]
4-methylpentan-2-one	4.28	2	0	0	0	0	0	0	0	1	7.16	-1.64	-3.19	-3.24	-3.05	[b]
decan-2-one	6.48	1	0	0	0	0	0	0	0	1	11.55	-1.75	-3.20	-3.34	-2.34	[b]
heptan-2-one	4.96	1	0	0	0	0	0	0	0	1	9.16	-0.77	-3.29	-3.34	-3.04	[b]
nonan-2-one	5.98	1	0	0	0	0	0	0	0	1	10.83	-1.36	-3.16	-3.27	-2.50	[b]
nonan-5-one	6.02	1	0	0	0	0	0	0	0	1	11.09	-1.13	-2.96	-3.02	-2.65	[b]
octan-2-one	5.48	1	0	0	0	0	0	0	0	1	9.98	-1.06	-3.24	-3.30	-2.88	[b]
undecan-2-one	6.99	1	0	0	0	0	0	0	0	1	12.26	-2.12	-3.25	-3.40	-2.16	[b]
ethers																
di-n-butyl ether	5.91	0	0	0	0	0	0	1	0	0	13.43	1.86	-0.78	-0.84	-0.83	[a]
di-n-propyl ether	4.93	0	0	0	0	0	0	1	0	0	12.26	3.06	-0.47	-0.26	-1.16	[a]
diethyl ether	3.85	0	0	0	0	0	0	1	0	0	9.38	2.36	-1.72	-1.65	-1.60	[a]

diisopropyl ether	4.56	2	0	0	0	0	0	1	0	0	9.04	-0.11	-1.00	-0.89	-0.53	[a]
dimethyl ether	2.69	0	0	0	0	0	0	1	0	0	6.79	1.98	-2.55	-2.76	-1.90	[a]
methyl tert-butylether	3.97	2	0	0	0	0	0	1	0	0	7.61	-0.37	-1.54	-1.52	-2.21	[a]
methylethyl ether	3.26	0	0	0	0	0	0	1	0	0	8.03	2.08	-2.18	-2.30	-2.01	[a]
polyfragment solutes																
1,1,1,2-tetrachloroethane	4.29	2	0	0	0	0	4	0	0	0	12.92	6.27	-3.42	-2.86	-1.28	[b]
1,1,1-trichloroethane	3.86	2	0	0	0	0	3	0	0	0	11.71	5.53	-1.89	-1.49	-0.19	[b]
1,1,2,2-tetrachloroethane	4.19	2	0	0	0	0	4	0	0	0	12.11	5.46	-4.07	-3.74	-2.47	[b]
1,1,2-trichloroethane	3.77	1	0	0	0	0	3	0	0	0	8.99	2.90	-5.52	-5.57	-1.99	[b]
1,1-dichloroethane	3.35	1	0	0	0	0	2	0	0	0	10.15	4.93	-1.64	-1.41	-0.85	[b]
1,1-dichloroethene	3.20	1	1	0	0	0	2	0	0	0	11.44	6.92	-1.03	-0.75	0.25	[b]
1,2,3,4-tetrachlorobenzene	5.13	4	0	1	0	0	4	0	0	0	11.88	3.66	-5.29	-4.19	-1.34	[b]
1,2,3,5-tetrachlorobenzene	5.14	4	0	1	0	0	4	0	0	0	12.37	4.07	-4.81	-3.78	-1.62	[b]
1,2,3-trichlorobenzene	4.75	3	0	1	0	0	3	0	0	0	12.09	4.60	-3.48	-2.50	-1.24	[b]
1,2,4,5-tetrachlorobenzene	5.17	4	0	1	0	0	4	0	0	0	12.59	4.26	-4.64	-3.57	-1.34	[b]
1,2,4-trichlorobenzene	4.83	3	0	1	0	0	3	0	0	0	12.90	5.31	-2.78	-1.74	-1.12	[b]
1,2-dichlorobenzene	4.34	2	0	1	0	0	2	0	0	0	11.77	4.95	-2.14	-1.41	-1.41	[b]
1,2-dichloroethane	3.34	0	0	0	0	0	2	0	0	0	10.03	4.66	-2.81	-3.07	-1.77	[b]
1,2-dichloropropane	3.78	1	0	0	0	0	2	0	0	0	9.67	3.25	-2.76	-2.82	-1.27	[b]
1,2-dimethoxyethane	4.01	0	0	0	0	0	0	2	0	0	8.10	0.09	-4.94	-5.19	-4.84	[b]
1,2-ethanediol	2.53	0	0	0	2	0	0	0	0	0	-2.83	-11.81	-8.77	-8.64	-7.75	[b]
1,3,5-trichlorobenzene	4.84	3	0	1	0	0	3	0	0	0	13.37	5.73	-2.33	-1.32	-0.78	[b]
1,3-dichlorobenzene	4.39	2	0	1	0	0	2	0	0	0	12.46	5.58	-1.51	-0.75	-0.98	[b]
1,3-dichloropropane	3.83	0	0	0	0	0	2	0	0	0	9.47	2.83	-4.11	-4.59	-1.90	[b]
1,4-dichlorobenzene	4.40	2	0	1	0	0	2	0	0	0	12.52	5.62	-1.49	-0.70	-1.01	[b]
1,4-dichloropentane	4.76	1	0	0	0	0	2	0	0	0	11.24	2.60	-2.67	-2.84	-2.32	[b]
2,3-dimethylbuta-1,3-diene	4.04	2	2	0	0	0	0	0	0	0	11.35	4.72	1.95	2.52	0.40	[b]
2-butoxyethanol	4.85	0	0	0	1	0	0	1	0	0	5.23	-6.73	-6.64	-6.78	-6.26	[b]
2-chlorophenol	3.95	2	0	1	1	1	1	0	0	0	3.35	-5.35	-8.66	-8.59	-2.82	[b]
2-chlorotoluene	4.51	2	0	1	0	0	1	0	0	0	11.94	4.54	-0.12	0.68	-1.14	[b]
2-ethoxyethanol	3.82	0	0	0	1	0	0	1	0	0	3.22	-6.47	-7.09	-7.17	-6.70	[b]
2-methoxyphenol	4.29	2	0	1	1	1	0	1	0	0	2.25	-7.71	-9.85	-9.72	-5.58	[b]
2-methylbuta-1,3-diene	3.65	1	2	0	0	0	0	0	0	0	11.01	5.25	1.13	1.42	0.68	[b]
2-methylstyrene	4.87	2	1	1	0	0	0	0	0	0	10.66	2.06	-0.77	-0.42	-1.24	[b]

2-phenylethanol	4.61	1	0	1	1	0	0	0	0	0	3.81	-7.25	-6.63	-6.72	-6.80	[b]
2-propoxyethanol	4.33	0	0	0	1	0	0	1	0	0	4.40	-6.37	-6.70	-6.75	-6.41	[b]
3-chlorophenol	4.00	2	0	1	1	1	1	0	0	0	4.95	-3.75	-7.13	-6.95	-6.61	[b]
3-hydroxybenzaldehyde	4.08	2	0	1	1	1	0	0	1	0	1.38	-8.71	-9.62	-9.55	-9.51	[b]
3-methoxyphenol	4.32	2	0	1	1	1	0	1	0	0	3.93	-6.01	-8.22	-7.99	-7.66	[b]
3-phenylpropanol	5.12	1	0	1	1	0	0	0	0	0	2.96	-9.42	-8.25	-8.57	-6.93	[b]
4-chloro-3-methylphenol	4.48	3	0	1	1	1	1	0	0	0	5.42	-4.38	-6.31	-5.89	-6.79	[b]
4-chlorophenol	4.03	2	0	1	1	1	1	0	0	0	5.08	-3.66	-7.04	-6.85	-7.04	[b]
4-hydroxybenzaldehyde	4.08	2	0	1	1	1	0	0	1	0	0.98	-9.15	-10.03	-9.98	-9.65	[b]
4-methoxyacetophenone	5.19	3	0	1	0	0	0	1	0	1	8.09	-2.29	-5.96	-5.34	-4.40	[b]
E-but-2-enal	3.37	0	1	0	0	0	0	0	1	0	7.04	0.77	-3.47	-3.47	-4.23	[b]
E-hex-2-enal	4.43	0	1	0	0	0	0	0	1	0	9.09	0.50	-3.01	-3.07	-3.68	[b]
E-oct-2-enal	5.43	0	1	0	0	0	0	0	1	0	10.58	-0.25	-3.04	-3.18	-3.44	[b]
acetophenone	4.47	2	0	1	0	0	0	0	0	1	7.59	-1.16	-4.75	-4.68	-4.54	[b]
allyl alcohol	2.88	0	1	0	1	0	0	0	0	0	1.81	-5.35	-6.31	-6.50	-5.10	[b]
benzyl alcohol	4.12	1	0	1	1	0	0	0	0	0	2.52	-7.21	-7.19	-6.99	-6.63	[b]
buta-1,3-diene	3.20	0	2	0	0	0	0	0	0	0	10.47	5.62	0.20	0.12	0.61	[b]
chlorobenzene	4.01	1	0	1	0	0	1	0	0	0	11.44	5.11	-0.93	-0.46	-1.09	[b]
cis-1,2-dichloroethene	3.12	0	1	0	0	0	2	0	0	0	9.47	4.83	-3.96	-4.28	-0.93	[b]
dichloromethane	2.84	0	0	0	0	0	2	0	0	0	8.93	4.63	-3.15	-3.43	-1.31	[b]
dimethoxymethane	3.52	0	0	0	0	0	0	2	0	0	6.66	-0.30	-5.63	-5.90	-2.97	[b]
ethyl phenyl ether	4.93	1	0	1	0	0	0	1	0	0	11.49	2.79	-1.87	-1.18	-2.22	[b]
hexa-1,5-diene	4.27	0	2	0	0	0	0	0	0	0	12.45	5.28	0.57	0.46	1.01	[b]
methyl_phenyl_ether	4.37	1	0	1	0	0	0	1	0	0	10.33	2.74	-2.19	-1.59	-2.45	[b]
penta-1,4-diene	3.76	0	2	0	0	0	0	0	0	0	11.74	5.73	0.62	0.58	0.93	[b]
pentachloroethane	4.70	3	0	0	0	0	5	0	0	0	14.01	6.77	-3.98	-3.11	-1.39	[b]
tetrachloroethene	4.08	2	1	0	0	0	4	0	0	0	13.88	8.09	-3.06	-2.41	0.10	[b]
tetrachloromethane	3.84	2	0	0	0	0	4	0	0	0	13.21	7.55	-2.45	-1.87	0.08	[b]
trans-1,2-dichloroethene	3.17	0	1	0	0	0	2	0	0	0	10.85	6.18	-2.65	-2.90	-0.78	[b]
trichloroethene	3.63	1	1	0	0	0	3	0	0	0	12.13	6.87	-3.10	-2.91	-0.44	[b]
trichloromethane	3.36	1	0	0	0	0	3	0	0	0	10.95	5.84	-2.94	-2.88	-1.08	[b]

[a] – training set; [b] – test set;

For ref. [1], [7] value of $\Delta\mu_{hyd}^{\exp}$ was recalculated from logP (water/gas) as: $\Delta\mu_{hyd}^{\exp} = -10^{-3} \ln 10 \cdot RT \log P(water/gas)$ (S-1)

PART II. Comparison of OPLS-AA and ESP partial charges for benzene, phenol, and 2-phenylethanol

To analyze the influence of substitutions in a benzene ring on its electron density distribution (and, correspondingly, the partial charges of carbon atoms) we calculated the partial charges by different methods for benzene, phenol, and 2-phenylethanol (Figure S-1). We compared the **OPLS-AA** partial charges with those calculated with the CHELPG procedure for the electrostatic potential surface **(ESP)** with using Gaussian03 at MP2/6-311G(d,p) level of theory.

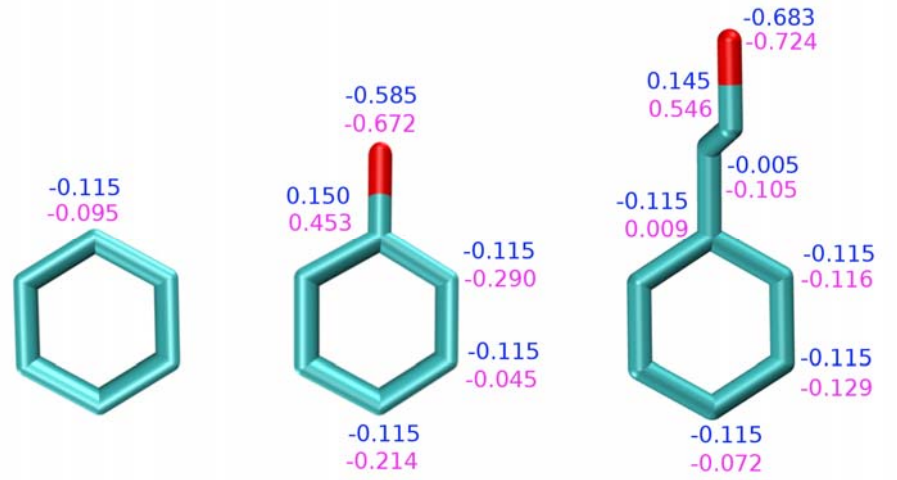

FIGURE S-1. Partial charges for heavy atoms of benzene, phenol, and benzyl alcohol. For simplicity the hydrogen atoms are not shown. There are the two numbers next to each symmetry-unique atom in the solute. The first number (in blue) is OPLS-AA charge and the second number (in magenta) is the ESP partial charge.

In the case of **benzene** each *type* of atoms has its own OPLS-AA partial charge. ESP partial charges differ for symmetric atoms in the third digit after point. For carbons the ESP charges are smaller than the corresponding OPLS-AA charges.

In the case of **phenol** the OPLS-AA partial charges take into account *only* the change of charge on the carbon atom which is closest to the OH-group (it has q = 0.150 instead of q = -0.115 for the benzene's carbon). Other five carbon atoms in phenol have the same charges as the benzene carbons. However, in reality, the partial charges of these carbon atoms also *change* because of the electron density redistribution. As a result, *meta*-, *orto*-, and *para*-positions in phenol ring become *distinguishable*. EPS charges are sensitive to the electron density redistribution. Thus, *orto*-, and *para*-positions in the phenol have more negative charges (about -0.25) than *meta*—positions (about -0.05). Transfer of the electron density from carbon atom to oxygen is fixed with both types of partial charges but it is more noticeable for the ESP ones.

In the case of **2-phenylethanol** there is a carbon *spacer* between the benzene ring and OH-group. As one can see (Figure S-1) 2-phenylethanol has the same OPLS-AA parameters of carbons in the ring as benzene. Thus, OPLS-AA partial charges reflect attenuation of the inductive effect by the carbon spacer. The ESP partial charges of 2-phenylethanol are similar to the corresponding OPLS-AA ones. The main difference is observed for the carbon spacer. As in the case of phenol, ESP charges are more sensitive to the electron transfer between the oxygen of hydroxyl group and the nearest carbon atom.

The comparison allows us to conclude that the OPLS-AA partial charges are good approximation to describe properties of benzene and 2-phenylethanol but *do not* describe phenol properties in a proper way.

PART III. Optimal set of descriptors of the SDC model

Values of coefficients of the SDC model with the considered set of descriptors were obtained by the multiple linear regression analysis with the training set of solutes. The main idea is the following. Using the SDC model, the errors of the RISM HFE expressions might be expressed as a linear combination of descriptors contributions (Eq. S-2). We vary coefficients a_0^{model},…, a_{10}^{model} to obtain the best agreement between the SDC model and the experimental measurements (in other words, to obtain minimum of the error ε):

$$\varepsilon = \Delta\mu_{hyd}^{\exp} - \Delta\mu_{hyd}^{\text{model}} \approx a_0^{\text{model}} + \sum_{i=1}^{10} a_i^{model} D_i \,, \qquad \text{(S-2)}$$

where $\Delta\mu_{hyd}^{\exp}$ is an experimental HFE, $\Delta\mu_{hyd}^{model}$ is the value of HFE obtained via the RISM calculations, the superscribe *model* denotes the HFE expression used within the RISM approach (e.g PW, GF), a_0^{model},…, a_{10}^{model} are coefficients, $D_1,\ldots,D_{10}$ are descriptors of the SDC model.

For this purpose we apply the standard least square method.[8] To obtain an *optimal combination of descriptors* we analyzed influence of each term of Eq. (S-2) on the correlation coefficient of multiple regression R^2 (see Table S-2). (The analysis was done for HFEs obtained with PW method and PLHNC closure).

Table S-2. Correlation coefficients of multiple linear regressions for different sets of descriptors

Set of descriptors	$\mathbf{R^2}$	$\mathbf{R^2_{total}-R^2_{current}}$
Total	0.947	0.000
without a_0^{model}	0.852	0.096
without $a_1^{model} D_1$	0.466	0.481
without $a_2^{model} D_2$	0.795	0.153
without $a_3^{model} D_3$	0.935	0.012
without $a_4^{model} D_4$	0.904	0.043
without $a_5^{model} D_5$	0.938	0.010
without $a_6^{model} D_6$	0.933	0.014
without $a_7^{\text{model}} D_7$	0.866	0.081
without $a_8^{\text{model}} D_8$	0.901	0.047
without $a_9^{\text{model}} D_9$	0.937	0.010
without $a_{10}^{\text{model}} D_{10}$	0.861	0.087

In the present work we chose the following criteria for the selection of descriptors. If the difference between correlation coefficients for the total set of descriptors and a reduced set (without one descriptor) is less than 0.005 then the descriptor may be excluded. As one can see (Table S-2) differences for all considered sets of descriptors are bigger that this criteria. Thus, final equation of the SDC model contains *all* eleven contributions.

PART IV. SDC hydration free energy calculations with GF expression

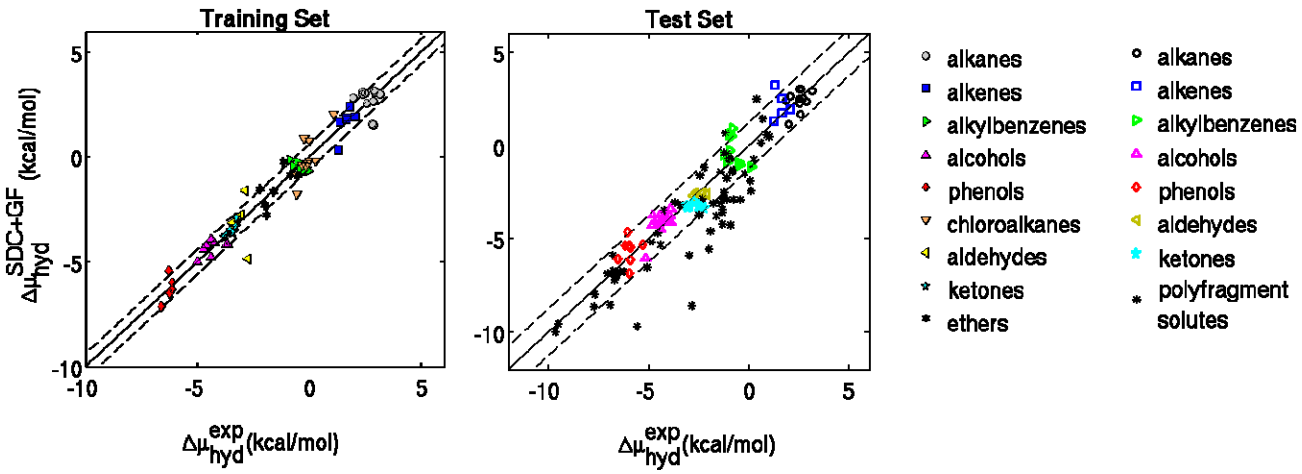

FIGURE S-2. HFEs calculated with the SDC model for the GF method versus the experimental values for training and test sets of solutes. Dashed lines indicate std of the HFEs.

In parallel with the PW data correction by the SDC model (see the main text of the paper) we perform correction of the HFEs calculated with the GF expression. Correlation coefficients were also obtained via the multiple linear regression analysis. One can find corresponding regression parameters in Table S-3. It is known that the GF method overestimates the specific interactions.[9,10] Indeed, the coefficient of the OH-group correction for the GF method ($a_5^{GF} = 3.33$ kcal/mol) is bigger than that for the PW method ($a_5^{PW} = 0.73$ kcal/mol). We obtained the mean value of the difference $\Delta\mu_{hyd}^{SDC+GF} - \Delta\mu_{hyd}^{\exp}$ to be -0.38 kcal/mol, (*std* = 1.26 kcal/mol) for the test set of solutes. Thus, the estimation of mean value of $\Delta\mu_{hyd}^{SDC+GF} - \Delta\mu_{hyd}^{\exp}$ is slightly biased with respect to zero value, and the accuracy of the predicted HFEs depends only on the standard deviation of calculated data.

It was found that for "simple" solutes the performance of the SDC model for PW HFE expression *is more efficient* for prediction of HFEs for small organic solutes. However, predictabilities of the models for polyfragment solutes are comparable.

Table S-3. Descriptors of the SDC model and the corresponding parameters of multiple linear regressions for the GF expression

Descriptor		**Coefficient (kcal/mol)**	**Descriptor**		**Coefficient (kcal/mol)**
Dimensionless partial molar volume	ρV	$a_1^{GF} = 0.64$	Number of phenol fragments	N_{ph}	$a_6^{GF} = -2.35$
Number of branches	N_{br}	$a_2^{GF} = 1.39$	Number of halogen atoms	N_{Hal}	$a_7^{GF} = -2.39$
Number of double bonds	N_{db}	$a_3^{GF} = -1.23$	Number of ether groups	N_{eth}	$a_8^{GF} = -1.38$
Number of benzene rings	N_{bz}	$a_4^{GF} = -2.04$	Number of aldehyde groups	N_{ald}	$a_9^{GF} = -0.07$
Number of OH-groups	N_{OH}	$a_5^{GF} = 3.33$	Number of ketone groups	N_{ket}	$a_{10}^{GF} = -2.02$

Correlation coefficient of the multiple regression (R^2) = 0.9773

PART V. SDC hydration free energy calculations with HNC closure

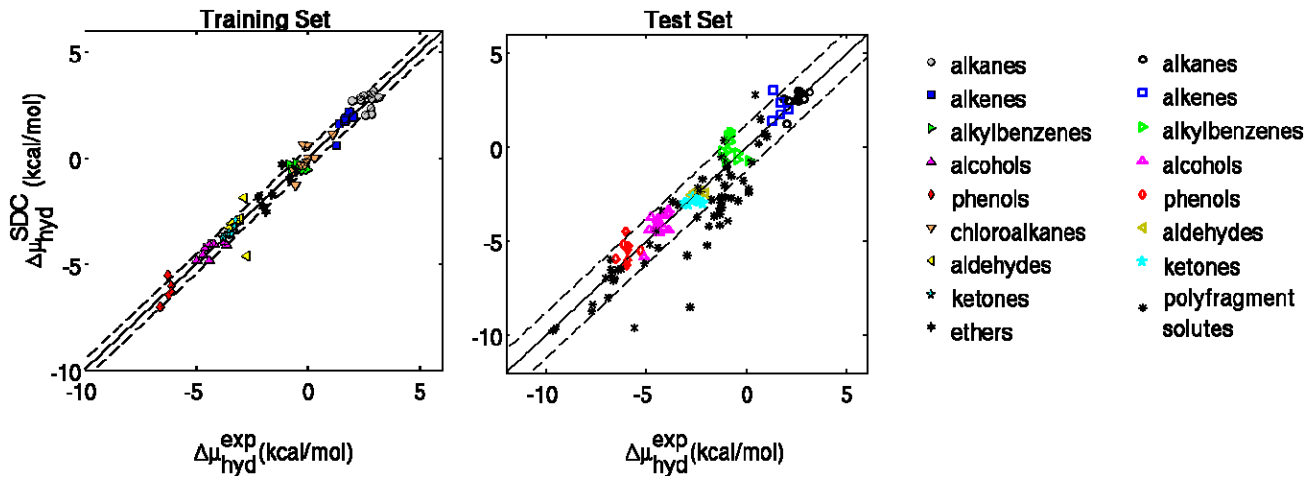

FIGURE S-3. HFEs calculated with the SDC model for the PW method versus the experimental values for training and test sets of solutes. Dashed lines indicate std of the HFE. HNC closure.

HFEs calculations can be performed with different closures (e.g. PLHNC, HNC). We analyzed the influence of the choice of closure on the results obtained with the SDC model. In parallel with the RISM/PLHNC data correction with the SDC model (see the present paper) we perform corrections of the HFEs calculated with RISM/HNC method.

As one can see, correlation coefficients of the multiple regression of the SDC model combined with the RISM/HNC method are bigger that the corresponding coefficient for the RISM/PLHNC method. In other words, HFEs obtained by the RISM/HNC method are *more* overestimated with respect to the experimental data than that calculated with RISM/PLHNC method. The difference between predicted and experimental data for RISM/HNC method equals -0.30 kcal/mol and the standard deviation equals 1.22 kcal/mol.

Table S-4. Descriptors of the SDC model and the corresponding parameters of multiple linear regressions for PW method. HNC closure.

Descriptor		**Coefficient (kcal/mol)**	**Descriptor**		**Coefficient (kcal/mol)**
Dimensionless partial molar volume	ρV	$a_1^{PW} = -1.56$	Number of phenol fragments	N_{ph}	$a_6^{PW} = -1.88$
Number of branches	N_{br}	$a_2^{PW} = 1.48$	Number of halogen atoms	N_{Hal}	$a_7^{PW} = -2.03$
Number of double bonds	N_{db}	$a_3^{PW} = -1.03$	Number of ether groups	N_{eth}	$a_8^{PW} = -1.78$
Number of benzene rings	N_{bz}	$a_4^{PW} = -2.25$	Number of aldehyde groups	N_{ald}	$a_9^{PW} = -0.77$
Number of OH-groups	N_{OH}	$a_5^{PW} = 0.97$	Number of ketone groups	N_{ket}	$a_{10}^{PW} = -2.76$

Correlation coefficient of the multiple regression (R^2) = 0.9598

However, the results for the test set of solutes were obtained with smaller λ coupling (see Table S-5) where λ_{coup} is the coupling iteration parameter[11] in the calculations. Calculations for OPLS-AA mixing rules can be performed with PLHNC closure only with low speed (very small λ_{coup}), but for the HNC closure calculations did not converge at all.

Table S-5. Convergence of RISM calculations for training and test sets with different mixing rules, closure relations, and coupling parameter (λ_{coup})

Set	mixing rules / closure	Lotentz-Berthelot	OPLS-AA
Training Set	HNC	converged with $\lambda_{coup} = 0.5$	does not converged
	PLHNC	converged with $\lambda_{coup} = 0.5$	converged with $\lambda_{coup} = 0.01$
Test Set	HNC	converged with $\lambda_{coup} = 0.2$	does not converged
	PLHNC	converged with $\lambda_{coup} = 0.5$	converged with $\lambda_{coup} = 0.01$

PART VI. 3D distributions

Three-dimensional water distribution $G_0(\mathbf{r})$ was built as the weighted superposition of the solute-water oxygen RDF functions:

$$G_0(\mathbf{r}) = \frac{\sum_s w_s(\mathbf{r}) g_{so}(\mathbf{r})}{\sum_s w_s(\mathbf{r})} \qquad (1)$$

where $\mathbf{r_s}$ is the position of the solute site s, $w_s(\mathbf{r}) = \left(|\mathbf{r} - \mathbf{r_s}|^6 + 1\right)^{-1}$. To treat the excluded volume effect correctly, we used the excluded volume indicator function $\sigma_{\text{ind}}(\mathbf{r})$ which is almost zero when at least one of the one dimensional RFDs is zero and almost unity otherwise:

$$\sigma_{\text{ind}}(\mathbf{r}) = \min_s \{\text{erf}(100 \cdot g_{so}(|\mathbf{r} - \mathbf{r_s}|))\} \qquad (2)$$

where $\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$ is the Gauss error function. The final three-dimensional water distribution $G(\mathbf{r})$ is obtained as:

$$G(\mathbf{r}) = \sigma_{\text{ind}}(\mathbf{r}) \cdot G_0(\mathbf{r}) \qquad (3)$$

Because the RISM treat hydrogen bonds inconsistentely, to obtain the realible 3D distribution Lennerd-Jones parameters of Oxygen and Hydrogen in OH groups were changed: σ_O = 2 Å, ε_O = 0.17 kcal/mol, σ_H =0.8 Å, ε_H = 0.03 kcal/mol

Support Information References:

(1) Abraham, M. A.; Whiting, G. S. *J. Chem. Soc. Perkin Trans 2* **1994**, 1777-1791.
(2) Cabani, S.; Gianni, P.; Mollica, V. *J. Sol. Chem.* **1981**, *10*, 563-595.
(3) Ben-Naim, A.; Marcus, Y. *J. Chem. Phys.* **1984**, *81*, 2016-2027.
(4) W.L, J.; Ulmschneider, J.P., J. W. *J. Phys. Chem. B* **2004**, *108*, 16264-16270.
(5) Cramer, C. J.; Truhlar, D. G. *JCAMD* **1992**, *6*, 629-666.
(6) Duffy, E. M.; Jorgensen, W. L. *JACS* **2000**, *122*, 2878-2888.
(7) McDonald, N. A.; W.L, J. *J. Phys. Org. Chem.* **1997**, *10*, 563-576.
(8) Barlow, R.J. *Statistics: A Guide to the Use of Statistical Methods in the Physical Sciences*, Reprint ed.: Wiley, 1989.
(9) Lee, P. H.; Maggiora, G. M. *J. Chem. Phys.* **1993,** *97*, 10175-10185
(10) Sato, K.; Chuman, H.; Ten-no, S. *J. Phys. Chem. B*, **2005,** *109*, 17290-17295.
(11) Fedorov, M.V., Chuev, G.N., Flad, H.-J., Grasedyck, L., Khoromskij, B.N. *Computing* **2007**, *80***,** 47-73.