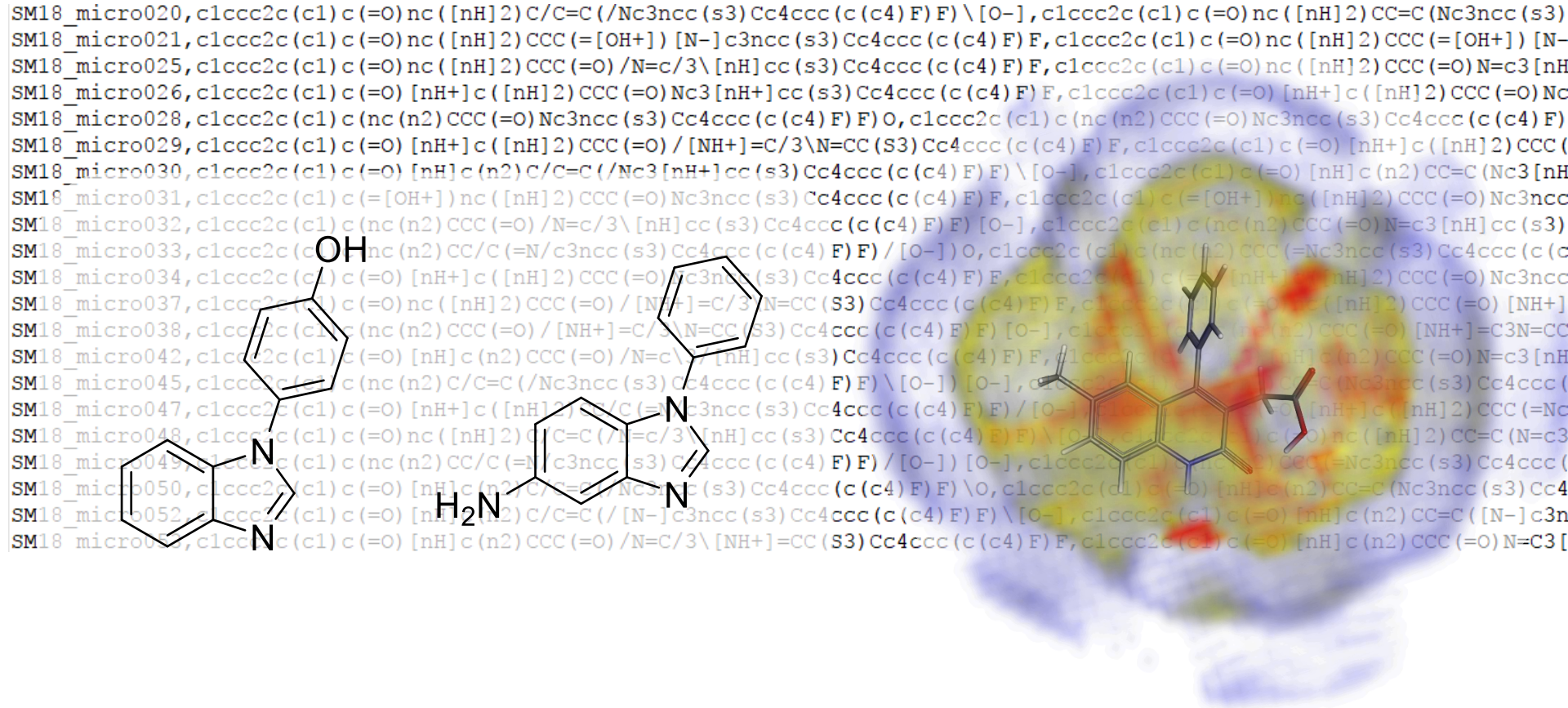
SAMPL6.2 - Prediction of partition coefficients for drug-like compounds

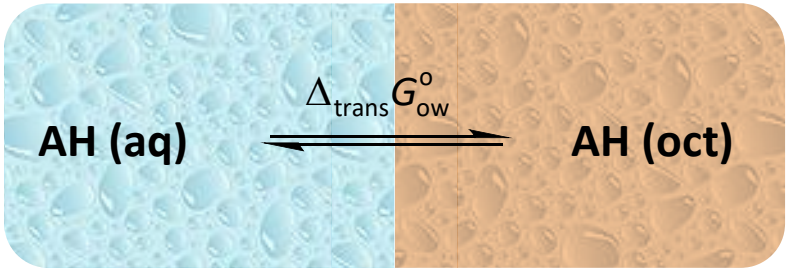


Nicolas Tielker, Lukas Eberlein, Stefan Güssregen, Stefan M. Kast

16.05.2019

Theoretical framework

Calculation of partition coefficients



$$\log P_{ow} = \lg \frac{[AH]_o}{[AH]_w} = \frac{\Delta_{trans} G^{o}_{ow}}{RT \ln 10} = \frac{G_o(AH) - G_w(AH)}{RT \ln 10}$$

- Free energies can be calculated directly via the Embedded Cluster Reference Interaction Site Model **(EC-RISM)**
- Same strategy used for SAMPL6 pK_a challenge

Microstate SMILES

↓

Find global minimum (Macromodel/OPLS3)

↓

QM optimization (Gaussian/PCM)

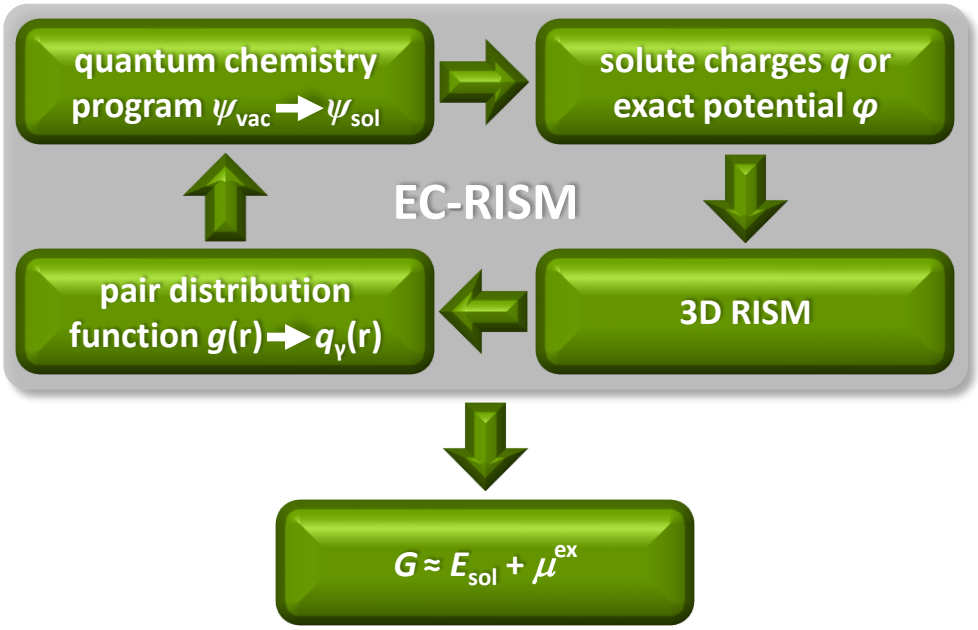
↓

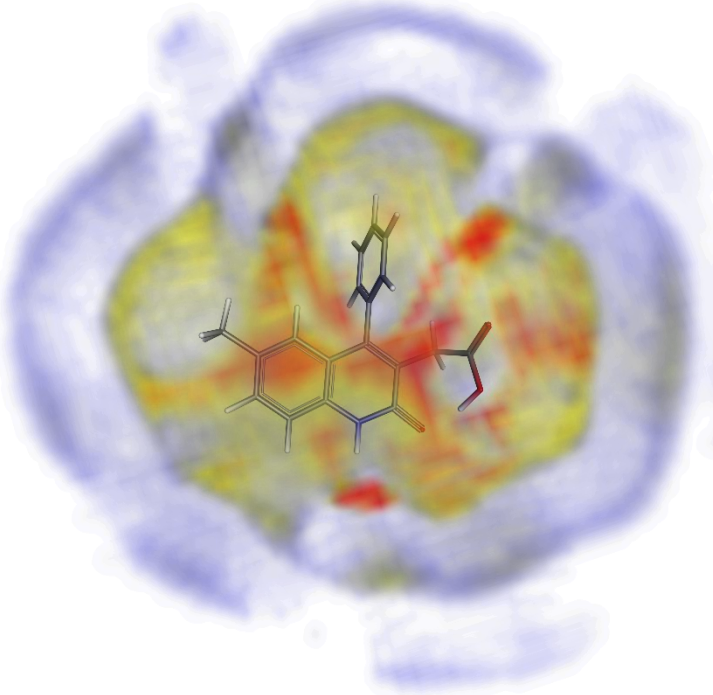
EC-RISM (Gaussian, GAFF)

Theoretical framework

Embedded Cluster Reference Interaction Site Model (EC-RISM)

- Solvent structure and the excess chemical potential are determined with 3D RISM
- QM calculations for self-consistent electronic and liquid structure
- Exact electrostatics, dispersion-repulsion: force field (GAFF)





D. Beglov, B. Roux, *J. Phys. Chem. B* **1997**, *101*, 7821; S. M. Kast, T. Kloss, *J. Chem. Phys.* **2008**, *129*, 236101
R. Frach, P. Kibies, S. Böttcher, T. Pongratz,…, S. M. Kast, *Angew. Chemie Int. Ed.* **2016**, *55*, 8757
N. Tielker, L. Eberlein, S. Güssregen, S. M. Kast, *J. Comput. Aided. Mol. Des.* **2018**, *32*, 1151

Preliminary work

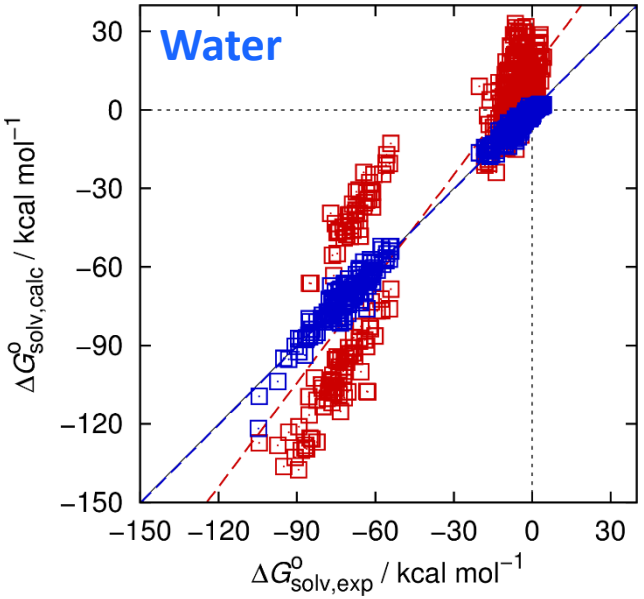
Partial molar volume (PMV) correction for EC-RISM

- Calculated chemical potential is typically too high due to overestimation of energy required for cavity formation
- Error shows an almost linear dependence on the solute's partial molar volume

$$\mu^{\mathrm{ex,corr}} = c_\mu \mu^{\mathrm{ex}} + c_V V_\mathrm{m} + c_q q$$

$$\Delta_{\mathrm{solv}} G^\mathrm{o} = E_{\mathrm{sol}} - E_{\mathrm{vac}} + \mu^{\mathrm{ex,corr}} + \Delta\mu^\mathrm{o}$$

MP2/6-311+G(d,p)/EC-RISM//B3LYP/6-311+G(d,p)/PCM

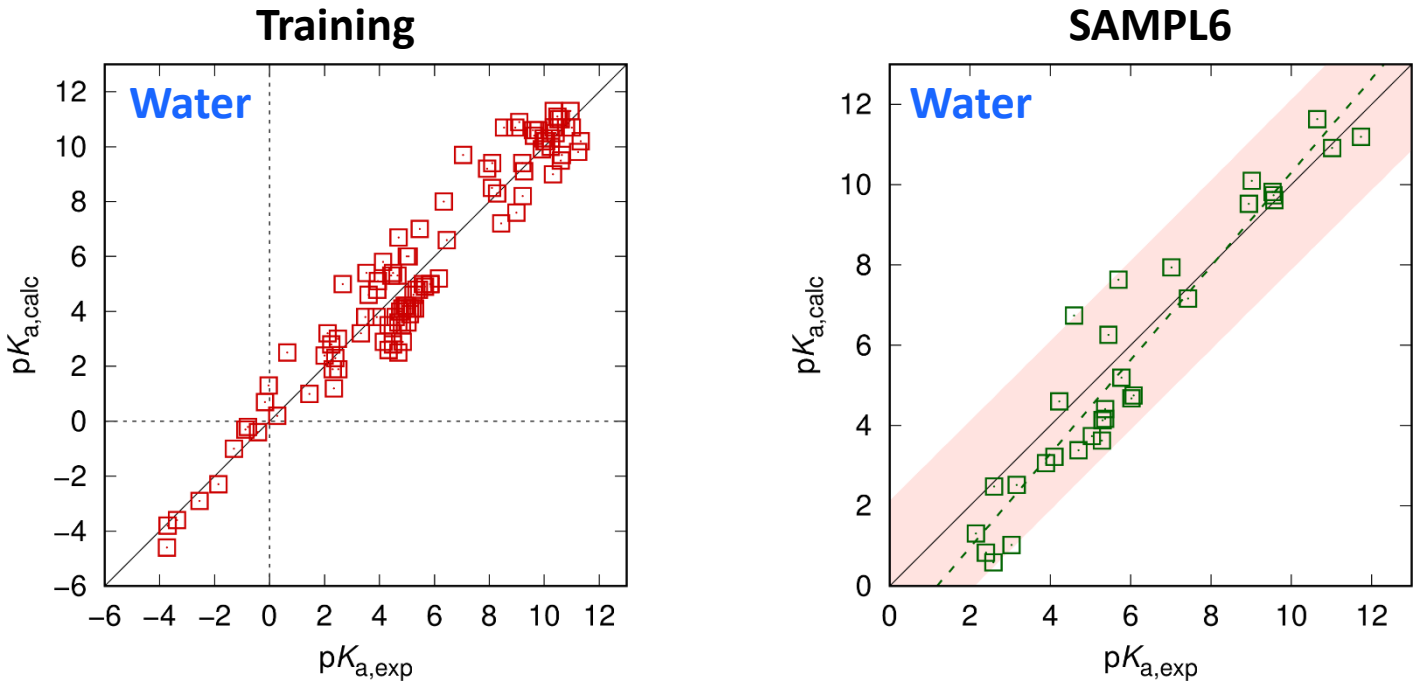


C. P. Kelly, C. J. Cramer, D.G. Truhlar, *J. Chem. Theory Comput.* **2005**, *1*, 1133
N. Tielker, L. Eberlein, S. Güssregen, S. M. Kast, *J. Comput. Aided. Mol. Des.* **2018**, *32*, 1151

SAMPL6 pK_a challenge

pK_a-prediction

- RMSE of approximately 1.13 log units for the best performing model
- Level of theory MP2/6-311+G(d,p) appears to be sufficient for accurate predictions
- Results of calculations for neutral compounds in water can be reused



J. J. Klicić, R. A. Friesner, S. Y. Liu, W. C. Guida, *J. Phys. Chem. A* **2002**, *106*, 1327

N. Tielker, L. Eberlein, S. Güssregen, S. M. Kast, *J. Comput. Aided. Mol. Des.* **2018**, *32*, 1151

Preliminary work

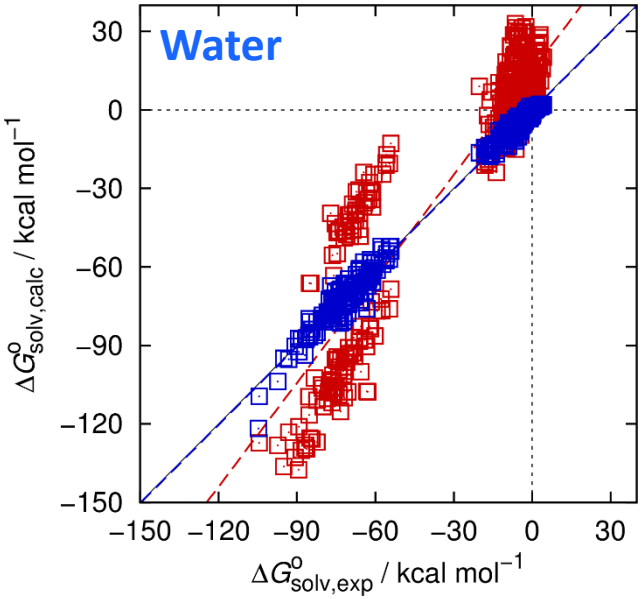
Partial molar volume (PMV) correction for EC-RISM

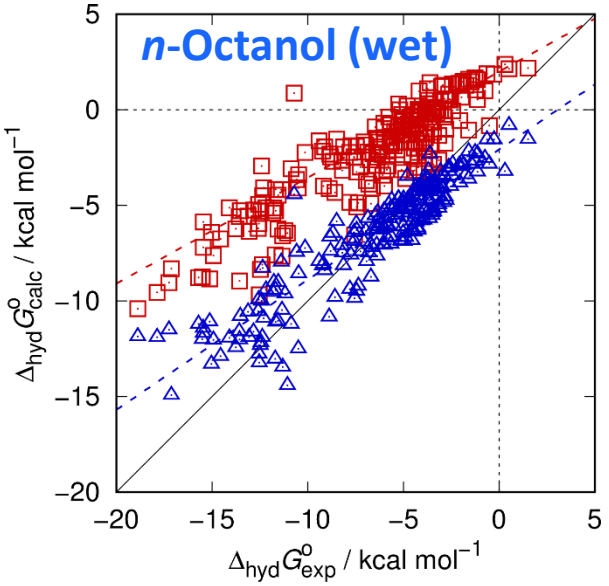
- Calculated chemical potential is typically too high due to overestimation of energy required for cavity formation
- Error shows an almost linear dependence on the solute's partial molar volume

$$\mu^{\mathrm{ex,corr}} = c_\mu \mu^{\mathrm{ex}} + c_V V_\mathrm{m} + c_q q$$

$$\Delta_{\mathrm{solv}} G^\mathrm{o} = E_{\mathrm{sol}} - E_{\mathrm{vac}} + \mu^{\mathrm{ex,corr}} + \Delta\mu^\mathrm{o}$$

MP2/6-311+G(d,p)/EC-RISM//B3LYP/6-311+G(d,p)/PCM





C. P. Kelly, C. J. Cramer, D.G. Truhlar, *J. Chem. Theory Comput.* **2005**, *1*, 1133
N. Tielker, L. Eberlein, S. Güssregen, S. M. Kast, *J. Comput. Aided. Mol. Des.* **2018**, *32*, 1151

Preliminary work

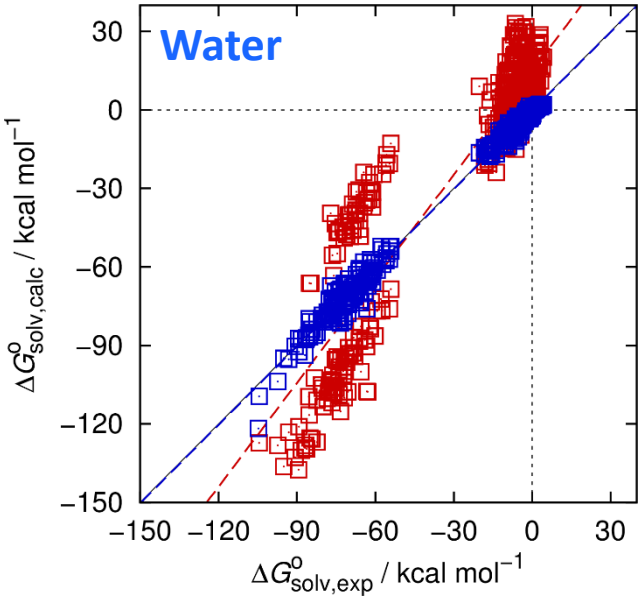
Partial molar volume (PMV) correction for EC-RISM

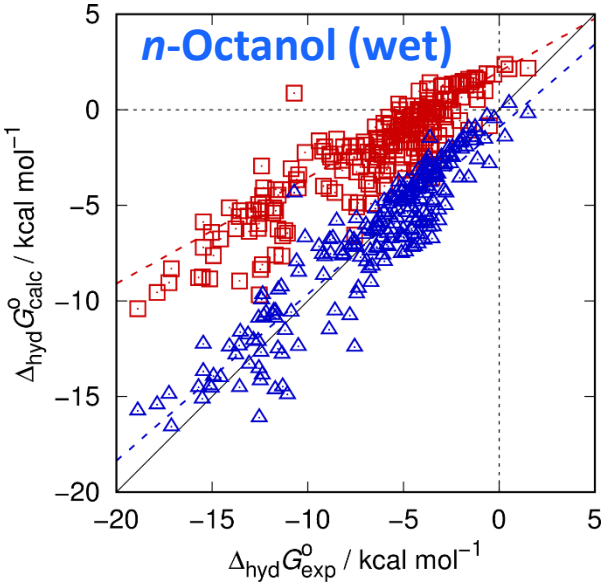
- Calculated chemical potential is typically too high due to overestimation of energy required for cavity formation
- Error shows an almost linear dependence on the solute's partial molar volume

$$\mu^{\mathrm{ex,corr}} = c_\mu \mu^{\mathrm{ex}} + c_V V_\mathrm{m} + c_q q$$

$$\Delta_{\mathrm{solv}} G^\mathrm{o} = E_{\mathrm{sol}} - E_{\mathrm{vac}} + \mu^{\mathrm{ex,corr}} + \Delta\mu^\mathrm{o}$$

MP2/6-311+G(d,p)/EC-RISM//B3LYP/6-311+G(d,p)/PCM



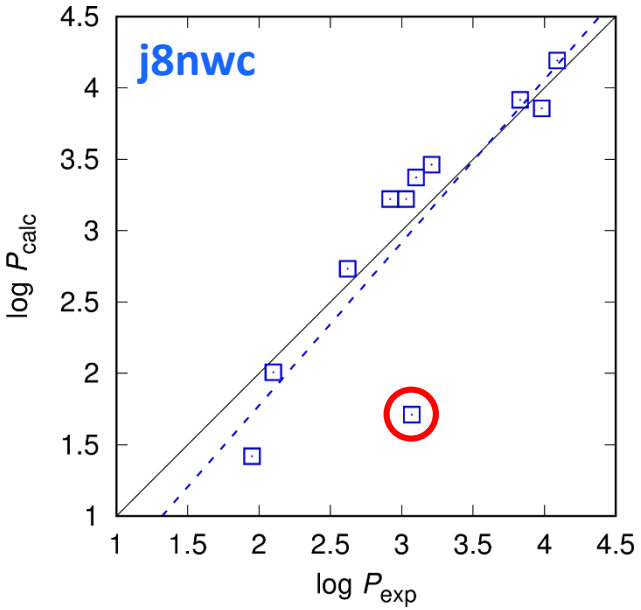


C. P. Kelly, C. J. Cramer, D.G. Truhlar, *J. Chem. Theory Comput.* **2005**, *1*, 1133
N. Tielker, L. Eberlein, S. Güssregen, S. M. Kast, *J. Comput. Aided. Mol. Des.* **2018**, *32*, 1151

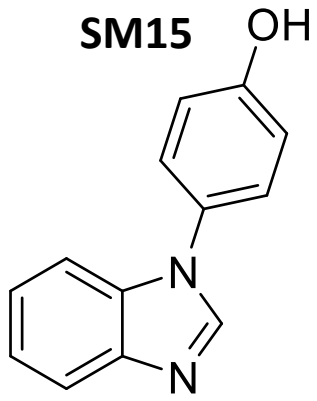
SAMPL6 log *P* challenge

log *P* prediction

- No further training w.r.t. experimental partition coefficients
- RMSE of **0.47 log units** for the best performing model
- Possible reasons for the single outlier?



SM15



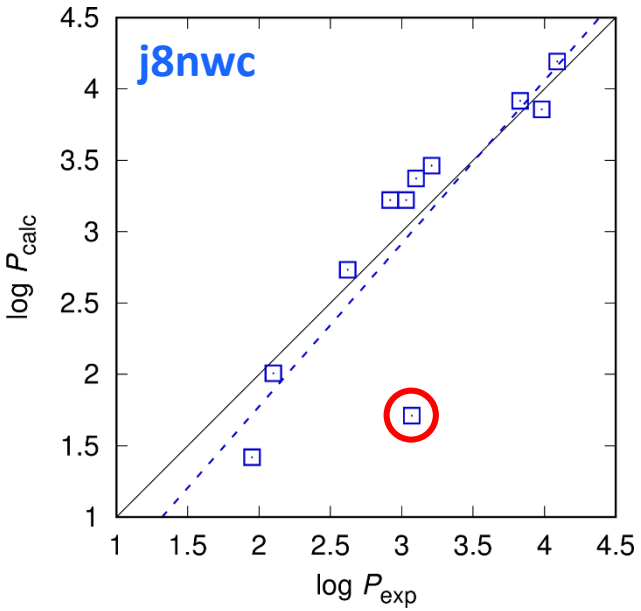
$\Delta \log P = -1.36$

	m	*b*	R^2	RMSE	MAE	MSE
dry octanol (qyzjx)	1.22	-0.51	0.73	0.54	0.45	0.15
wet octanol (j8nwc)	**1.14**	**-0.51**	**0.77**	**0.47**	**0.31**	**0.07**

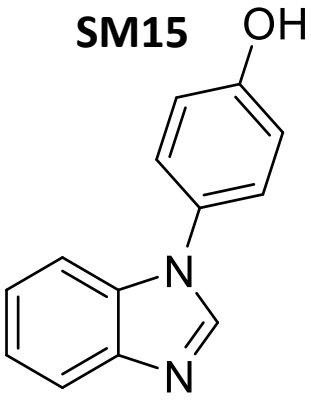
SAMPL6 log *P* challenge

log *P* prediction

- No further training w.r.t. experimental partition coefficients
- RMSE of **0.47 log units** for the best performing model
- Possible reasons for the single outlier?

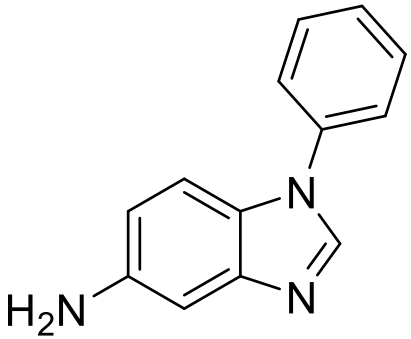


SM15



$\Delta \log P = -1.36$

SM14



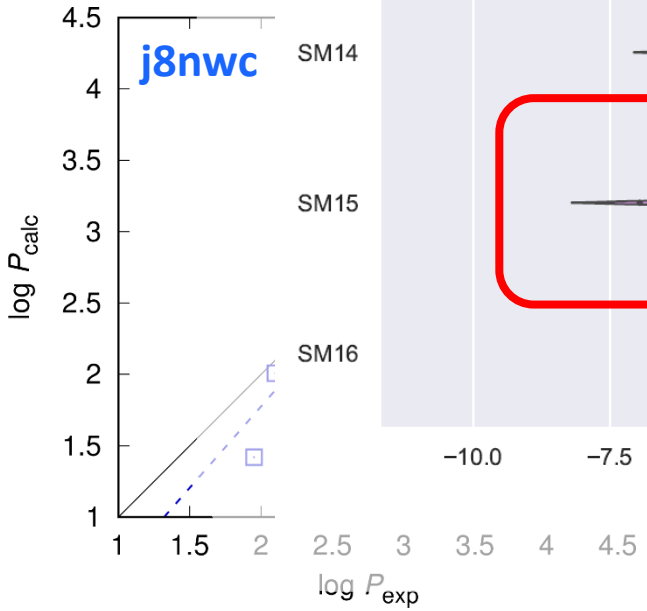
$\Delta \log P = -0.53$

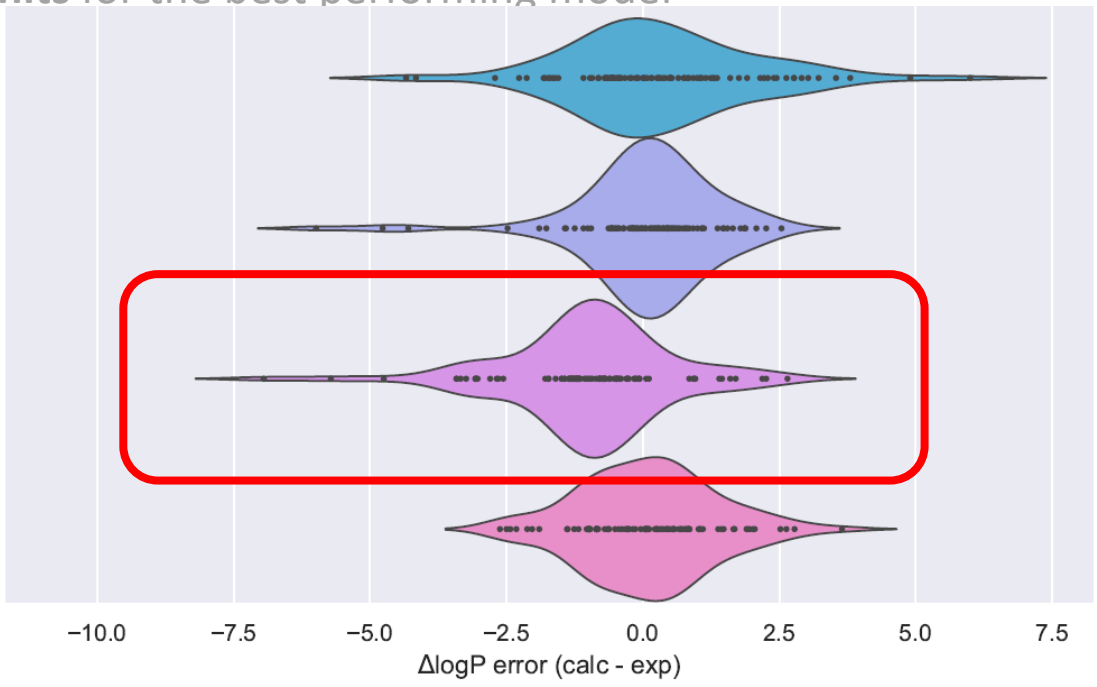
	m	*b*	R^2	RMSE	MAE	MSE
dry octanol (qyzjx)	1.22	-0.51	0.73	0.54	0.45	0.15
wet octanol (j8nwc)	**1.14**	**-0.51**	**0.77**	**0.47**	**0.31**	**0.07**

SAMPL6 log *P* challenge

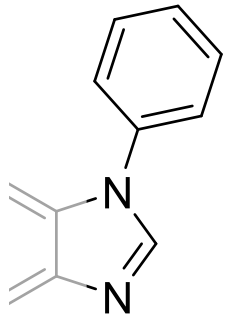
log *P* prediction

- No further training w.r.t. experimental partition coefficients
- RMSE of **0.47 log units** for the best performing model
- Possible reaso





M14



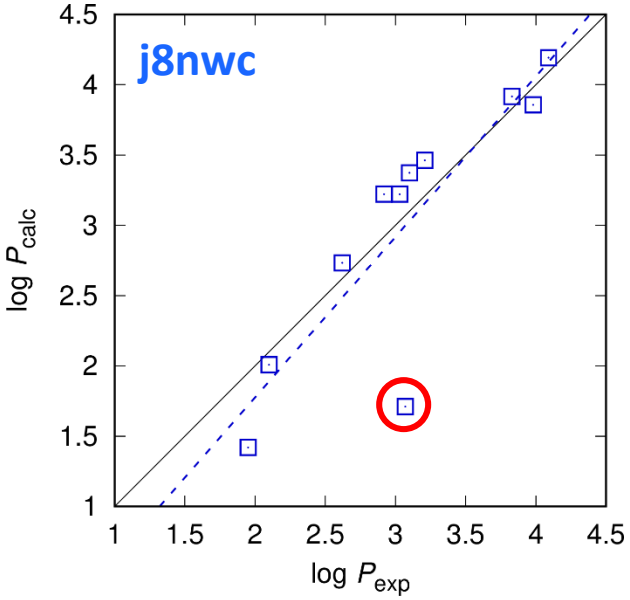
= −0.53

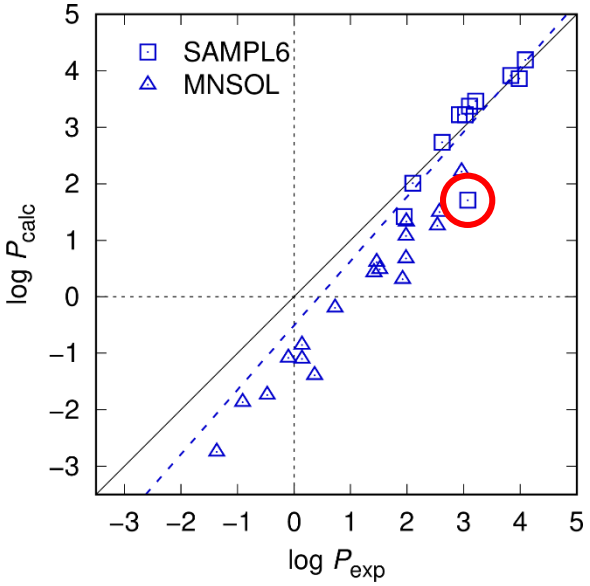
					MAE	MSE
dry octanol (qyzjx)	1.22	-0.51	0.73	0.54	0.45	0.15
wet octanol (j8nwc)	**1.14**	**-0.51**	**0.77**	**0.47**	**0.31**	**0.07**

SAMPL6 log *P* challenge

log *P* prediction

- Other alcohols from Minnesota Solvation Database transfer free energy dataset show similar error as SM15
- Is the error a result of bad prediction of free energy of solvation in water, octanol or both for these molecules?

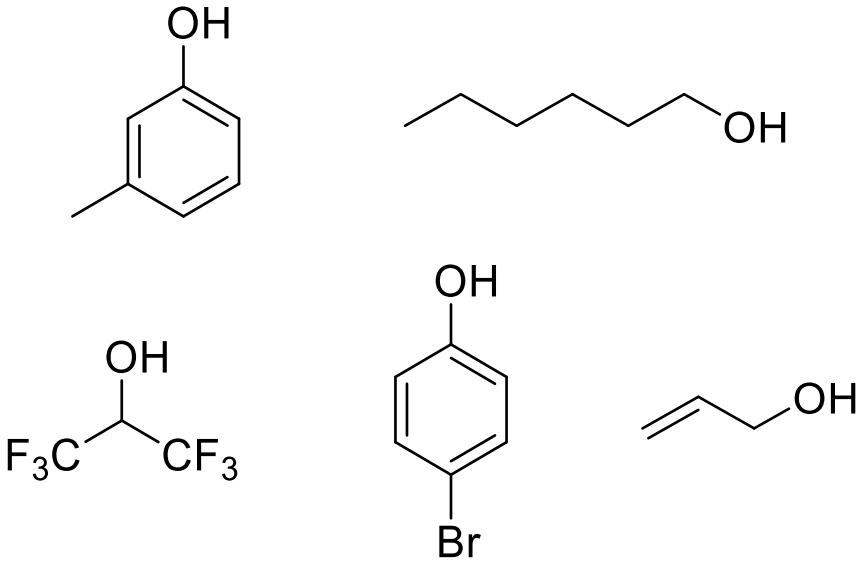


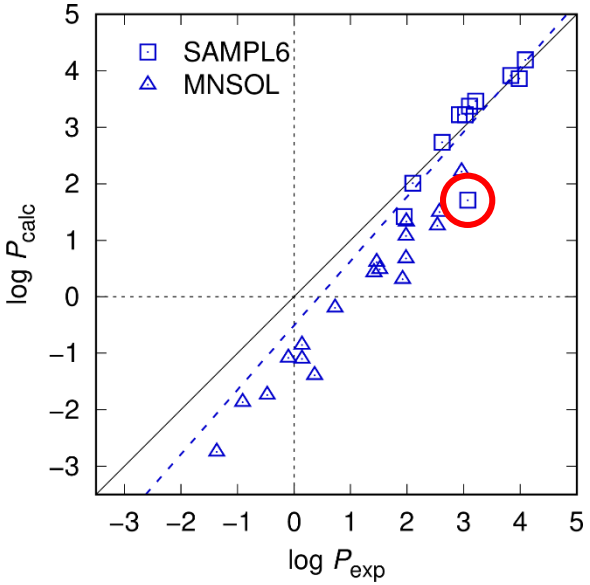


SAMPL6 log *P* challenge

log *P* prediction

- Other alcohols from Minnesota Solvation Database transfer free energy dataset show similar error as SM15
- Is the error a result of bad prediction of free energy of solvation in water, octanol or both for these molecules?



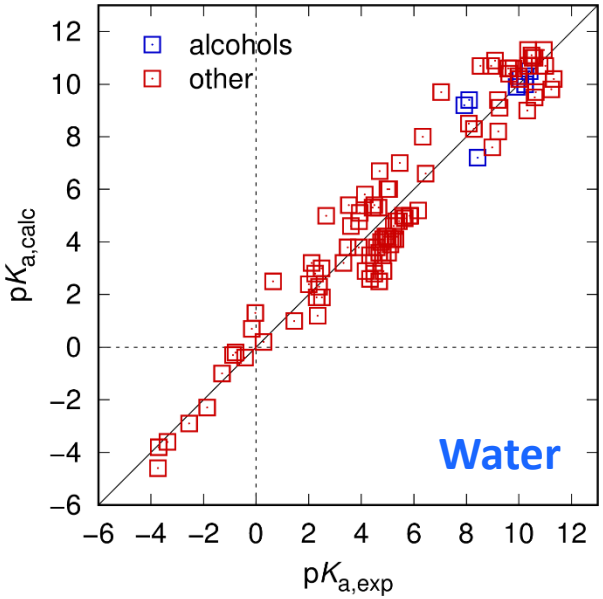


SAMPL6 pK_a challenge

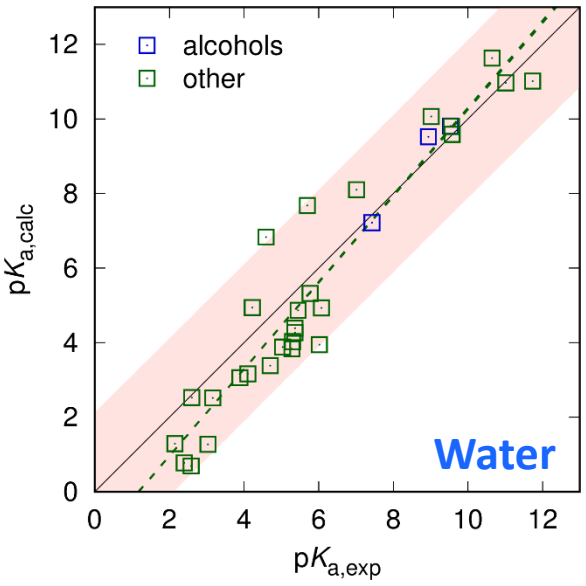
log *P* prediction

- Other alcohols from Minnesota Solvation Database transfer free energy dataset show similar error as SM15
- Is the error a result of bad prediction of free energy of solvation in water, octanol or both for these molecules?

Training



SAMPL6



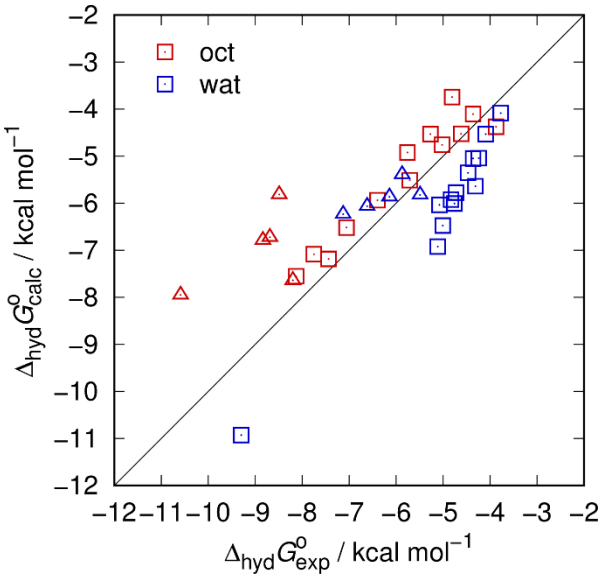
J. J. Klicić, R. A. Friesner, S. Y. Liu, W. C. Guida, *J. Phys. Chem. A* **2002**, *106*, 1327

N. Tielker, L. Eberlein, S. Güssregen, S. M. Kast, *J. Comput. Aided. Mol. Des.* **2018**, *32*, 1151

SAMPL6 log *P* challenge

log *P* prediction

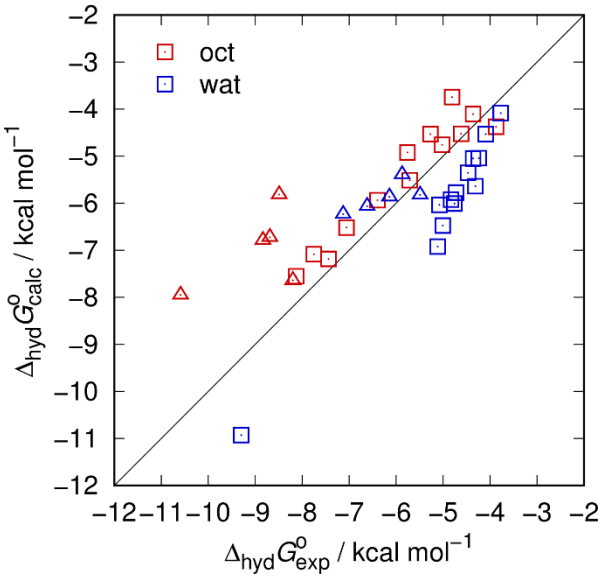
- Other alcohols from Minnesota Solvation Database transfer free energy dataset show similar error as SM15
- Is the error a result of bad prediction of free energy of solvation in water, octanol or both for these molecules?

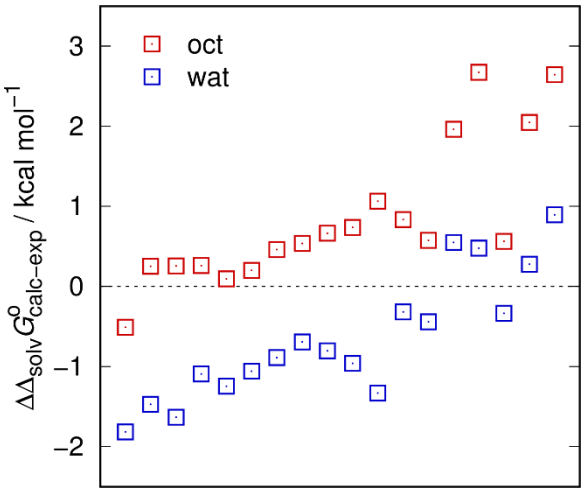


SAMPL6 log *P* challenge

log *P* prediction

- Other alcohols from Minnesota Solvation Database transfer free energy dataset show similar error as SM15
- Is the error a result of bad prediction of free energy of solvation in water, octanol or both for these molecules?





Conclusion

log *P* prediction

- Successful **blind prediction** of experimental log *P* to within 0.5 log units
- Additional parameter necessary for *n*-octanol to achieve chemical accuracy
- "wet" *n*-octanol shows consistently better results than "dry" *n*-octanol

Outlook

- Further investigation of worse performance for aliphatic and aromatic alcohols (GAFF re-parametrization?)
- Should fully *ab initio* prediction of water/*n*-octanol **distribution coefficients** (log *D*) take into account ions in the octanol phase?
- Experimental uncertainty vs. incompletely captured physics?

MAE by submission