"Predicting solvation energies for kinetic modeling"
Amrit Jalan, Robert W. Ashcraft, Richard H. West and William H. Green.
*Annu. Rep. Prog. Chem., Sect. C*, 2010, **106**, DOI: 10.1039/b811056p

# Supplementary Material

**1 Testing Abraham and Platts methods. ..... 2**

1.1 SAMPL-0 ..... 3
1.2 Guthrie and Povar ..... 4
1.3 SAMPL-1 ..... 6

**2 References ..... 8**

## 1 Testing Abraham and Platts methods.

A version of the Platts[1] fragment method described in Section 4.3.5 of the accompanying article is implemented in the *Absolv* module of the commercial software package *ADME Boxes*,[2] with the fragment parameters modified and retrained on a larger database of experimentally obtained descriptors. Until recently a version of the software was also accessible online for free at http://pharma-algorithms.com/webboxes/. The calculations reported in this review were obtained using the Absolv module implemented in ADME Boxes version 4.95 distributed by Advanced Chemistry Development, Inc. (ACD/Labs) (http://www.acdlabs.com/). In many cases experimentally derived solute descriptors are available in the *Absolv* database accompanying the software, suggesting they are part of the fragment training set. However, calculated descriptors were used in this study, except where explicitly noted. The free energy of solvation is calculated from the Abraham linear solvation energy relationship (LSER) for the decadic logarithm of the gas-water partition coefficient $\log_{10} K_{aw}$:[3]

$$\log_{10} K_{aw} = \frac{-\Delta G^{o}_{solv}}{2.303RT} = -1.271 + 0.822E + 2.743S + 3.904A + 4.814B - 0.213L \qquad (1)$$

where *A, B, E, S, L* are the solute descriptors predicted by *Absolv*. Note that although this is the technically correct expression for air/water partitioning, it is different from the one found to perform slightly better by Schüürmann *et al.* in their study.[4]

For each compound in the test sets, the corresponding InChI key was generated, which was then used to obtain the SMILES representation. Solute descriptors were then obtained from these SMILES strings using the *Absolv* routine of *ADME Boxes*.

## 1.1 SAMPL-0

Table 1 shows the experimental values of the hydration free energy for the SAMPL0 test set as reported by Nicholls *et al.*[5] The second column of numbers lists the estimates using the *Absolv* fragment method to estimate the Abraham solute descriptors and the LSER shown in equation (1) above. Because experimentally-derived solute descriptors are available for most of the compounds in the sample set, the hydration free energy was also estimated using these descriptors, as reported in the third column of numbers.

**Table 1. Experimental hydration free energies (kcal/mol) and *Absolv* predictions for the SAMPL0 challenge set**

| Species | Experiment target | Calculated from estimated descriptors | Calculated from experimental descriptors |
|---|---|---|---|
| glycerol triacetate | -8.89 | -8.84 | -10.47 |
| benzyl bromide | -1.91 | -2.38 | -2.87 |
| benzyl chloride | -1.57 | -1.93 | -2.06 |
| m-bistrifluoro methyl benzene | 1.72 | 1.07 | 1.28 |
| bis-2-chloroethyl ether | -1.55 | -4.23 | -1.89 |
| 1,1-diethoxyethane | -2.40 | -3.28 | -3.51 |
| 1,4 dioxane | -2.84 | -5.05 | -4.80 |
| diethyl propanedioate | -5.42 | -6.00 | -5.85 |
| dimethoxymethane | -2.63 | -2.93 | -2.96 |
| ethylene glycol diacetate | -5.53 | -6.34 | -6.25 |
| 1,2-diethoxyethane | -2.30 | -3.54 | -5.23 |
| diethyl sulfide | -0.08 | -1.43 | -1.40 |
| imidazole | -7.11 | -9.81 | -8.44 |
| N,N-dimethyl-p-methoxybenzamide | -9.63 | -11.01 | ** |
| N,N-4-trimethylbenzamide | -7.72 | -9.76 | ** |
| 1,1-diacetoxyethane | -5.67 | -4.97 | ** |
| phenyl formate | -4.71 | -3.82 | ** |

All $\Delta G$ values are hydration free energies in kcal/mol

** Compounds for which experimental solute descriptors are **not** found in *Absolv* database

## 1.2 Guthrie and Povar

Table 2 shows the experimental and estimated hydration free energies of species in the Guthrie and Povar test set.[6] The compounds in this set, like the SAMPL0, are not as complex as those in the SAMPL1 test set. Most of the SAMPL0 compounds can be found in the Guthrie test set (indicated with a '*' in Table 2). We also observed that the *Absolv* database contained experimental descriptors for 42 out of the 54 compounds in the test set (the ones that did not have experimental descriptors have been marked with a '‡' in Table 2). All estimates were performed using estimated descriptors only.

In their original paper, Guthrie and Povar compared the performance of several variants of the PCM family of methods on the test set, which required proper treatment of the different standard states assumed by this class of methods (1 atm ideal gas and 1 mol/L ideal solution) compared to the more common 1 mol/L ideal gas and ideal solution phase).[6] The free energy obtained after correction was labeled $\Delta G_t$ while that obtained from conventional standard state definitions was labeled $\Delta G_s$ and the two are related at 298 K by the following relation: [6]

$$\Delta G_t = \Delta G_s + 1.894 \text{ kcal/mol}$$

This relation was used to obtain the experimental values listed in Table 2 so that the standard states are comparable.

**Table 2. Experimental hydration free energies (kcal/mol) and predictions using descriptors estimated by *Absolv*, for the Guthrie and Povar test set.**

| Species | Experimental | Calculated from estimated descriptors |
|---|---|---|
| 2,2,2-trifluoroethanol | -4.31 | -3.00 |
| methyl formate | -2.78 | -2.52 |
| methyl methanesulfonate | -4.87 | -7.01 |
| dimethyl sulfate | -5.10 | -8.80 |
| dimethyl sulfoxide | -8.71 | -7.65 |
| dimethyl sulfide | -1.54 | -0.37 |
| methyl trifluoroacetate | -1.10 | -0.73 |
| methly chloroacetate | -4.00 | -2.81 |
| methyl acetate | -3.35 | -2.30 |
| dimethoxy methane | -2.93 | -2.63 |
| trimethyl orthophosphate | -8.46 | -7.93 |
| 1,2-dimethoxyethane | -4.84 | -2.49 |
| trimethyl orthoformate | -4.42 | -4.57 |
| methyl cyanoacetate | -6.72 | -5.43 |
| sulfolane | -8.68 | -5.65 |

| | | |
|---|---|---|
| N-methyl morpholine | -6.34 | -4.79 |
| methyl t-butyl ether | -2.21 | -0.54 |
| trimethyl orthoacetate | -4.42 | -4.33 |
| N,N'-dimethyl piperazine | -7.58 | -6.70 |
| nitrobenzene | -4.12 | -4.00 |
| diethyl malonate | -6.00 | -5.42 |
| Cyanobenzene | -4.22 | -3.95 |
| methyl phenyl sulfide | -2.74 | -2.48 |
| diethyl succinate | -5.71 | -5.27 |
| methyl p-nitrobenzoate | -6.88 | -7.07 |
| methyl benzoate | -3.78 | -4.48 |
| N,N-dimethyl p-nitrobenzamide | -11.95 | -10.72 |
| methyl p-methoxybenzoate | -5.33 | -6.06 |
| Nitromethane | -4.20 | -2.31 |
| N,N-dimethylbenzamide | -9.29 | -8.06 |
| trimethyl orthobenzoate‡ | -4.04 | -6.52 |
| 3,3,3-trimethoxypropionitrile‡ | -6.40 | -7.46 |
| methane sulfonyl chloride‡ | -4.87 | -6.10 |
| trimethyl orthotrifluoroacetate‡ | -0.80 | -2.76 |
| trimethyl orthotrichloroacetate‡ | -4.59 | -4.68 |
| trimethyl orthomethoxyacetate‡ | -5.73 | -6.17 |
| phenyl trifluoroethyl ether‡ | -1.29 | -1.26 |
| N,N-dimethyl-p-methylbenzamide‡ | -9.76 | -7.72 |
| phenyl diethyl orthoformate‡ | -5.23 | -6.49 |
| N-methyl-N-(2,2,2-trifluoroethyl)-aniline‡ | -1.92 | -2.55 |
| glycerol triacetate* | -8.84 | -8.89 |
| benzyl bromide* | -2.38 | -1.91 |
| benzyl chloride* | -1.93 | -1.57 |
| m-bistrifluoro methyl benzene* | 1.07 | 1.72 |
| bis-2-chloroethyl ether* | -4.23 | -1.55 |
| 1,1-diethoxyethane* | -3.28 | -2.40 |
| 1,4 dioxane* | -5.05 | -2.85 |
| Dimethoxymethane* | -2.93 | -2.63 |
| ethylene glycol diacetate* | -6.34 | -5.53 |
| 1,2-diethoxyethane* | -3.54 | -2.30 |
| diethyl sulfide* | -1.43 | -0.08 |
| Imidazole* | -9.81 | -7.11 |
| 1,1-diacetoxyethane*,‡ | -4.97 | -5.67 |
| phenyl formate*,‡ | -3.82 | -4.71 |

All $\Delta G$ values are hydration free energies in kcal/mol

* Compounds included in the SAMPL0 challenge set

‡ Compounds for which experimental solute descriptors were not found in *Absolv* database

## 1.3 SAMPL-1

The SAMPL1 test set originally consisted of 63 drug-like compounds.[7] However, seven of these (cup08037, cup08042, cup08058, cup08059, cup08060, cup08061, cup08062) had to be omitted from the challenge due to errors and/or difficulties with the data preparation resulting in a reduced set of 56 compounds. The aqueous solvation energy of each compound in the dataset was calculated as a mean of all available experimental values. Details about data sources and the associated uncertainties may be obtained from Table 2 of ref [7]. Table 3 lists the 56 compounds in the test set using the nomenclature defined by Guthrie with the experimental value of the aqueous solvation free energy and the value obtained using *Absolv* estimates of the Abraham solute descriptors and the LSER shown in equation (1) above.

**Table 3. Experimental hydration free energies (kcal/mol) and predictions using descriptors estimated by *Absolv* for the SAMPL1 challenge set**

| Key | Species | Experiment | Calculated from estimated descriptors |
|---|---|---|---|
| cup08001 | nitroglycol | -5.73 | -4.29 |
| cup08002 | 1,2-dinitroxypropane | -4.95 | -4.35 |
| cup08003 | butyl nitrate † | -2.09 | -1.40 |
| cup08004 | 2-butyl nitrate | -1.82 | -1.50 |
| cup08005 | isobutyl nitrate | -1.88 | -1.50 |
| cup08006 | ethyleneglycol mononitrate | -8.18 | -6.03 |
| cup08007 | alachlor | -8.21 | -9.06 |
| cup08008 | aldicarb | -9.84 | -7.74 |
| cup08009 | ametryn | -7.65 | -11.34 |
| cup08010 | azinphosmethyl | -10.03 | -14.82 |
| cup08011 | benefin | -3.51 | -7.15 |
| cup08012 | bensulfuron ‡ | -17.17 | -29.92 |
| cup08013 | bromacil | -9.73 | -12.67 |
| cup08014 | captan | -9.01 | -10.85 |
| cup08015 | carbaryl | -9.45 | -9.76 |
| cup08016 | carbofuran | -9.61 | -9.36 |
| cup08017 | carbophenothion | -6.5 | -8.49 |
| cup08018 | chlordane | -3.44 | -5.55 |
| cup08019 | chlorfenvinphos | -7.07 | -11.08 |
| cup08020 | chlorimuronethyl ‡ | -14.01 | -22.97 |
| cup08021 | chloropicrin † | -1.45 | -2.31 |
| cup08022 | chlorpyrifos | -5.04 | -9.57 |
| cup08023 | dialifor * | -5.74 | -15.97 |
| cup08024 | diazinon | -6.48 | -10.05 |
| cup08025 | dicamba | -9.86 | -9.76 |

| | | | |
|---|---|---|---|
| cup08026 | dichlobenil | -4.71 | -4.08 |
| cup08027 | dinitramine | -5.66 | -10.68 |
| cup08028 | dinoseb | -6.23 | -8.58 |
| cup08029 | endosulfan alpha * | -4.23 | -16.37 |
| cup08030 | endrin † | -4.82 | -6.67 |
| cup08031 | ethion | -6.1 | -10.30 |
| cup08032 | fenuron | -9.13 | -10.00 |
| cup08033 | heptachlor † | -2.55 | -5.31 |
| cup08034 | isophorone † | -5.18 | -3.30 |
| cup08035 | lindane † | -5.44 | -4.52 |
| cup08036 | malathion | -8.15 | -11.31 |
| cup08038 | methyparathion | -7.19 | -10.18 |
| cup08039 | metsulfuronmethyl ‡ | -15.54 | -26.11 |
| cup08040 | nitralin * | -7.98 | -15.48 |
| cup08041 | nitroxyacetone | -5.99 | -5.96 |
| cup08043 | parathion | -6.74 | -9.93 |
| cup08044 | pebulate | -3.64 | -6.68 |
| cup08045 | phorate | -4.37 | -5.87 |
| cup08046 | profluralin | -2.45 | -7.56 |
| cup08047 | prometryn | -8.43 | -11.40 |
| cup08048 | propanil | -7.78 | -6.13 |
| cup08049 | pyrazon | -16.43 | -14.73 |
| cup08050 | simazine | -10.22 | -10.45 |
| cup08051 | sulfometuron-methyl ‡ | -20.25 | -23.57 |
| cup08052 | terbacil | -11.14 | -12.15 |
| cup08053 | terbutryn | -6.68 | -11.11 |
| cup08054 | thifensulfuron ‡ | -16.23 | -28.76 |
| cup08055 | trichlorfon † | -12.74 | -11.88 |
| cup08056 | trifluralin | -3.25 | -7.15 |
| cup08057 | vernolate | -4.13 | -6.68 |
| cup08063 | pirimor (pirimicarb) † | -9.41 | -11.29 |

All $\Delta G$ values are hydration free energies in kcal/mol

* One of three compounds with largest difference between Platts' prediction and experimental value

† Compounds for which experimental solute descriptors found in *Absolv* database

‡ Compounds of the sulfonyl-urea family

The RMS error of Platts predictions using equation (1) over the 56 compound data set is 4.55 kcal/mol and the mean absolute deviation (MAD) is found to be 3.25 kcal/mol. It is worth noting that a large part of this error comes from three compounds: nitralin (cup08040), endosulfan (cup08029), dialifor (cup08023) (indicated with a '*' in Table 1) whose absolute deviations from the experimental values are 7.5, 12.14 and 10.23 kcal/mol respectively. The sulfonyl-urea family of compounds (indicated with a '‡' in Table 1) also contributes significantly to the error. As described in the accompanying article, these were found to be outliers by many other methods.

## 2 References


1. J. A. Platts, D. Butina, M. H. Abraham and A. Hersey, *J. Chem. Inf. Comput. Sci.*, 1999, **39**, 835-845.
2. *ACD/ADME Boxes, version 4.95*, (2009) Advanced Chemistry Development, Inc., Toronto, ON, Canada, www.acdlabs.com.
3. M. H. Abraham, A. Ibrahim and A. M.Zissimos, *J. Chromatogr. A*, 2004, **1037**, 29-47.
4. G. Schüürmann, R.-U. Ebert and R. Künhe, *Chimia*, 2006, **60**, 691-698.
5. A. Nicholls, D. L. Mobley, J. P. Guthrie, J. D. Chodera, C. I. Bayly, M. D. Cooper and V. S. Pande, *J. Med. Chem.*, 2008, **51**, 769-779.
6. J. P. Guthrie and I. Povar, *Can. J. Chem.*, 2009, **87**, 1154-1162.
7. J. P. Guthrie, *J. Phys. Chem. B*, 2009, **113**, 4501-4507.