12 Supplementary Information

12.1 Detailed methods

12.1.1 Physical reference calculations - Direct Transfer Free Energy Approach

log *P* can be estimated directly from the transfer free energy of a solute moving from the organic to the aqueous layer. Specifically, we calculate the transfer free energy from the difference in solvation free energy into octanol and hydration free energy. log *P* is directly proportional to the difference between the solvation free energy for the solute into each solvent

$$\log P = -\frac{\Delta G_{transfer}}{RT \ln 10} = \frac{\Delta G_{solvation} - \Delta G_{hydration}}{RT \ln 10} \tag{4}$$

where $\Delta G_{transfer}$ is the transfer free energy, $\Delta G_{solvation}$ is the solvation free energy of the solute going from the gas to the octanol phase, $\Delta G_{hydration}$ is the hydration free energy going from the gas to the water phase, R is the gas constant (8.314 J / mol · K) and T is the temperature (298.15 K).

The direct transfer free energy (DFE) protocol that was used for the physical reference calculations (*REF01*-REF08, *EXT02*, *EXT05*, *EXT07*, *EXT08*) directly computed the transfer free energy between solvents without any gas phase calculation, whereas the IFE protocol (discussed in Section 12.1.2) computed gas-to-solution phase solvation free energies in water and octanol separately and then subtracted to obtain the transfer free energy.

To explore how solvent mixing would effect predicted values, water was included in the octanol phase for the majority of the reference calculations. A portion of the calculations treated the octanol and water phase as completely immiscible for comparison. The experimental mole fraction of water in octanol was measured as 0.2705 [58]. The solutions and calculations modeled each phase at infinite dilution, with only a single solute molecule in each solvent.

The initial input files were made using the Solvation Toolkit (https://github.com/MobleyLab/SolvationToolkit), which converts SMILES strings to parameterized molecules and builds topology and coordinate files for use in molecular dynamics software packages. Solvation Toolkit is a driver utility that utilizes the OpenEye toolkits (version 2018.10.1) for cheminformatics (specifically file conversion and handling of molecular identities), and OEChem for reading and writing files. AmberTools [111] was used to parameterize systems with the General AMBER Force Field for organic molecules (version 2017.1.81) and water was parameterized with the TIP3P water model, AM1-BCC charges were assigned via Antechamber, Packmol (version 18.169) [112] was used to build boxes, and lastly AMBER topology and coordinate files were made with LEaP.

The SMILES string and the mole fraction of each compound in the system were used as input. The “wet” octanol systems were generated using a mole fraction of 0.7295 for octanol and 0.2705 for water, producing systems with about 200 octanol molecules and 74 water molecules, depending on the solute size. The “dry” octanol systems had no water component and about 211 octanol molecules. All of the water systems had 1497 molecules. The box dimensions were about 40x40 Å in all cases.

The following equilibration stages were carried out using the GAFF forcefield (version 2017.1.81), the TIP3P water model and OpenMM (version 7.3.1) [66, 113], a molecular simulation toolkit.

For minimization, an energy tolerance of 10 kilojoules/mole was used and the systems were minimized until convergence was reached. A Langevin integrator was used with a 0.5 fs timestep. Minimization was followed by 100 ps of NVT using a Langevin integrator and 1.0 fs timestep, 100 ps of NPT using a Langevin integrator and 2.0 fs timestep, and lastly 500 ns of NPT using a Langevin integrator and 2.0 fs timestep.

Three independent equilibrations were run starting from water and octanol phase systems of the initial setup, in order to obtain three different sets of starting coordinates for replicate transfer free energy calculations with YANK (version 0.24.0 [63]). The protocol for creating systems with different force field and/or water model conditions (1) is detailed below.

Following equilibration, the resulting systems were saved to PDBs. For each solvent system, a ParmEd Structure was created using the topology and positions from the equilibrated PDB, with parameters coming from the original GAFF/TIP3P OpenMM System. The ParmEd structure of the system was split into individual components or structures and then used to create newly parameterized OpenMM systems. The water was parameterized with either the TIP3P, TIP3P-FB or OPC water model, and the solute and solvent were parameterized with the SMIRNOFF force field (smirnoff99Frosst version 1.0.7) or remained parameterized with GAFF. In just the OPC case, a dummy atom was added to the water component structure. After parameterization, the OpenMM systems of the solute-octanol and water were converted back to ParmEd structures which maintained their new parameters. The final OpenMM system was created using the particle mesh Ewald (PME) method for periodic boundary conditions, an error tolerance of 1e-4 and a cutoff for nonbonded interactions was set to 11 Å.

The resulting OpenMM Systems were saved as XMLs for use later on. Prior to using YANK, the new systems were briefly equilibrated using the same setup described previously, excluding the 500 ns of NPT. The final equilibrated PDB and system XML

files were used as input files for solvation and transfer free energy calculations with YANK [61], a toolkit that uses Hamiltonian replica exchange and can compute solvation free energies. For the YANK simulations, hydrogen mass repartitioning (HMR) was used to allow a 3 fs timestep. HMR works by slowing down the fastest motions in the simulation by reallocating mass from the connected heavy atom to the hydrogens [114]. The temperature was set to 298.15 K (the experimental temperature), the pressure to 1.0 atm, and an anisotropic dispersion cutoff of 12.0 Å was used. There were 5000 iterations total and 335 steps per iteration. The overall length of the YANK simulations were 5 ns for each replica.

In the octanol and water phase the electrostatic interactions of the solute with the solvent were scaled off through a λ (lambda) parameter using the following lambda values where λ = [1.00, 0.75, 0.50, 0.25, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00], and steric interactions were scaled using λ = [1.00, 1.00, 1.00, 1.00, 1.00, 0.95, 0.90, 0.80, 0.70, 0.60, 0.50, 0.40, 0.35, 0.30, 0.25, 0.20, 0.15, 0.10, 0.05, 0.00].

The direct transfer free energy was obtained from YANK [61], where the $\Delta G_{transfer}$ was equivalent to $\Delta G_{octanol} - \Delta G_{water}$. This was then converted to log *P* using Equation 4. The uncertainties of the log *P* predictions were calculated as the standard error of the mean (SEM) of three replicate predictions. The SEM was estimated as $SEM = \sigma/\sqrt{N}$ where σ is the sample standard deviation and N is the size of the sample (in this case the number of replicate predictions made). The model uncertainty was reported as 1.6 log units, based on similar previous work [103].

12.1.2 Physical Reference Calculations - Indirect Solvation-Based Transfer Free Energy Approach

We ran an additional set of reference calculations using a more traditional indirect solvation-based transfer free energy method to see how it would compare to the direct transfer free energy method (described in Section 12.1.1). Specifically, the IFE protocol calculates the transfer free energy as the difference between the solvation free energy of the solute going from the gas to the octanol phase, and the hydration free energy going from the gas to the water phase. The direct transfer free energy method that was run with YANK had not computed gas-to-water transfer free energies as previous work had done when computing log *P* and log *D* values and, while in principle this should be an unimportant methodological detail, we wanted to assess whether this choice had negatively impacted results. Thus, we ran the indirect solvation-based transfer free energy protocol described below.

The set of indirect solvation-based alchemical free energy calculations were run using OpenEye's Orion cloud computing platform, also with the YANK software but with an alternate, fully automated workflow. The Orion workflow utilizes a very similar approach to that utilized above, except that it employs solvation free energy calculations for each molecule in each phase, rather than computing transfer free energies. Details of equilibration and simulation length are also different, as described below – with the largest difference being equilibration protocol.

On Orion, the input for each calculation was the target solute (SMILES) and the target solvent (SMILES), along with the temperature (298.15 K) and a guessed initial density for each solution (here 1.0 g/mL for solutes in water, 0.83 g/mL for octanol, to match experiment roughly). These settings are used on Orion, to prepare initial simulations via an internal Orion workflow based on that used in SolvationToolkit. The GAFF version used during parameterization on Orion was 1.8. In this Orion workflow, we also tested several additional potential tautomers for some molecules. For each molecule, we conducted a solvation free energy calculation of the solute in pure water and another in octanol. After parameterization, equilibration stages were run with OpenMM (version 7.2.2.dev-32bc79a) and free energy calculations were done with YANK (version 0.23.7 [115]). A cutoff for nonbonded interactions was set to 9 Å, electrostatic interactions were computed using PME, bonds involving hydrogen were constrained and HMR was used to allow for a 4 fs timestep.

The equilibration was carried out with OpenMM on Orion. The first step was 200 ps of NVT simulation with the solute heavy atoms harmonically restrained with 2.0 kcal/(mol·Å^2) spring constants. The second step of equilibration was 200 ps of NPT simulation with harmonic heavy atom restraints with a 0.1 kcal/(mol·Å^2) spring constant. These equilibrated structures were then used in YANK [61] simulations. The length of the YANK simulations were 5 ns for each replica. The electrostatic and steric interactions of the solute with the solvent were scaled using the same λ parameters listed in the transfer free energy protocol previously. The OpenEye workflow was also different in that it employed the ELF10 AM1-BCC charging engine (https://docs.eyesopen.com/toolkits/python/quacpactk/OEProtonClasses/OEAM1BCCELF10Charges.html, https://docs.eyesopen.com/applications/quacpac/theory/molcharge_theory.html), and only *syn* conformers of neutral carboxylic acids were retained for charging because, in OpenEye's view, *anti* comformers result in incorrect charges dominated by strong internal interactions which are not well suited for MM applications. The only carboxylic acid studied was SM08, but the modification of the charging procedure in this case (relative to that employed in our direct solvation free energy approach) appears to have significantly impacted employed partial charges, likely for the better, as performance on SM08 was markedly different with this protocol.

12.1.3 Empirical reference calculations

For all empirical calculations, the compounds were stripped of counter ions and neutralized. The pyridone tautomer of SM08 was used, as given, and as it is assumed to be the most stable tautomer.

The MOE/logP(o/w) model, the MOE/h_logP model, and the MOE/S_logP model are all available within the graphical modeling program MOE (MOE, available from the Chemical Computing Group, Montreal, www.chemcomp.com). The MOE/logP(o/w) model is based on 95 atom types, plus a few corrections for geminal halogens, 1-4 aromatic nitrogens, ethylene-glycol ethers, alkane carbons, and amino acids. The individual contributions were obtained from fitting to a data set of 1827 measurements, yielding an R^2 of 0.931 and an RMSE of 0.393 (P. Labute, logP(o/w) model, unpublished).

The MOE/h_logP model uses 8 2D-descriptors derived from Extended Huckel Theory (the descriptors used are the sum of atomic EHT donor and acceptor strengths, the sum over log(1 + pi-bond order), the sum over log(1+ d-orbital bond order), the Gerber ring number and Gerber atomic surface area [116], and the number of hydrogens and number of hydrophobic carbons (carbons with no heteroatom within 3 bonds). The contributions of these descriptors were obtained by fitting to 1836 molecules yielding a model with an R^2 of 0.084 and an RMSE of 0.59 (P. Labute, MOE h_mr, h_logP, and h_logS models, unpublished). The MOE program is available from the Chemical Computing Group, Montreal (www.chemcomp.com). The MOE/S_logP model is described in this reference [35]. In this model, 68 different atom types were defined based on element and nearest neighbors, e.g. 27 different carbon types or 14 different nitrogen types. Then the atomic contributions were determined by fitting to a training set of almost 10000 molecules.

The MoKa/logP methodology [MoKa-3.2.2, Molecular Discovery Ltd, London, www.moldiscovery.com] builds on a similar approach as the corresponding pK_a prediction [117]. The procedure starts by calculating molecular interaction fields based on the GRID force field on a large number of molecular fragments. The 3D energy fields of these fragments are then stored and used to recompute any molecule as a summation of appropriate 3D fragments. Therefore any molecule can be quickly approximated by 3D fields describing polar and hydrophobic interaction with water and n-octanol. From these fields, VolSurf descriptors are computed and used in a training scheme using a database of about 20000 compounds. From the training model, a final model is computed to make external predictions (G. Cruciani, personal communication).

References for the Supplementary Information

111 Case D, Berryman J, Betz R, Cerutti D, Cheatham Iii T, Darden T, Duke R, Giese T, Gohlke H, Goetz A et al (2015) AMBER 2015. University of California, San Francisco.

112 Martínez L, Andrade R, Birgin EG, Martínez JM (2009) PACKMOL: a package for building initial conigurations for molecular dynamics simulations. J Comput Chem 30(13):2157–2164. https://doi.org/10.1002/jcc.21224

113 Eastman P, Swails J, Chodera JD, McGibbon RT, Zhao Y, Beauchamp KA, Wang LP, Simmonett AC, Harrigan MP, Stern CD, Wiewiora RP, Brooks BR, Pande VS (2017) OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. PLoS Comput Biol 13(7):e1005659. https://doi.org/10.1371/journal.pcbi.1005659

114 Hopkins CW, Le Grand S, Walker RC, Roitberg AE (2015) Longtime- step molecular dynamics through hydrogen mass repartitioning. J Chem Theory Comput 11(4):1864–1874. https://doi.org/10.1021/ct5010406

115 Rizzi A, Chodera J, Naden L, Beauchamp K, Albanese S, Grinaway P, Rustenburg B, Saladi S, Boehm K (2018) choderalab/yank: Bugix release. Zenodo. https://doi.org/10.5281/zenodo.1447109.

116 Gerber PR (1998) Charge distribution from a simple molecular orbital type calculation and non-bonding interaction terms in the force ield MAB. J Comput Aided Mol Des 12(1):37–51. https://doi.org/10.1023/A:1007902804814

117 Milletti F, Storchi L, Sforna G, Cruciani G (2007) New and original pKa prediction method using grid molecular interaction fields. J Chem Inf Model 47(6):2172–2181. https://doi.org/10.1021/ci700018y

12.2 Supplementary figures and tables

Table S1. Method details of log *P* predictions with MM-based physical methods. Force fields, water models, and octanol phase choice are reported. A dry octanol phase indicates the octanol phase was treated as consisting of pure octanol. A wet octanol phase indicates the octanol phase was treated as a mixture of octanol and water. RMSE and Kendall's Tau values are reported as mean and 95% confidence intervals. A CSV version of this table can be found in *SAMPL6-supplementary-documents.tar.gz*.

Submission ID	Method Name	Force Field	Water Model	Octanol Phase	RMSE	Kendall's Tau (τ)
nh6c0	Molecular-Dynamics-Expanded-Ensembles	AMBER/OPLS like force field with manually adjusted parameters	modified Toukan-Rahman	wet	0.74 [0.56, 0.93]	0.49 [0.02, 0.87]
ujsgv	Alchemical-CGenFF	CGenFF	TIP3P	wet	0.82 [0.56, 1.06]	0.35 [-0.14, 0.79]
2mi5w	Alchemical-CGenFF	CGenFF	TIP3P	wet	0.95 [0.64, 1.24]	0.24 [-0.22, 0.71]
y0xxd	FS-GM (Fast switching Growth Method)	CGenFF	OPC3	dry	1.04 [0.42, 1.50]	0.42 [-0.14, 0.91]
2ggir	FS-AGM (Fast switching Annihilation/Growth Method)	CGenFF	OPC3	dry	1.04 [0.84, 1.24]	0.49 [-0.02, 0.92]
3wvyh	Alchemical-CGenFF	CGenFF	TIP3P	dry	1.13 [0.48, 1.75]	0.55 [0.11, 0.95]
25s67	FS-AGM (Fast switching Annihilation/Growth Method)	OPLS-AA	OPC3	dry	1.21 [0.84, 1.54]	0.45 [-0.14, 0.88]
v2q0t	InterX_GAFF_WET_OCTANOL	GAFF	TIP3P	wet	1.31 [0.94, 1.65]	0.64 [0.14, 1.00]
ggm6n	FS-GM (Fast switching Growth Method)	OPLS-AA	OPC3	dry	1.32 [0.95, 1.64]	0.53 [0.08, 0.87]
jjd0b	MD/S-MBIS-GAFF-TIP3P/MBAR/	GAFF (parameters refined w.r.t to QM	TIP3P	dry	1.35 [0.89, 1.74]	0.53 [0.02, 0.91]
sqosi	MD-AMBER-dryoct	GAFF	TIP3P	dry	1.69 [1.14, 2.18]	0.45 [-0.06, 0.84]
ke5gu	MD/S-MBIS-GAFF-SPCE/MBAR/	GAFF (parameters refined w.r.t to QM	SPCE	dry	1.82 [1.31, 2.25]	0.53 [-0.02, 0.91]
mwuua	MD-LigParGen-wetoct	OPLS-AA	TIP4P	wet	1.83 [1.48, 2.12]	0.48 [0.02, 0.84]
fyx45	LogP-prediction-Drude-FEP-HuangLab	Drude	unknown	unknown	1.85 [0.63, 2.70]	0.67 [0.14, 1.00]
6nmtt	MD-AMBER-wetoct	GAFF	TIP3P	wet	1.87 [1.33, 2.45]	0.6 0[0.06, 1.00]
eufcy	MD-LigParGen-dryoct	OPLS-AA	TIP4P	dry	1.99 [1.62, 2.33]	0.66 [0.21, 0.96]
tzzb5	Alchemical-CGenFF	CGenFF (parameters refined w.r.t to QM)	TIP3P	wet	2.12 [1.55, 2.57]	-0.20 [-0.63, 0.29]
3oqhx	MD-CHARMM-dryoct	CGenFF	TIP3P	dry	2.14 [1.24, 2.86]	0.00 [-0.5, 0.51]
bzeez	FS-AGM (Fast switching Annihilation/Growth Method)	GAFF2	OPC3	dry	2.20 [1.83, 2.51]	0.53 [0.00, 0.91]
5svjv	FS-GM (Fast switching Growth Method)	GAFF2	OPC3	dry	2.26 [1.84, 2.66]	0.44 [-0.15, 0.92]
odex0	InterX_ARROW_2017_PIMD_SOLVENT2_WET_OCTANOL	ARROW FF	PIMD	wet	2.29 [1.63, 2.82]	-0.09 [-0.61, 0.50]
padym	InterX_ARROW_2017_PIMD_WET_OCTANOL	ARROW FF	PIMD	wet	2.29 [1.63, 2.81]	-0.13 [-0.69, 0.48]
pnc4j	LogP-prediction-Drude-Umbrella-HuangLab	Drude	unknown	unknown	2.29 [1.68, 2.88]	0.20 [-0.37, 0.70]
REF02	YANK-GAFF-tip3p-wet-oct	GAFF	TIP3P	wet	2.29 [1.07, 3.53]	0.53 [0.06, 0.92]
REF05	YANK-SMIRNOFF-tip3p-wet-oct	SMIRNOFF	TIP3P	wet	2.31 [1.20, 3.47]	0.45 [-0.04, 0.85]
REF08	YANK-SMIRNOFF-tip3p-dry-oct	SMIRNOFF	TIP3P	dry	2.34 [1.04, 3.65]	0.42 [-0.04, 0.75]
REF07	YANK-GAFF-tip3p-dry-oct	GAFF	TIP3P	dry	2.38 [1.03, 3.73]	0.53 [0.09, 0.88]
fcspk	ARROW_2017_PIMD_SOLVENT2	ARROW FF	ARROW FF	dry	2.40 [1.72, 2.95]	-0.16 [-0.65, 0.40]
6cm6a	ARROW_2017_PIMD	ARROW FF	ARROW FF	dry	2.41 [1.75, 2.93]	-0.27 [-0.72, 0.29]
623c0	MD-OPLSAA-wetoct	OPLS-AA	TIP4P	wet	2.67 [2.13, 3.20]	0.38 [-0.14, 0.84]
4nfzz	MD/S-HI-GAFF-TIP3P/MBAR/	GAFF (parameters refined w.r.t to QM)	TIP3P	dry	2.67 [1.98, 3.35]	0.42 [-0.13, 0.88]
eg52i	ARROW_2017	ARROW FF	ARROW FF	dry	2.86 [2.01, 3.56]	-0.16 [-0.59, 0.35]
cp8kv	MD-OPLSAA-dryoct	OPLS-AA	TIP4P	dry	2.88 [2.31, 3.60]	0.59 [0.11, 1.00]
5585v	Alchemical-CGenFF	CGenFF (parameters refined w.r.t to QM)	TIP3P	wet	2.88 [2.02, 3.67]	-0.2 [-0.76, 0.32]
REF04	YANK-SMIRNOFF-TIP3P-FB-wet-oct	SMIRNOFF	TIP3P-FB	wet	3.22 [2.04, 4.48]	0.42 [-0.08, 0.84]
hf4wj	MD/S-HI-GAFF-SPCE/MBAR/	GAFF (parameters refined w.r.t to QM)	SPCE	dry	3.28 [2.49, 4.11]	0.38 [-0.16, 0.84]
REF01	YANK-GAFF-TIP3P-FB-wet-oct	GAFF	TIP3P-FB	wet	3.33 [2.08, 4.72]	0.49 [0.08, 0.83]
REF06	YANK-SMIRNOFF-OPC-wet-oct	SMIRNOFF	OPC	wet	3.64 [2.37, 4.97]	0.31 [-0.14, 0.72]
REF03	YANK-GAFF-opc-wet-oct	GAFF	OPC	wet	4.01 [2.74, 5.34]	0.42 [-0.06, 0.79]

Table S2. SMILES and InChI identifiers of SAMPL6 log *P* Challenge molecules. Experimental log *P* values can be found in a separate paper reporting measurements [9]. A CSV version of this table can be found in *SAMPL6-supplementary-documents.tar.gz*.

SAMPL6 Molecule ID	isomeric SMILES	InChI
SM02	c1ccc2c(c1)c(ncn2)Nc3cccc(c3)C(F)(F)F	InChI=1S/C15H10F3N3/c16-15(17,18)10-4-3-5-11(8-10)21-14-12-6-1-2-7-13(12)19-9-20-14/h1-9H,(H,19,20,21)
SM04	c1ccc2c(c1)c(ncn2)NCc3ccc(cc3)Cl	InChI=1S/C15H12ClN3/c16-12-7-5-11(6-8-12)9-17-15-13-3-1-2-4-14(13)18-10-19-15/h1-8,10H,9H2,(H,17,18,19)
SM07	c1ccc(cc1)CNc2c3ccccc3ncn2	InChI=1S/C15H13N3/c1-2-6-12(7-3-1)10-16-15-13-8-4-5-9-14(13)17-11-18-15/h1-9,11H,10H2,(H,16,17,18)
SM08	Cc1ccc2c(c1)c(c(c(=O)[nH]2)CC(=O)O)c3ccccc3	InChI=1S/C18H15NO3/c1-11-7-8-15-13(9-11)17(12-5-3-2-4-6-12)14(10-16(20)21)18(22)19-15/h2-9H,10H2,1H3,(H,19,22)(H,20,21)
SM09	COc1cccc(c1)Nc2c3ccccc3ncn2.Cl	InChI=1S/C15H13N3O.ClH/c1-19-12-6-4-5-11(9-12)18-15-13-7-2-3-8-14(13)16-10-17-15;/h2-10H,1H3,(H,16,17,18);1H
SM11	c1ccc(cc1)n2c3c(cn2)c(ncn3)N	InChI=1S/C11H9N5/c12-10-9-6-15-16(11(9)14-7-13-10)8-4-2-1-3-5-8/h1-7H,(H2,12,13,14)
SM12	c1ccc2c(c1)c(ncn2)Nc3cccc(c3)Cl.Cl	InChI=1S/C14H10ClN3.ClH/c15-10-4-3-5-11(8-10)18-14-12-6-1-2-7-13(12)16-9-17-14;/h1-9H,(H,16,17,18);1H
SM13	Cc1cccc(c1)Nc2c3cc(c(cc3ncn2)OC)OC	InChI=1S/C17H17N3O2/c1-11-5-4-6-12(7-11)20-17-13-8-15(21-2)16(22-3)9-14(13)18-10-19-17/h4-10H,1-3H3,(H,18,19,20)
SM14	c1ccc(cc1)n2cnc3c2ccc(c3)N	InChI=1S/C13H11N3/c14-10-6-7-13-12(8-10)15-9-16(13)11-4-2-1-3-5-11/h1-9H,14H2
SM15	c1ccc2c(c1)ncn2c3ccc(cc3)O	InChI=1S/C13H10N2O/c16-11-7-5-10(6-8-11)15-9-14-12-3-1-2-4-13(12)15/h1-9,16H
SM16	c1cc(c(c(c1)Cl)C(=O)Nc2ccncc2)Cl	InChI=1S/C12H8Cl2N2O/c13-9-2-1-3-10(14)11(9)12(17)16-8-4-6-15-7-5-8/h1-7H,(H,15,16,17)

Table S3. SMILES and InChI identifiers of extra molecules included in the evaluation of reference methods. A CSV version of this table can be found in *SAMPL6-supplementary-documents.tar.gz*. Experimental log *P* values can be found in a separate paper reporting measurements [108] and in machine readable format in https://github.com/samplchallenges/SAMPL6/blob/master/physical_properties/logP/analysis_of_extra_molecules/logP_experimental_values.csv.

Extra Molecule ID	Isomeric SMILES	InChI
3,4-Dichlorophenol	c1cc(c(cc1O)Cl)Cl	InChI=1S/C6H4Cl2O/c7-5-2-1-4(9)3-6(5)8/h1-3,9H
3,5-Dichlorophenol	c1c(cc(cc1Cl)Cl)O	InChI=1S/C6H4Cl2O/c7-4-1-5(8)3-6(9)2-4/h1-3,9H
3-Bromoquinoline	c1ccc2c(c1)cc(cn2)Br	InChI=1S/C9H6BrN/c10-8-5-7-3-1-2-4-9(7)11-6-8/h1-6H
3-Chlorophenol	c1cc(cc(c1)Cl)O	InChI=1S/C6H5ClO/c7-5-2-1-3-6(8)4-5/h1-4,8H
4-Butoxyphenol	CCCCOc1ccc(cc1)O	InChI=1S/C10H14O2/c1-2-3-8-12-10-6-4-9(11)5-7-10/h4-7,11H,2-3,8H2,1H3
4-Chlorophenol	c1cc(ccc1O)Cl	InChI=1S/C6H5ClO/c7-5-1-3-6(8)4-2-5/h1-4,8H
4-Ethoxyphenol	CCOc1ccc(cc1)O	InChI=1S/C8H10O2/c1-2-10-8-5-3-7(9)4-6-8/h3-6,9H,2H2,1H3
4-Iodophenol	c1cc(ccc1O)I	InChI=1S/C6H5IO/c7-5-1-3-6(8)4-2-5/h1-4,8H
4-Methoxyphenol	COc1ccc(cc1)O	InChI=1S/C7H8O2/c1-9-7-4-2-6(8)3-5-7/h2-5,8H,1H3
4-Pentoxyphenol	CCCCCOc1ccc(cc1)O	InChI=1S/C11H16O2/c1-2-3-4-9-13-11-7-5-10(12)6-8-11/h5-8,12H,2-4,9H2,1H3
4-Propoxyphenol	CCCOc1ccc(cc1)O	InChI=1S/C9H12O2/c1-2-7-11-9-5-3-8(10)4-6-9/h3-6,10H,2,7H2,1H3
Acebutolol	CCCC(=O)Nc1ccc(c(c1)C(=O)C)OCC(CNC(C)C)O	InChI=1S/C18H28N2O4/c1-5-6-18(23)20-14-7-8-17(16(9-14)13(4)21)24-11-15(22)10-19-12(2)3/h7-9,12,15,19,22H,5-6,10-1
Amylobarbitone	CCC1(C(=O)NC(=O)NC1=O)CCC(C)C	InChI=1S/C11H18N2O3/c1-4-11(6-5-7(2)3)8(14)12-10(16)13-9(11)15/h7H,4-6H2,1-3H3,(H2,12,13,14,15,16)
Benzoicacid	c1ccc(cc1)C(=O)O	InChI=1S/C7H6O2/c8-7(9)6-4-2-1-3-5-6/h1-5H,(H,8,9)
Butobarbitone	CCCCC1(C(=O)NC(=O)NC1=O)CC	InChI=1S/C10H16N2O3/c1-3-5-6-10(4-2)7(13)11-9(15)12-8(10)14/h3-6H2,1-2H3,(H2,11,12,13,14,15)
Celiprolol	CCN(CC)C(=O)Nc1ccc(c(c1)C(=O)C)OCC(CNC(C)(C)C)O	InChI=1S/C20H33N3O4/c1-7-23(8-2)19(26)22-15-9-10-18(17(11-15)14(3)24)27-13-16(25)12-21-20(4,5)6/h9-11,16,21,25H,7-8,12-13H2,1-6H3,(H,22,26)
Chlorpromazine	CN(C)CCCN1c2ccccc2Sc3c1cc(cc3)Cl	InChI=1S/C17H19ClN2S/c1-19(2)10-5-11-20-14-6-3-4-7-16(14)21-17-9-8-13(18)12-15(17)20/h3-4,6-9,12H,5,10-11H2,1-2H3
Ketoprofen	CC(c1cccc(c1)C(=O)c2ccccc2)C(=O)O	InChI=1S/C16H14O3/c1-11(16(18)19)13-8-5-9-14(10-13)15(17)12-6-3-2-4-7-12/h2-11H,1H3,(H,18,19)
N-Methylaniline	CNc1ccccc1	InChI=1S/C7H9N/c1-8-7-5-3-2-4-6-7/h2-6,8H,1H3
Pentobarbitone	CCCC(C)C1(C(=O)NC(=O)NC1=O)CC	InChI=1S/C11H18N2O3/c1-4-6-7(3)11(5-2)8(14)12-10(16)13-9(11)15/h7H,4-6H2,1-3H3,(H2,12,13,14,15,16)
Pericyazine	c1ccc2c(c1)N(c3cc(ccc3S2)C#N)CCCN4CCC(CC4)O	InChI=1S/C21H23N3OS/c22-15-16-6-7-21-19(14-16)24(18-4-1-2-5-20(18)26-21)11-3-10-23-12-8-17(25)9-13-23/h1-2,4-7,14,17,25H,3,8-13H2
Phenobarbitone	CCC1(C(=O)NC(=O)NC1=O)c2ccccc2	InChI=1S/C12H12N2O3/c1-2-12(8-6-4-3-5-7-8)9(15)13-11(17)14-10(12)16/h3-7H,2H2,1H3,(H2,13,14,15,16,17)
Phenol	c1ccc(cc1)O	InChI=1S/C6H6O/c7-6-4-2-1-3-5-6/h1-5,7H
Quinalbarbitone	CCCC(C)C1(C(=O)NC(=O)NC1=O)CC=C	InChI=1S/C12H18N2O3/c1-4-6-8(3)12(7-5-2)9(15)13-11(17)14-10(12)16/h5,8H,2,4,6-7H2,1,3H3,(H2,13,14,15,16,17)
Quinoline	c1ccc2c(c1)cccn2	InChI=1S/C9H7N/c1-2-6-9-8(4-1)5-3-7-10-9/h1-7H
(±)-Propranolol	CC(C)NCC(COc1cccc2c1cccc2)O	InChI=1S/C16H21NO2/c1-12(2)17-10-14(18)11-19-16-9-5-7-13-6-3-4-8-15(13)16/h3-9,12,14,17-18H,10-11H2,1-2H3
Sulfamethazine	Cc1cc(nc(n1)NS(=O)(=O)c2ccc(cc2)N)C	InChI=1S/C12H14N4O2S/c1-8-7-9(2)15-12(14-8)16-19(17,18)11-5-3-10(13)4-6-11/h3-7H,13H2,1-2H3,(H,14,15,16)

Table S4. Evaluation statistics calculated for all methods. Methods are represented via their SAMPL6 submission IDs which can be cross referenced with Table 3 for method details. There are six error metrics reported: the root-mean-squared error (RMSE), mean absolute error (MAE), mean (signed) error (ME), coefficient of determination (R^2), linear regression slope (m), and Kendall's Rank Correlation Coefficient (τ). This table is ranked by increasing RMSE. A CSV version of this table can be found in *SAMPL6-supplementary-documents.tar.gz*.

Submission ID	RMSE	MAE	ME	R^2	m	Kendall's Tau
hmz0n	0.38 [0.23,0.55]	0.31 [0.19,0.46]	-0.17 [-0.38,0.03]	0.77 [0.34,0.94]	0.94 [0.59,1.15]	0.64 [0.16,0.96]
gmoq5	0.39 [0.28,0.49]	0.34 [0.23,0.46]	0.01 [-0.21,0.25]	0.74 [0.39,0.92]	0.99 [0.66,1.33]	0.59 [0.10,0.88]
3vqbi	0.41 [0.28,0.53]	0.36 [0.24,0.48]	-0.08 [-0.30,0.17]	0.66 [0.27,0.93]	0.78 [0.50,1.10]	0.56 [0.12,0.91]
sq07q	0.47 [0.33,0.58]	0.41 [0.28,0.54]	0.03 [-0.24,0.31]	0.64 [0.21,0.89]	0.92 [0.51,1.30]	0.56 [0.10,0.88]
j8nwc	0.47 [0.17,0.75]	0.31 [0.15,0.55]	0.07 [-0.16,0.38]	0.74 [0.33,0.97]	1.14 [0.84,1.38]	0.81 [0.44,1.00]
xxh4i	0.49 [0.34,0.62]	0.43 [0.29,0.57]	0.18 [-0.09,0.43]	0.54 [0.14,0.86]	0.60 [0.29,1.03]	0.51 [0.00,0.88]
hdpuj	0.49 [0.37,0.61]	0.44 [0.32,0.57]	-0.29 [-0.51,-0.05]	0.74 [0.40,0.94]	1.02 [0.69,1.35]	0.67 [0.23,1.00]
dqxk4	0.49 [0.33,0.62]	0.42 [0.26,0.57]	0.30 [0.06,0.53]	0.69 [0.37,0.91]	0.83 [0.50,1.26]	0.67 [0.25,0.96]
vzgyt	0.50 [0.27,0.68]	0.38 [0.21,0.58]	-0.35 [-0.57,-0.15]	0.72 [0.28,0.95]	0.76 [0.48,0.98]	0.64 [0.25,0.92]
ypmr0	0.50 [0.36,0.63]	0.44 [0.31,0.58]	0.07 [-0.23,0.35]	0.61 [0.25,0.89]	0.93 [0.54,1.52]	0.64 [0.23,0.92]
yd6ub	0.51 [0.32,0.66]	0.41 [0.23,0.59]	0.09 [-0.21,0.38]	0.63 [0.21,0.89]	0.99 [0.47,1.41]	0.53 [-0.02,0.87]
7egyc	0.52 [0.35,0.66]	0.44 [0.28,0.60]	0.27 [0.01,0.52]	0.57 [0.22,0.85]	0.50 [0.32,0.77]	0.45 [0.06,0.83]
0a7a8	0.53 [0.34,0.69]	0.43 [0.25,0.62]	0.32 [0.07,0.56]	0.62 [0.13,0.90]	0.74 [0.34,1.02]	0.45 [-0.14,0.84]
7dhtp	0.54 [0.33,0.70]	0.44 [0.26,0.62]	0.06 [-0.27,0.36]	0.49 [0.06,0.88]	0.73 [0.26,1.15]	0.56 [0.04,0.96]
qyzjx	0.54 [0.34,0.75]	0.46 [0.31,0.65]	-0.15 [-0.41,0.19]	0.73 [0.33,0.97]	1.22 [0.89,1.50]	0.78 [0.45,1.00]
REF11	0.54 [0.25,0.80]	0.39 [0.19,0.64]	0.19 [-0.09,0.50]	0.59 [0.37,0.89]	0.90 [0.37,1.62]	0.67 [0.33,0.96]
REF13	0.55 [0.37,0.71]	0.47 [0.31,0.64]	-0.27 [-0.55,0.02]	0.69 [0.31,0.93]	1.06 [0.55,1.55]	0.60 [0.08,0.96]
w6jta	0.56 [0.33,0.76]	0.46 [0.28,0.66]	0.32 [0.06,0.61]	0.53 [0.12,0.89]	0.62 [0.34,0.86]	0.51 [0.02,0.88]
REF12	0.60 [0.42,0.76]	0.52 [0.36,0.70]	-0.08 [-0.43,0.26]	0.67 [0.23,0.90]	1.21 [0.76,1.53]	0.55 [0.06,0.88]
ji2zm	0.60 [0.43,0.75]	0.53 [0.38,0.70]	0.45 [0.22,0.67]	0.66 [0.32,0.90]	0.66 [0.43,0.96]	0.51 [0.11,0.84]
5krdi	0.60 [0.39,0.81]	0.51 [0.33,0.71]	-0.30 [-0.60,0.01]	0.63 [0.24,0.91]	1.03 [0.59,1.51]	0.60 [0.14,0.92]
REF10	0.60 [0.39,0.83]	0.51 [0.33,0.72]	-0.04 [-0.42,0.30]	0.38 [0.01,0.82]	0.65 [-0.03,1.21]	0.35 [-0.27,0.8]
gnxuu	0.61 [0.39,0.80]	0.51 [0.31,0.72]	0.40 [0.13,0.68]	0.53 [0.12,0.91]	0.57 [0.34,0.79]	0.51 [0.04,0.88]
tc4xa	0.62 [0.41,0.80]	0.51 [0.31,0.73]	0.17 [-0.18,0.53]	0.66 [0.17,0.90]	1.21 [0.52,1.65]	0.49 [-0.02,0.84]
6cdyo	0.65 [0.42,0.83]	0.54 [0.33,0.76]	-0.24 [-0.60,0.10]	0.52 [0.20,0.81]	0.93 [0.48,1.70]	0.53 [0.17,0.87]
dbmg3	0.70 [0.47,0.89]	0.60 [0.39,0.82]	0.42 [0.09,0.74]	0.47 [0.03,0.80]	0.75 [0.12,1.29]	0.38 [-0.18,0.80]
kxsp3	0.74 [0.49,0.94]	0.62 [0.39,0.86]	0.48 [0.14,0.80]	0.36 [0.02,0.77]	0.54 [0.04,1.15]	0.35 [-0.20,0.80]
nh6c0	0.74 [0.56,0.93]	0.67 [0.48,0.87]	0.09 [-0.35,0.53]	0.62 [0.16,0.88]	1.34 [0.52,1.91]	0.49 [0.02,0.87]
kivfu	0.78 [0.35,1.08]	0.56 [0.27,0.90]	-0.03 [-0.51,0.40]	0.41 [0.03,0.88]	0.97 [0.30,1.43]	0.45 [-0.02,0.84]
NULL0	0.79 [0.50,1.03]	0.66 [0.42,0.92]	0.42 [0.02,0.81]	0.00 [0.00,0.00]	0.00 [0.00,0.00]	0.00 [0.00,0.00]
ujsgv	0.82 [0.56,1.06]	0.67 [0.39,0.95]	-0.31 [-0.76,0.15]	0.33 [0.01,0.81]	0.80 [-0.02,1.45]	0.35 [-0.14,0.79]
REF09	0.82 [0.52,1.10]	0.68 [0.43,0.97]	-0.26 [-0.73,0.20]	0.46 [0.08,0.89]	1.09 [0.44,1.78]	0.48 [-0.02,0.86]
wu52s	0.83 [0.57,1.05]	0.72 [0.49,0.97]	0.70 [0.43,0.97]	0.55 [0.11,0.99]	0.54 [0.24,0.88]	0.56 [-0.06,1.00]
g6dwz	0.85 [0.56,1.08]	0.72 [0.45,0.99]	0.35 [-0.11,0.80]	0.52 [0.07,0.85]	1.18 [0.47,1.70]	0.45 [-0.07,0.84]
5mahv	0.85 [0.42,1.19]	0.62 [0.31,0.99]	-0.02 [-0.53,0.47]	0.34 [0.03,0.79]	0.90 [0.28,1.37]	0.24 [-0.33,0.72]
bqeuh	0.87 [0.51,1.17]	0.66 [0.34,1.01]	0.25 [-0.24,0.73]	0.01 [0.00,0.53]	-0.05 [-0.42,0.49]	0.02 [-0.57,0.57]
d7vth	0.87 [0.62,1.10]	0.78 [0.56,1.01]	-0.65 [-0.96,-0.30]	0.63 [0.21,0.93]	1.11 [0.73,1.39]	0.49 [0.02,0.85]
2mi5w	0.95 [0.64,1.24]	0.81 [0.54,1.12]	-0.3 [-0.83,0.23]	0.18 [0.00,0.64]	0.61 [-0.12,1.25]	0.24 [-0.22,0.71]
kuddg	0.97 [0.73,1.18]	0.89 [0.67,1.12]	0.89 [0.67,1.12]	0.67 [0.26,0.95]	0.71 [0.44,1.04]	0.53 [-0.02,0.96]
qz8d5	0.97 [0.71,1.19]	0.84 [0.56,1.12]	0.77 [0.42,1.10]	0.53 [0.18,0.84]	0.93 [0.49,1.58]	0.48 [0.06,0.82]
y0xxd	1.04 [0.42,1.50]	0.72 [0.32,1.21]	0.37 [-0.18,1.00]	0.33 [0.00,0.93]	1.03 [-0.20,2.00]	0.42 [-0.14,0.91]
2ggir	1.04 [0.84,1.24]	0.98 [0.76,1.19]	-0.36 [-0.88,0.27]	0.31 [0.00,0.93]	0.98 [-0.33,1.88]	0.49 [-0.02,0.92]
dyxbt	1.07 [0.79,1.34]	0.96 [0.70,1.23]	0.96 [0.70,1.23]	0.55 [0.11,0.9]	0.68 [0.22,1.15]	0.56 [0.12,0.92]
mm0jf	1.09 [0.91,1.24]	1.03 [0.81,1.22]	1.03 [0.81,1.22]	0.75 [0.44,0.98]	0.60 [0.39,0.82]	0.75 [0.38,1.00]
h83sb	1.12 [0.59,1.59]	0.87 [0.50,1.33]	-0.21 [-0.91,0.40]	0.00 [0.00,0.57]	-0.02 [-1.06,0.84]	-0.16 [-0.69,0.42]
3wvyh	1.13 [0.48,1.75]	0.77 [0.35,1.33]	0.26 [-0.32,0.99]	0.37 [0.03,0.93]	1.24 [0.32,2.29]	0.55 [0.11,0.95]
f3dpg	1.17 [0.74,1.52]	0.92 [0.50,1.36]	-0.85 [-1.33,-0.38]	0.11 [0.00,0.47]	0.36 [-0.18,0.85]	0.15 [-0.33,0.51]
25s67	1.21 [0.84,1.54]	1.06 [0.72,1.42]	-0.97 [-1.39,-0.55]	0.63 [0.16,0.90]	1.33 [0.43,2.34]	0.45 [-0.14,0.88]
zdj0j	1.21 [0.98,1.41]	1.13 [0.86,1.37]	1.13 [0.86,1.37]	0.64 [0.26,0.94]	0.86 [0.41,1.31]	0.64 [0.18,0.96]
7gg6s	1.27 [0.81,1.62]	1.00 [0.55,1.47]	-1.00 [-1.47,-0.55]	0.10 [0.00,0.46]	0.31 [-0.17,0.77]	0.16 [-0.33,0.55]
hwf2k	1.28 [0.57,1.90]	0.93 [0.49,1.50]	-0.09 [-0.92,0.57]	0.12 [0.00,0.84]	0.68 [-0.77,1.60]	0.31 [-0.32,0.79]
pcv32	1.28 [1.00,1.53]	1.17 [0.84,1.47]	1.17 [0.84,1.47]	0.50 [0.14,0.89]	0.75 [0.26,1.38]	0.44 [-0.04,0.81]
v2q0t	1.31 [0.94,1.65]	1.16 [0.82,1.52]	-1.15 [-1.52,-0.79]	0.70 [0.25,0.98]	1.31 [0.92,1.57]	0.64 [0.14,1.00]

Table S5. [Table S4 continued.] Evaluation statistics calculated for all methods. Methods are represented via their SAMPL6 submission IDs which can be cross referenced with Table 3 for method details. There are six error metrics reported: the root-mean-squared error (RMSE), mean absolute error (MAE), mean (signed) error (ME), coefficient of determination (R^2), linear regression slope (m), and Kendall's Rank Correlation Coefficient (τ). This table is ranked by increasing RMSE. A CSV version of this table can be found in *SAMPL6-supplementary-documents.tar.gz*.

Submission ID	RMSE	MAE	ME	R^2	m	Kendall's Tau
rdsnw	1.32 [0.88,1.71]	1.15 [0.80,1.54]	1.15 [0.80,1.54]	0.78 [0.39,0.97]	1.51 [1.16,1.77]	0.75 [0.37,1.00]
ggm6n	1.32 [0.95,1.64]	1.16 [0.79,1.54]	-1.15 [-1.53,-0.76]	0.53 [0.12,0.84]	1.04 [0.46,1.67]	0.53 [0.08,0.87]
jjd0b	1.35 [0.89,1.74]	1.13 [0.71,1.57]	-1.09 [-1.56,-0.63]	0.66 [0.24,0.91]	1.51 [0.81,2.04]	0.53 [0.02,0.91]
2tzb0	1.38 [0.94,1.79]	1.21 [0.85,1.62]	1.21 [0.85,1.62]	0.79 [0.42,0.97]	1.58 [1.21,1.86]	0.75 [0.36,1.00]
cr3hs	1.39 [0.58,2.10]	0.96 [0.46,1.61]	0.80 [0.21,1.52]	0.40 [0.01,0.79]	1.36 [-0.19,2.63]	0.35 [-0.33,0.84]
arw58	1.41 [0.81,1.89]	1.09 [0.60,1.63]	1.01 [0.45,1.61]	0.09 [0.00,0.54]	-0.24 [-0.75,0.26]	-0.20 [-0.64,0.36]
ahmtf	1.41 [1.13,1.69]	1.33 [1.07,1.62]	1.33 [1.07,1.62]	0.55 [0.11,0.89]	0.70 [0.23,1.16]	0.56 [0.11,0.92]
o7djk	1.42 [1.14,1.70]	1.34 [1.07,1.62]	1.34 [1.07,1.62]	0.55 [0.12,0.89]	0.70 [0.24,1.16]	0.56 [0.10,0.92]
fmf7r	1.44 [1.03,1.76]	1.25 [0.83,1.66]	0.26 [-0.56,1.08]	0.05 [0.00,0.57]	0.47 [-0.89,2.08]	0.10 [-0.5,0.64]
4p2ph	1.44 [0.81,1.94]	1.12 [0.61,1.68]	1.04 [0.47,1.64]	0.09 [0.00,0.55]	-0.26 [-0.77,0.25]	-0.26 [-0.68,0.29]
6fyg5	1.50 [1.27,1.70]	1.44 [1.18,1.67]	1.44 [1.18,1.67]	0.69 [0.32,0.96]	0.93 [0.50,1.51]	0.71 [0.28,1.00]
sqosi	1.69 [1.14,2.18]	1.42 [0.89,1.97]	-1.40 [-1.97,-0.86]	0.51 [0.05,0.87]	1.40 [0.39,2.02]	0.45 [-0.06,0.84]
rs4ns	1.71 [1.12,2.23]	1.44 [0.92,2.01]	1.44 [0.92,2.01]	0.06 [0.00,0.50]	-0.19 [-0.71,0.29]	-0.22 [-0.69,0.34]
c7t5j	1.73 [1.14,2.24]	1.47 [0.94,2.02]	1.47 [0.94,2.02]	0.05 [0.00,0.49]	-0.18 [-0.72,0.30]	-0.16 [-0.66,0.40]
jc68f	1.74 [1.13,2.25]	1.47 [0.94,2.03]	1.47 [0.94,2.03]	0.05 [0.00,0.48]	-0.18 [-0.71,0.30]	-0.16 [-0.65,0.40]
03cyy	1.75 [0.57,2.73]	1.11 [0.43,2.01]	0.03 [-0.89,1.16]	0.00 [0.00,0.53]	0.12 [-1.15,1.49]	0.09 [-0.56,0.71]
hsotx	1.81 [1.22,2.32]	1.56 [1.03,2.11]	1.56 [1.03,2.11]	0.07 [0.00,0.49]	-0.19 [-0.66,0.25]	-0.20 [-0.67,0.36]
ke5gu	1.82 [1.31,2.25]	1.59 [1.07,2.09]	-1.59 [-2.09,-1.07]	0.62 [0.17,0.89]	1.54 [0.74,2.16]	0.53 [-0.02,0.91]
mwuua	1.83 [1.48,2.12]	1.73 [1.39,2.07]	-1.73 [-2.07,-1.39]	0.41 [0.01,0.77]	0.67 [0.07,1.13]	0.48 [0.02,0.84]
fe8ws	1.83 [1.24,2.34]	1.58 [1.06,2.13]	1.58 [1.06,2.13]	0.06 [0.00,0.48]	-0.18 [-0.67,0.26]	-0.16 [-0.64,0.41]
5t0yn	1.85 [1.26,2.37]	1.61 [1.09,2.15]	1.61 [1.09,2.15]	0.06 [0.00,0.49]	-0.18 [-0.67,0.27]	-0.16 [-0.65,0.41]
fyx45	1.85 [0.63,2.70]	1.25 [0.51,2.14]	0.65 [-0.3,1.74]	0.63 [0.17,0.92]	2.63 [1.09,3.88]	0.67 [0.14,1.00]
6nmtt	1.87 [1.33,2.45]	1.65 [1.16,2.20]	-1.65 [-2.20,-1.16]	0.42 [0.02,0.92]	1.10 [0.23,1.56]	0.60 [0.06,1.00]
eufcy	1.99 [1.62,2.33]	1.88 [1.49,2.25]	-1.77 [-2.25,-1.17]	0.54 [0.18,0.88]	1.43 [0.49,2.41]	0.66 [0.21,0.96]
tzzb5	2.12 [1.55,2.57]	1.87 [1.26,2.44]	1.43 [0.50,2.31]	0.20 [0.00,0.63]	-0.76 [-1.61,0.17]	-0.20 [-0.63,0.29]
3oqhx	2.14 [1.24,2.86]	1.64 [0.86,2.49]	1.11 [0.06,2.22]	0.03 [0.00,0.41]	-0.44 [-1.90,1.03]	0.00 [-0.50,0.51]
bzeez	2.20 [1.83,2.51]	2.07 [1.57,2.46]	-2.07 [-2.46,-1.57]	0.63 [0.17,0.95]	1.39 [0.77,2.03]	0.53 [0.00,0.91]
ynquk	2.26 [1.87,2.59]	2.13 [1.67,2.54]	2.13 [1.67,2.54]	0.08 [0.00,0.76]	0.25 [-0.25,0.61]	0.38 [-0.06,0.80]
5svjv	2.26 [1.84,2.66]	2.14 [1.69,2.58]	-2.03 [-2.57,-1.36]	0.39 [0.03,0.91]	1.20 [0.44,1.77]	0.44 [-0.15,0.92]
odex0	2.29 [1.63,2.82]	1.98 [1.31,2.65]	1.73 [0.82,2.57]	0.09 [0.00,0.64]	-0.53 [-1.76,0.68]	-0.09 [-0.61,0.50]
padym	2.29 [1.63,2.81]	1.99 [1.31,2.64]	1.72 [0.78,2.57]	0.12 [0.00,0.69]	-0.60 [-1.92,0.73]	-0.13 [-0.69,0.48]
pnc4j	2.29 [1.68,2.88]	2.03 [1.42,2.67]	2.03 [1.42,2.67]	0.04 [0.00,0.64]	0.31 [-0.81,1.30]	0.20 [-0.37,0.70]
REF02	2.29 [1.07,3.53]	1.68 [0.95,2.73]	-1.68 [-2.73,-0.95]	0.23 [0.00,0.91]	1.26 [0.02,2.29]	0.53 [0.06,0.92]
REF05	2.31 [1.20,3.47]	1.80 [1.15,2.76]	-1.80 [-2.76,-1.15]	0.20 [0.00,0.91]	1.07 [-0.08,2.18]	0.45 [-0.04,0.85]
REF08	2.34 [1.04,3.65]	1.66 [0.92,2.77]	-1.66 [-2.77,-0.92]	0.13 [0.00,0.81]	0.95 [-0.39,2.05]	0.42 [-0.04,0.75]
REF07	2.38 [1.03,3.73]	1.65 [0.84,2.80]	-1.65 [-2.80,-0.84]	0.24 [0.01,0.93]	1.43 [0.07,2.65]	0.53 [0.09,0.88]
fcspk	2.40 [1.72,2.95]	2.10 [1.41,2.79]	1.97 [1.12,2.76]	0.11 [0.00,0.65]	-0.50 [-1.60,0.61]	-0.16 [-0.65,0.40]
6cm6a	2.41 [1.75,2.93]	2.10[1.40,2.78]	1.94 [1.04,2.74]	0.19 [0.00,0.69]	-0.66 [-1.77,0.32]	-0.27 [-0.72,0.29]
bq6fo	2.58 [1.68,3.34]	2.15 [1.35,3.01]	1.55 [0.30,2.74]	0.10 [0.00,0.56]	1.05 [-0.88,2.73]	0.09 [-0.39,0.60]
623c0	2.67 [2.13,3.20]	2.53 [2.08,3.04]	-2.53 [-3.04,-2.08]	0.22 [0.00,0.80]	0.64 [-0.05,1.09]	0.38 [-0.14,0.84]
4nfzz	2.67 [1.98,3.35]	2.44 [1.83,3.10]	-2.44 [-3.10,-1.83]	0.40 [0.05,0.87]	1.30 [0.56,1.85]	0.42 [-0.13,0.88]
eg52i	2.86 [2.01,3.56]	2.41 [1.52,3.32]	2.06 [0.88,3.21]	0.15 [0.00,0.55]	-0.94 [-2.15,0.19]	-0.16 [-0.59,0.35]
cp8kv	2.88 [2.31,3.60]	2.72 [2.27,3.35]	-2.72 [-3.35,-2.27]	0.24 [0.01,0.93]	0.78 [-0.01,1.47]	0.59 [0.00.11,1]
5585v	2.88 [2.02,3.67]	2.55 [1.81,3.36]	2.40 [1.46,3.31]	0.04 [0.00,0.55]	-0.41 [-1.97,0.62]	-0.2 [-0.76,0.32]
j4nb3	2.89 [2.32,3.34]	2.63 [1.84,3.26]	2.63 [1.84,3.26]	0.01 [0.00,0.73]	0.12 [-0.74,0.90]	0.16 [-0.35,0.76]
REF04	3.22 [2.04,4.48]	2.76 [1.93,3.85]	-2.76 [-3.84,-1.93]	0.19 [0.00,0.82]	1.20 [0.01,2.22]	0.42 [-0.08,0.84]
hf4wj	3.28 [2.49,4.11]	3.04 [2.36,3.83]	-3.04 [-3.82,-2.36]	0.34 [0.03,0.85]	1.31 [0.48,1.95]	0.38 [-0.16,0.84]
REF01	3.33 [2.08,4.72]	2.82 [1.99,4.02]	-2.82 [-4.02,-1.99]	0.24 [0.01,0.90]	1.46 [0.05,2.63]	0.49 [0.08,0.83]
REF06	3.64 [2.37,4.97]	3.10 [2.08,4.34]	-3.10 [-4.33,-2.08]	0.16 [0.00,0.68]	1.24 [-0.50,2.68]	0.31 [-0.14,0.72]
REF03	4.01 [2.74,5.34]	3.58 [2.66,4.78]	-3.58 [-4.78,-2.66]	0.17 [0.00,0.84]	1.20 [-0.53,2.54]	0.42 [-0.06,0.79]
pku5g	4.87 [4.06,5.68]	4.68 [3.90,5.49]	4.68 [3.90,5.49]	0.49 [0.03,0.90]	1.80 [0.28,2.99]	0.56 [0.00,0.96]
po4g2	5.46 [4.35,6.63]	5.17 [4.17,6.28]	5.17 [4.17,6.28]	0.51 [0.04,0.88]	2.33 [0.36,3.75]	0.56 [0.00,1.00]

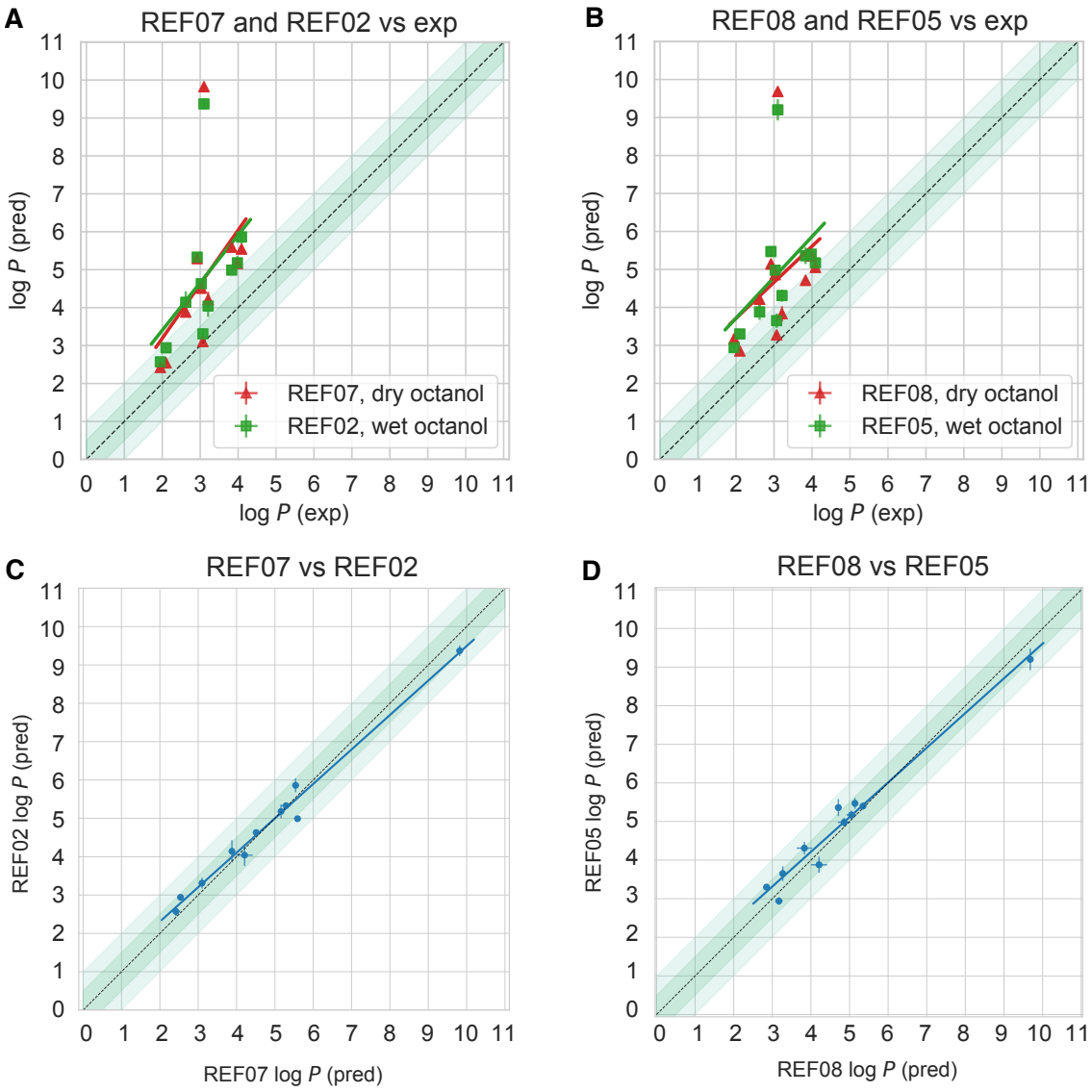

Figure S1. Varying the amount of water in the octanol phase has no significant effect on the predicted log *P* in reference calculations, as discussed in section 4.2.1. Comparison of predicted log *P* values to the experimental values using wet (27% water) and dry octanol phases and the (A) GAFF and (B) SMIRNOFF force field, from non-blinded reference calculations performed for this paper, shows no statistically significant difference in performance of methodologies. Comparison of the calculated log *P* using dry and wet octanol phases for (C) the GAFF force field and (D) the SMIRNOFF force field shows a small systematic difference.

Table S6. Comparison of force field parameters of the TIP3P, TIP3P-FB and OPC water models.

Water model	q_H(e)[1]	q_O(e)[1]	∠ HOH (deg)	l_1(Å)[2]	l_2(Å)[3]	σ_{LJ}(Å)[4]	ϵ_{LJ}(kJ/mol)[4]
TIP3P	0.417	-0.834	104.52	0.9572	-	3.151	0.636
TIP3P-FB	0.424	-0.848	108.15	1.0118	-	3.178	0.652
OPTIMAL POINT CHARGE	0.679	-1.358	103.6	0.8724	0.1594	3.167	0.89

[1] Corresponds to the hydrogen and oxygen charges.

[2] Corresponds to the bond length between the oxygen and hydrogen atoms.

[3] Corresponds to the length between the oxygen atom and virtual site.

[4] Corresponds to the Lennard-Jones (LJ) parameters of the oxygen.

Table S7. Comparison of the charges assigned to the syn and anti conformation of SM08_micro011 in the DFE protocol.

anti conformation			
atom number	atom type	atom name	charge
1	C1	C1	0.1252
2	C2	C2	-0.0073
3	C2	C3	0.0000
4	C2	C4	0.0000
5	C2	C5	-0.0074
6	C2	C6	-0.0118
7	C2	C7	0.0000
8	C2	C8	0.0236
9	C2	C9	-0.0196
10	C2	C10	0.3597
11	O1	O1	-0.2755
12	N1	N1	-0.1461
13	C1	C11	0.0310
14	C2	C12	0.3291
15	O1	O2	-0.1890
16	O2	O3	-0.2911
17	C2	C13	-0.0118
18	C2	C14	0.0000
19	C2	C15	0.0000
20	C2	C16	0.0000
21	C2	C17	0.0000
22	C2	C18	0.0000
23	H1	H1	-0.0393
24	H1	H2	-0.0393
25	H1	H3	-0.0393
26	H2	H4	0.0000
27	H2	H5	0.0000
28	H2	H6	0.0000
29	H3	H7	0.0865
30	H1	H8	-0.0393
31	H1	H9	-0.0393
32	H4	H10	0.2010
33	H2	H11	0.0000
34	H2	H12	0.0000
35	H2	H13	0.0000
36	H2	H14	0.0000
37	H2	H15	0.0000

syn conformation			
atom number	atom type	atom name	charge
1	C1	C1	-0.1205
2	C1	C2	-0.1322
3	C1	C3	-0.1322
4	C1	C4	-0.1112
5	C1	C5	-0.1112
6	C1	C6	-0.0742
7	C1	C7	-0.1856
8	C1	C8	-0.0653
9	C1	C9	-0.0839
10	C1	C10	-0.1199
11	C1	C11	-0.1160
12	C1	C12	0.0947
13	C1	C13	0.1024
14	C1	C14	-0.1731
15	C1	C15	0.6987
16	C1	C16	0.6458
17	C2	C17	-0.0498
18	C2	C18	-0.0780
19	N1	N1	-0.4371
20	O1	O1	-0.6386
21	O1	O2	-0.5474
22	O2	O3	-0.6145
23	H1	H1	0.1352
24	H1	H2	0.1377
25	H1	H3	0.1377
26	H1	H4	0.1459
27	H1	H5	0.1459
28	H1	H6	0.1381
29	H1	H7	0.1406
30	H1	H8	0.1469
31	H2	H9	0.0455
32	H2	H10	0.0455
33	H2	H11	0.0455
34	H2	H12	0.1028
35	H2	H13	0.1028
36	H3	H14	0.3362
37	H4	H15	0.4428

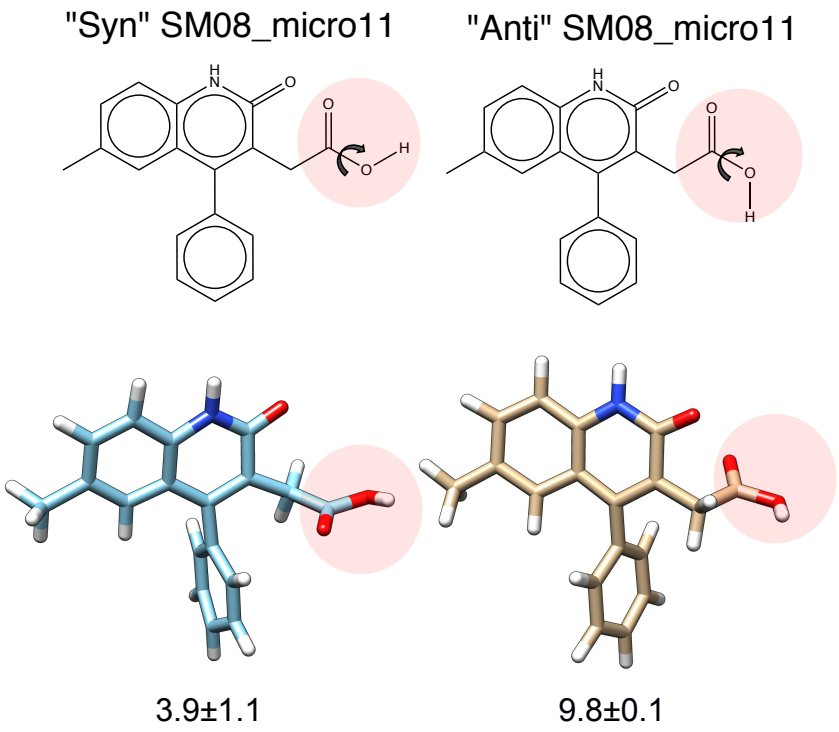

Figure S2. Shown here are the 2- and 3D structures of SM08_micro011 with the carboxylic acid in "anti" and "syn" conformation. The dihedral angle is indicated by the arrow around the carbon and oxygen atom. The calculated log *P* is included for comparison. The charges pertaining to each conformation are listed in Figure S7 and transition data is available in Figure S3

.

Figure S3. For the DFE method, the starting conformation impacts the number of C-O dihedral transitions for SM08_micro011, influencing sampling. Here is the transition data for the C-O dihedral in Figure S2, with charges listed in Table S7, for the DFE method (run in triplicate). In the “anti starting position” the torsion remains “anti” throughout the simulation, while the “syn starting position” allows transitions.