SAMPL6 pKa Challenge

Mehtap Isik

Chodera Lab, MSKCC

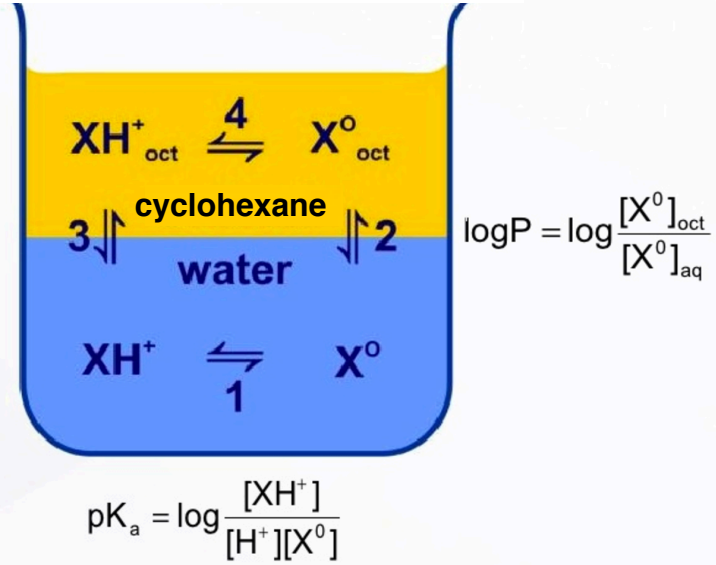
D3R 2018 Workshop

February 22nd, 2018

Why did we decide to organize a blind pKa prediction challenge?

SAMPL5 logD challenge indicated the impact of prediction of the ionization state distribution on the accuracy of logD predictions.

$$logD = log \frac{[X^0]_{oct} + [XH^+]_{oct}}{[X^0]_{aq} + [XH^+]_{aq}}$$



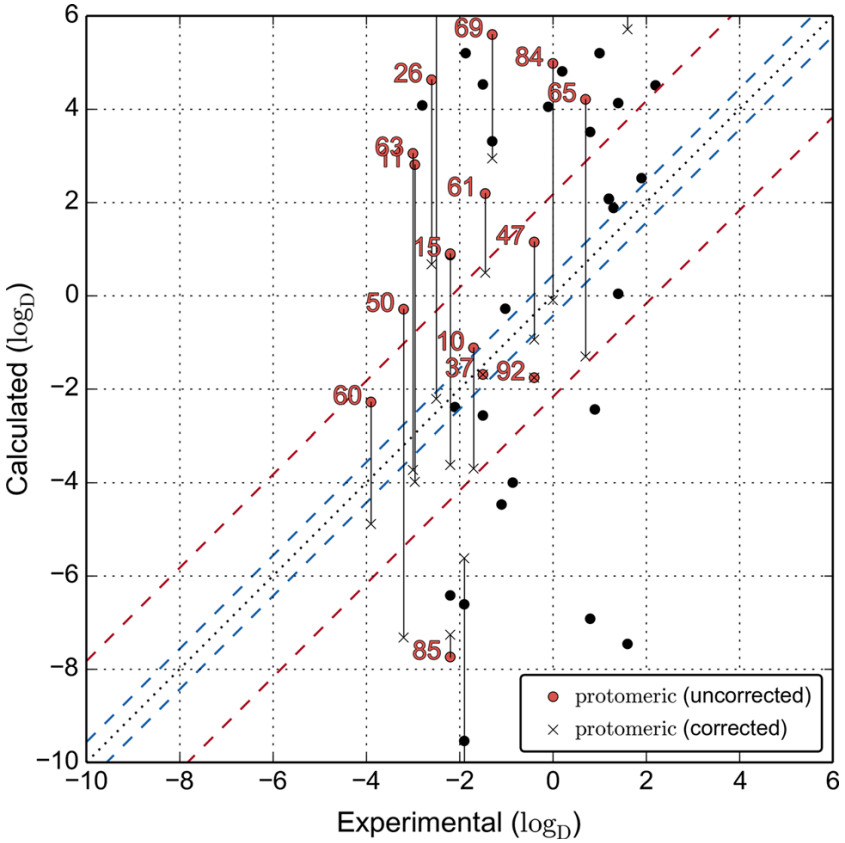
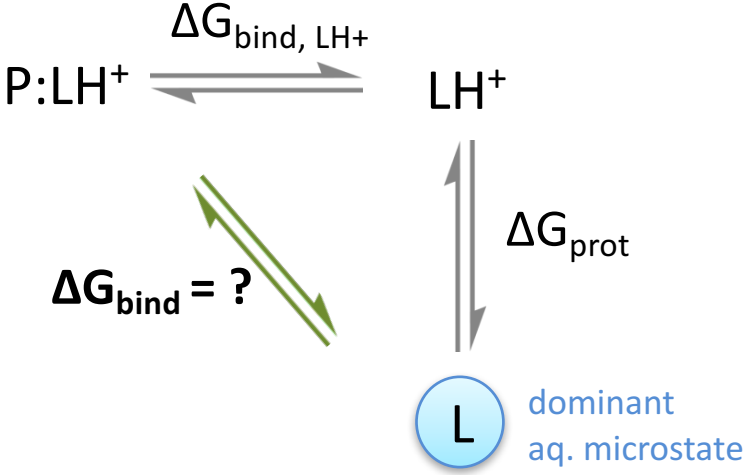


Fig. 4 Our partition estimations from MM BAR (submission 38) plotted against experiment. We have applied our QM based free energy corrections (adiabatic/absolute scheme, submission 10), shifting the predicted values towards more hydrophilic values. These corrections account for multiple protomeric states and for ligand ionization due to the presence of protonizable groups. These corrections substantially reduce the RMSD and increase the correlation of these predictions with respect to experimentally determined values

Pickard, F. C. et al. Blind prediction of distribution in the SAMPL5 challenge with QM based protomer and pKa corrections. Journal of Computer-Aided Molecular Design 30, 1087–1100 (2016).

pKa predictions contribute to the errors in binding free energy predictions.

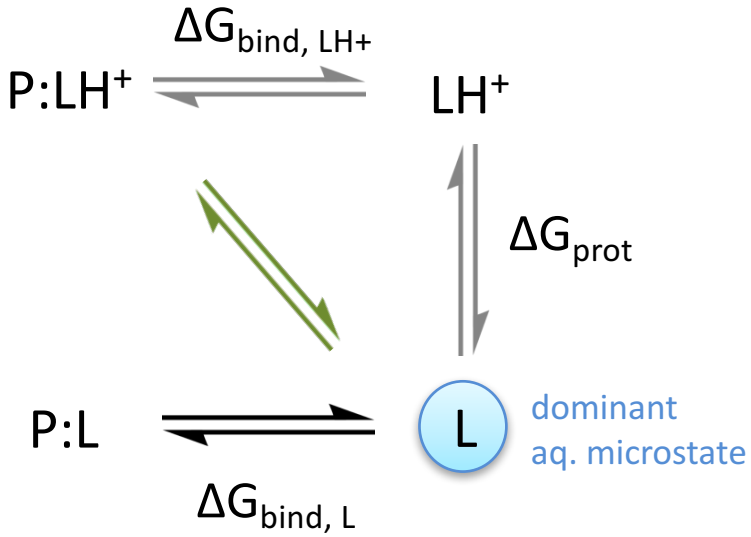
Case 1: Failing to correct binding free energy with pKa penalty



$$\Delta G_{bind} = \Delta G_{bind, LH+} + \Delta G_{prot}$$

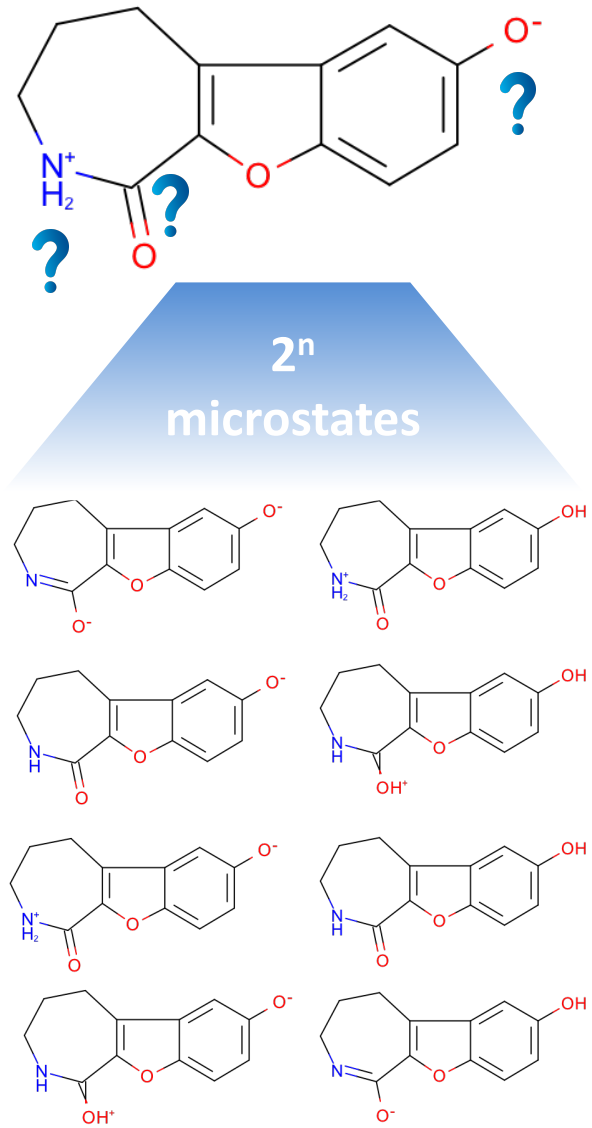
1.38 kcal/mol error to free energy per 1 unit of error in pKa

Case 2: Failing to predict ligand protonation state in the complex



Modeling the wrong ionization state in complex can cause severe errors.

Accurate prediction of pKas is a useful tool for computer aided drug design and lead optimization.



During lead optimization pKa values guide

- Improving target potency
- Reducing potency against undesired target
- Modulating solubility and lipophilicity
- Improving ADME properties

Predicting pKas of of drug-like compounds are challenging due to

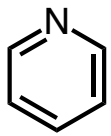
- multiple protonation sites
- conjugated systems and heterocycles

We selected fragment-like molecules with heterocycles common in kinase inhibitors for SAMPL6.

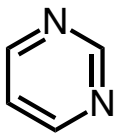
Starting point: **ZINC15 kinase subset** and **anodyne** compounds

Frequent heterocycles found in FDA-approved kinase inhibitors

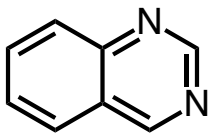
Murcko ring fragmentation



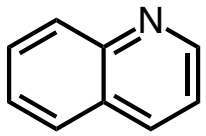
pyridine



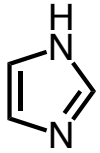
pyrimidine



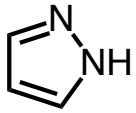
quinazoline



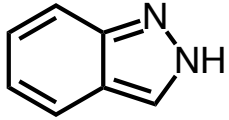
quinoline



imidazole

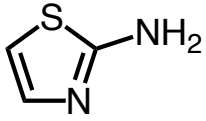


pyrazole



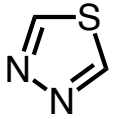
indazole

Other heterocycles in the SAMPL6 set

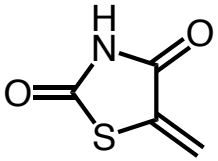


aminothiazole

thiophene

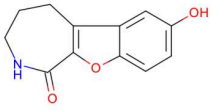


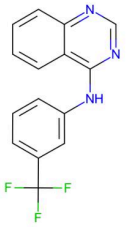
1,3,4-thiadiazole

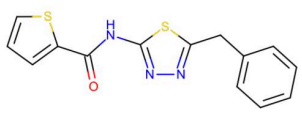


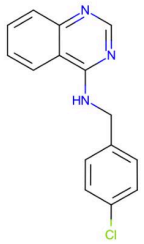
5-methylenethiazolidine-2,4-dione

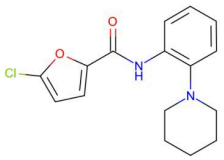
24 compounds are present in SAMPL6 pKa challenge

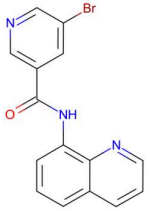

SM01

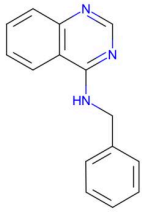

SM02

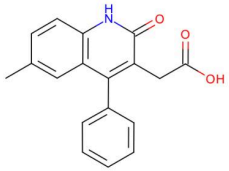

SM03

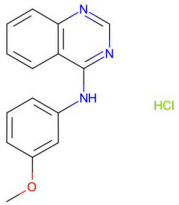

SM04

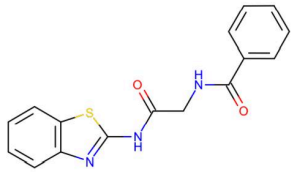

SM05

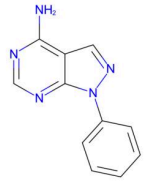

SM06

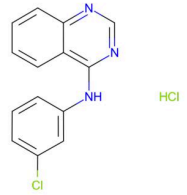

SM07

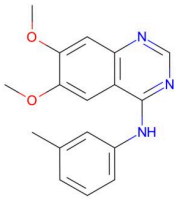

SM08

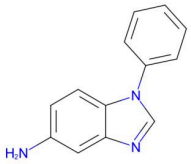

SM09

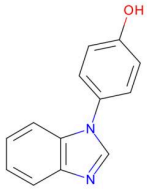

SM10

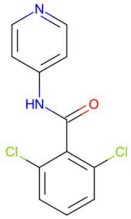

SM11

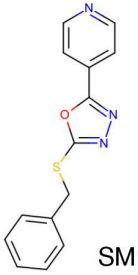

SM12

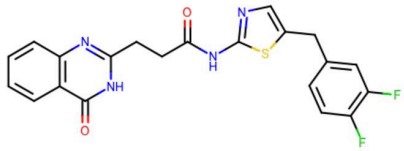

SM13

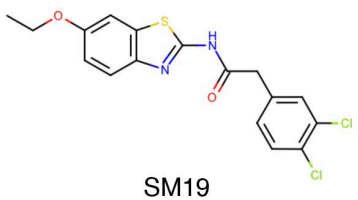

SM14

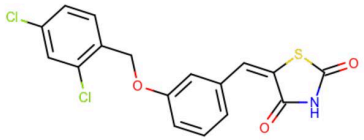

SM15

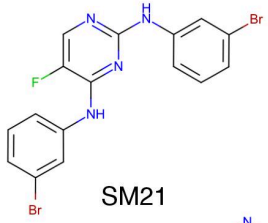

SM16

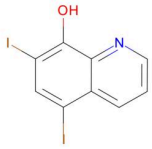

SM17

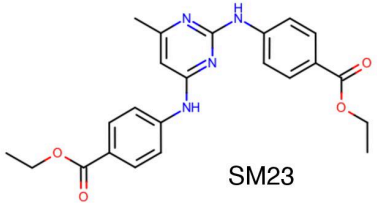

SM18

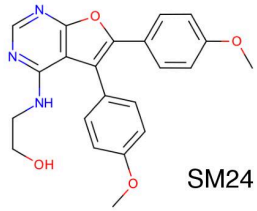

SM19


SM20


SM21


SM22

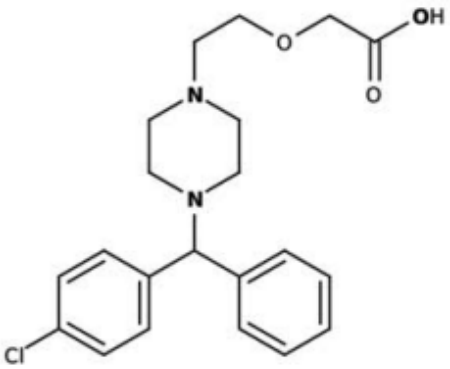

SM23


SM24

In multiprotic compounds it is important to differentiate between macroscopic and microscopic pKas.

Macroscopic pKas

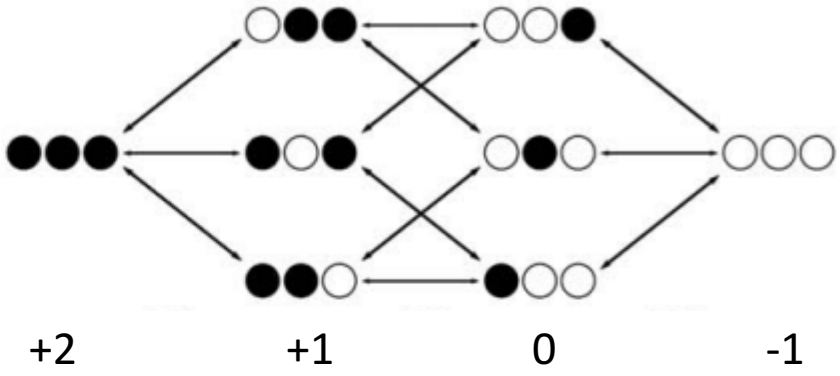
$[HCet^{0}]$ =



$[Cet^{-}]$ ⟷ 8.17 ⟷ $[HCet^{0}]$ ⟷ 3.01 ⟷ $[H_2Cet^{+}]$ ⟷ 2.10 ⟷ $[H_3Cet^{++}]$

UV-metric pKas may fail to capture all macroscopic pKas.

Microscopic pKas



Microequilibrium	pK_a	Microequilibrium	pK_a
●●● ⇌ ○●●	2.59	●○● ⇌ ●○○	5.87
●●● ⇌ ●○●	4.03	●●○ ⇌ ○●○	2.83
●●● ⇌ ●●○	2.28	●●○ ⇌ ●○○	7.62
○●● ⇌ ○○●	4.58	○○● ⇌ ○○○	6.11
○●● ⇌ ○●○	2.52	○●○ ⇌ ○○○	8.17
●○● ⇌ ○○●	3.14	●○○ ⇌ ○○○	3.38

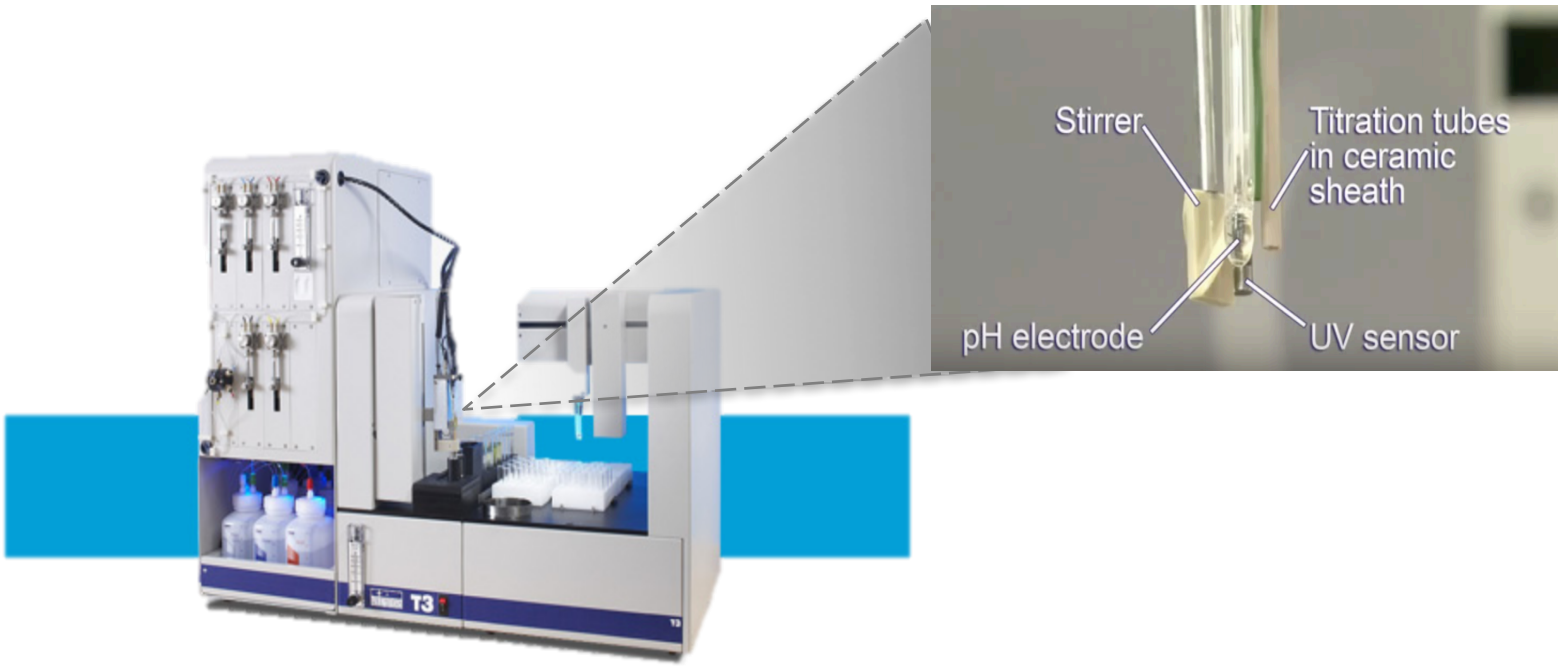
(d) Microconstants. All values are experimentally determined [30].

Fraczkiewicz R. (2013) In Silico Prediction of Ionization. In: Reedijk, J. (Ed.) Elsevier Reference Module in Chemistry, Molecular Sciences and Chemical Engineering. Waltham, MA: Elsevier.

Rupp M. et al. Predicting the pKa of Small MoleculesCombinatorial Chemistry & High Throughput Screening, 2011, 14, 307-327.

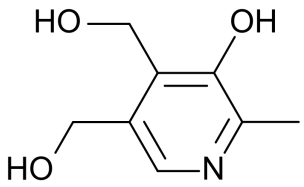
Experimental pKa values for SAMPL6 were measured with Sirius T3.

Dorothy Levorse
Timothy Rhodes

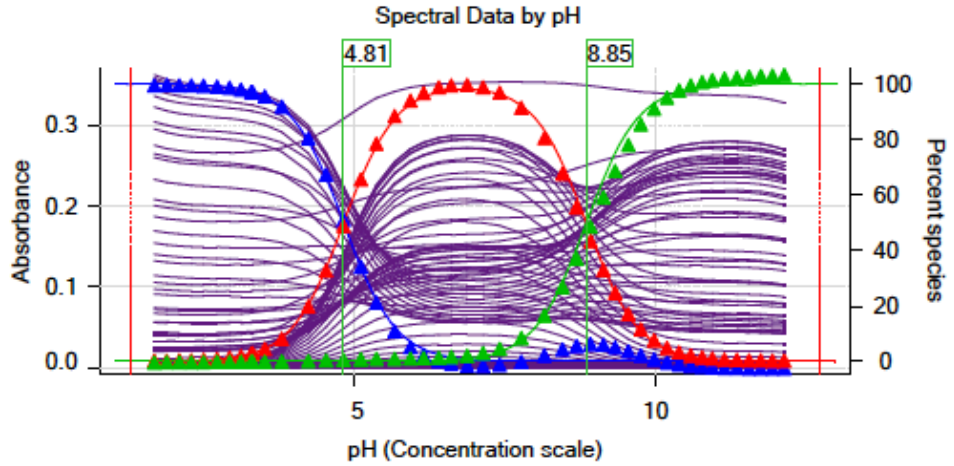


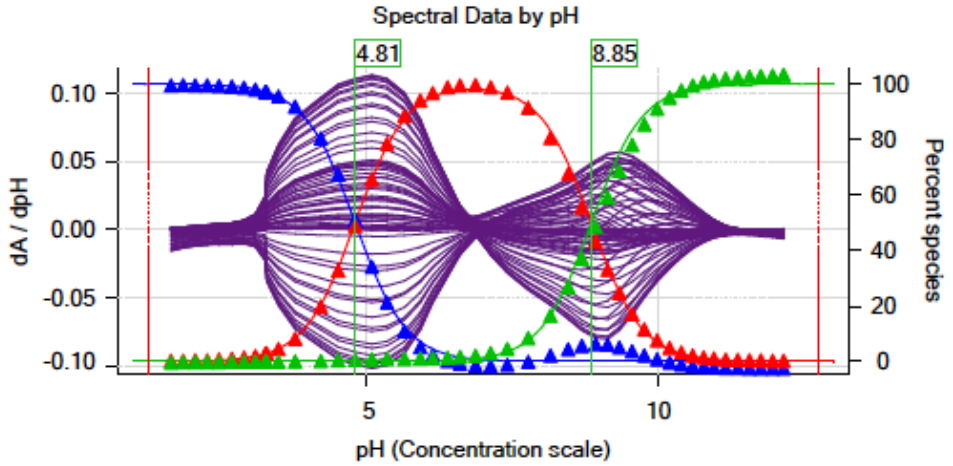
- Method: UV-absorbance spectra based pKa measurement
- Measurement range: 2-12
- 24 small kinase inhibitor fragment-like molecules
- Temperature: 25°C
- Ionic strength: 150 mM KCl solution
- 3 independent replicates (from the same DMSO stock)

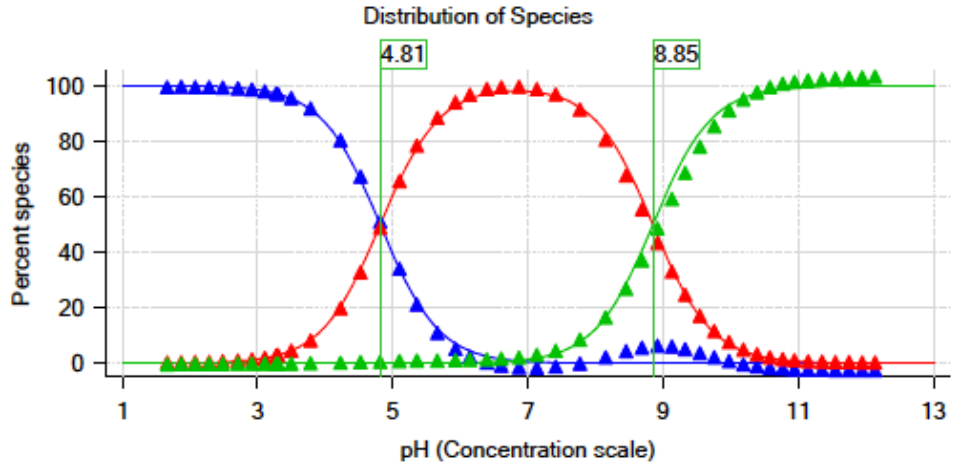
UV-metric pKa measurements of multiprotic compounds lead to determination of macroscopic pKa values.

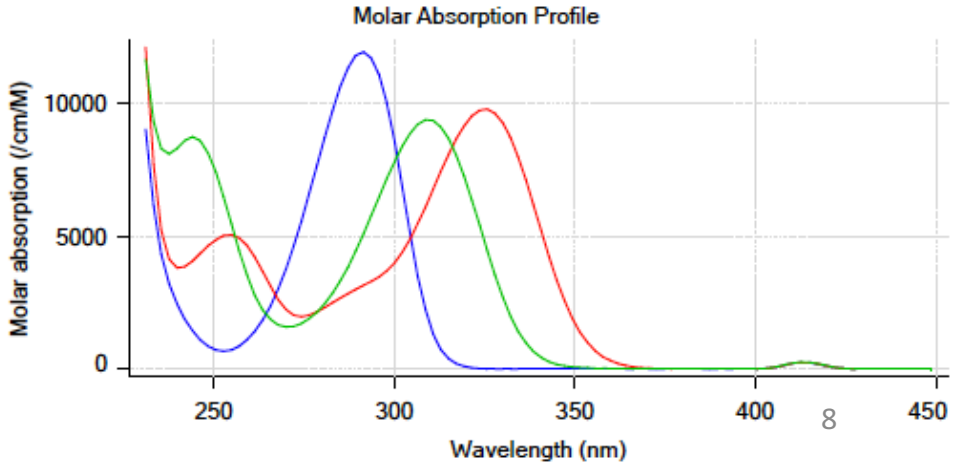


Pyridoxine HCl



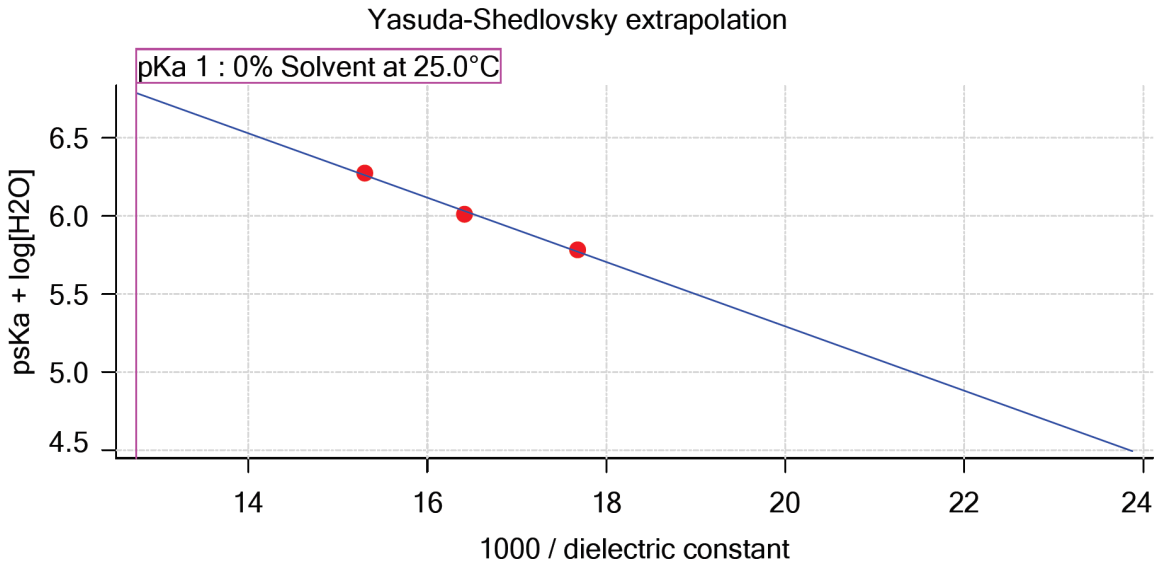






pKas of water insoluble compounds were determined by extrapolation from multiple cosolvent experiments.

Apparent pKa is measured at 3 cosolvent concentrations: 30%, 40%, and 50% MeOH



pKa is determined by **Yasuda-Shedlovsky extrapolation** to 0% cosolvent.

Acid/base assignment based on pKa shift with cosolvent does not provide reliable evidence for assigning pKa values to ionizable groups, especially in multiprotic compounds.

Avdeef, A. et al. PH-metric logP 11. pK a determination of water-insoluble drugs in organic solvent–water mixtures. Journal of pharmaceutical and biomedical analysis 20, 631–641 (1999).

Suggestions for future pKa experimental data collection

- UV-metric pKa measurements with Sirius T3 do not provide any structural information about microstates.
- Acid/base assignment based on pKa shift with cosolvent is not reliable in multiprotic compounds.
- Monoprotic compounds should be prefered if UV-metric or potentiometric methods for pKa measurements will be used.
- Compound purity is critical for accuracy.
- Compound solubility is the limiting factor for pKa measurements with Sirius T3.
- For future pKa challenges with multiprotic compounds, it is ideal to use experimental methods that can measure microscopic pKas, such as NMR.

Submission types and participation to pKa prediction challenge

Type I - microscopic pKas and microstates

Predicting microscopic pKa values and related microstate structures.

32 submissions

Type II - microstate populations as a function of pH

Predicting fractional microstate populations between pH 2 to 12 in 0.1 pH increments.

27 submissions

Type III - macroscopic pKas

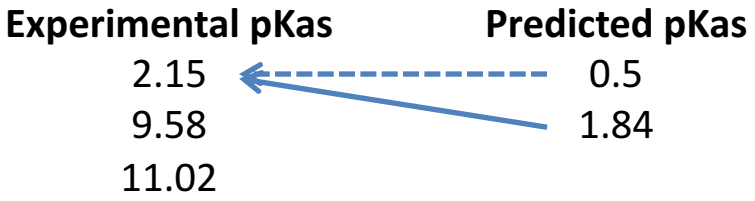
Predicting the value of macroscopic pKas between 2 and 12.

34 submissions

Analysis of macroscopic pKa predictions requires mapping of experimental pKas to predicted pKas.

Closest Method

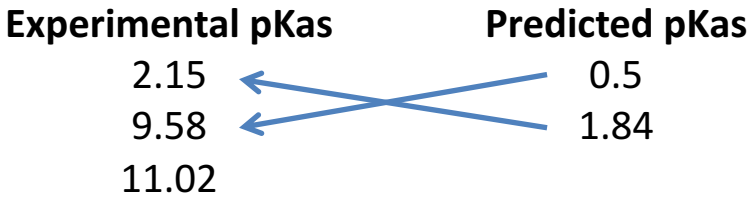
- Each predicted pKa is matched to closest experimental pKa value (min absolute error).
- When more than one predicted pKa match to the same experimental pKa, only the predicted pka that has the lowest absolute error is kept.
- Extra predicted or experimental pKas are ignored.



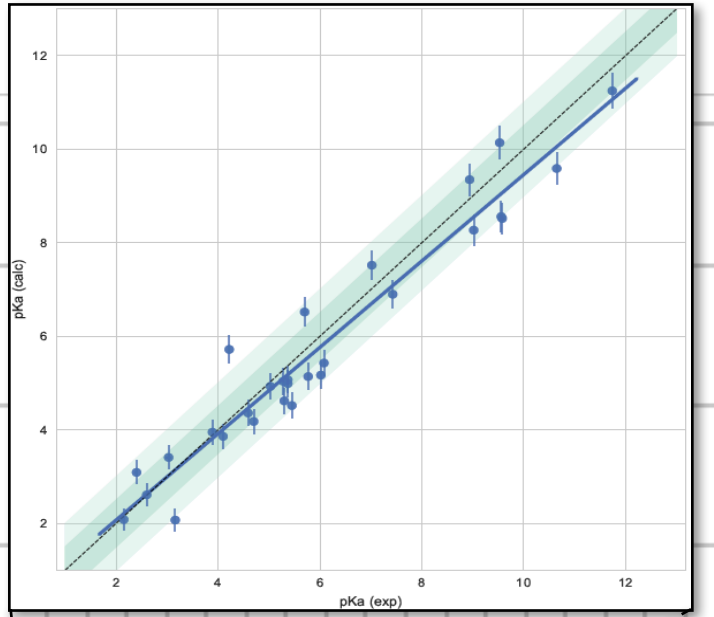
Hungarian Method

- Experimental pKas and predicted pKas are matched following Hungarian algorithm.
- Optimum global assignment that minimizes linear sum of squared errors of all pairwise matches.

Kiril Lanevskij



Overall performance of macroscopic pKa predictions



RMSE values span the range of 0.7-5 pKa units.

RMSE

0
1
2
3
4
5
6
7

xvxzd gyuhx xmyhm yqkga nb007 8xt50 p0jba 37xm8 hytjn q3pfp mkhqa 2ii2g nb001 35bdm nb002 nyzue yc70m 5byn6 y75vj np6b4 w4iyd pwn3m f0gew xikp8 5nm4j ad5pu 0hxtm ds62k ttjd0 mpwiy nb004 nb003 nb005 nb006

Submission ID

Analysis of pKa predictions and future directions

Macroscopic pKa analysis results can be found in SAMPL6 GitHub repository:

https://github.com/MobleyLab/SAMPL6

- Experimental vs predicted pKa value correlation plots
- Error distribution plots for each molecule
- Performance statistics (RMSE, MAE, ME, R^2, slope)

We will keep updating the SAMPL6 repository with:

- Additional performance criteria for pKa predictions
- Analysis of microscopic pKa values and microstate populations

4 participants of the pKa challenge will present us their perspectives.

NEXT

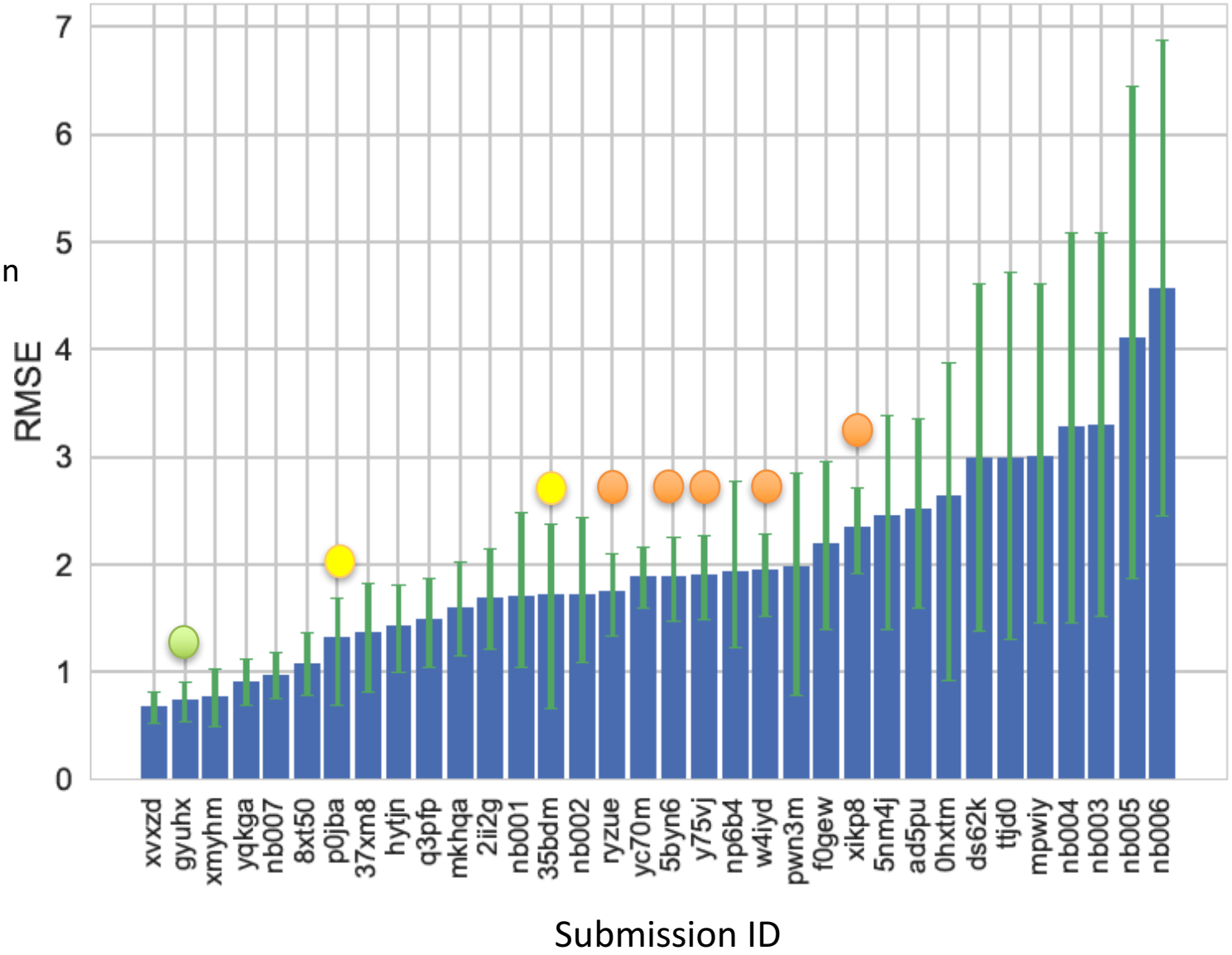
Samarjeet Prasad (Type I)

Qiao Zeng

Marvin Waldman

TOMORROW

Bogdan Iorga



Acknowledgments

SAMPL6 organizers and advisors

David Mobley
John Chodera
Andrea Rizzi
Caitlin Bannan
Bas Rustenburg
Michael Chiu
Michael Gilson
Michael Shirts
Paul Czodrowski

Merck Preformulation Department

Timothy Rhodes
Dorothy Levorse
Brad Sherborne
Heather Wang

Participants of pKa challenge

Caitlin Bannan
Robert Fraczkiewicz
Bogdan Iorga
Stefan Kast
Kiril Lanevskij
Chris Loschen
Philipp Pracht
Samarjeet Prasad
Geoff Skillman
Rainer Wilcken
Qiao Zeng

Tri-Institutional PhD Program in Chemical Biology

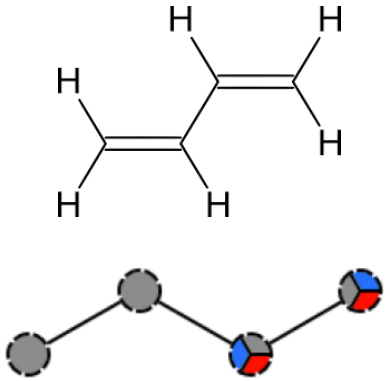
Doris J. Hutchison Fellowship

Selecting kinase inhibitor-like compounds

Chemical and Availability Criteria

- Tier1 compound
- Availability of least 100 mg
- Cheaper compounds in logP bins are prioritized.
- Non-hazardous.
- Anodyne (PAINs and reactive groups removed)

- At least 1 pKa in the interval 3 ≤ pKa ≤ 11
- Multiple pKa's at least 1 log unit part in selected pKa interval.
- Minimum number of UV-chromophore unit: 8
- -1 < XlogP < 6



SMARTS: [n,o,c][c,n,o]cc

Fragment-like

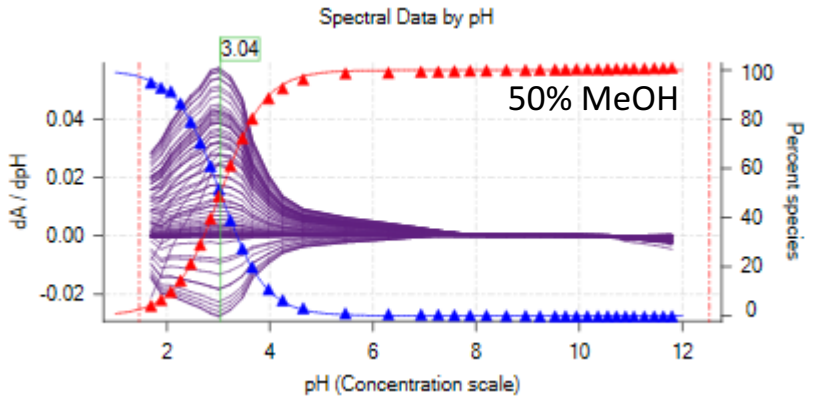
Number of rotatable bonds ≤ 3

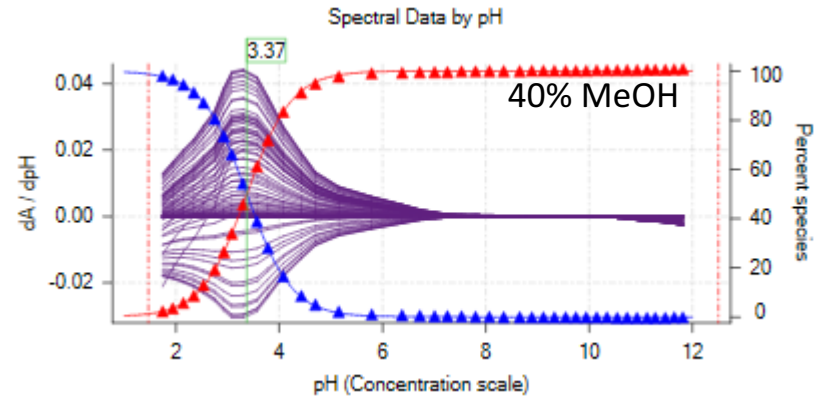
- 150 ≤ mw < 350

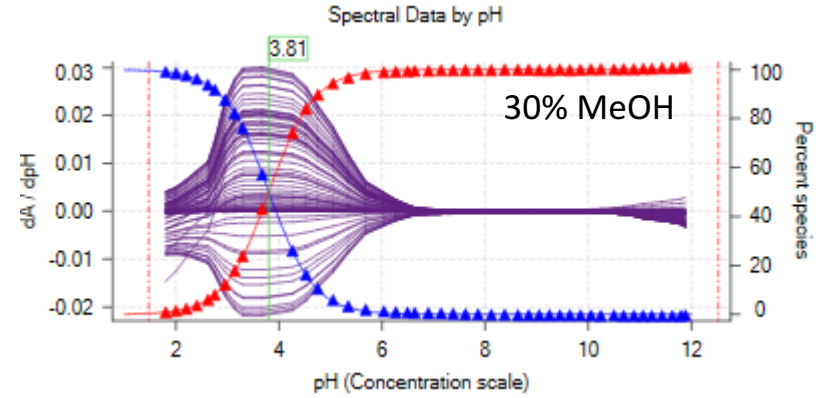
Drug-like

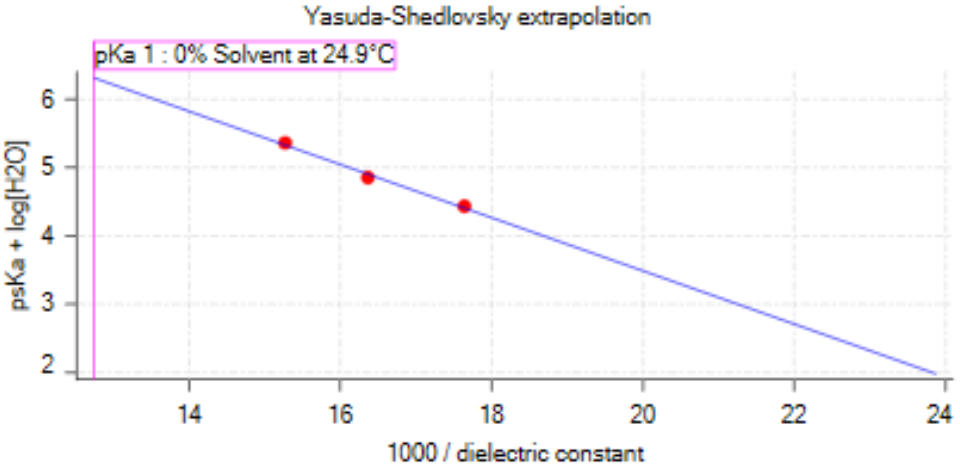
- Number of rotatable bonds ≤ 8
- 350 ≤ mw ≤ 500

pKas of water insoluble compounds were determined by extrapolation from multiple cosolvent experiments.









pKa is determined by **Yasuda-Shedlovsky extrapolation** to 0% cosolvent.

Acid/base assignment based on pKa shift with cosolvent does not provide reliable evidence for assigning pKa values to ionizable groups, especially in multiprotic compounds.

Analysis of macroscopic pKa predictions

Overall performance of macroscopic pKa predictions