**SAMPL6 pKa Challenge Instructions for Human Experts**

Submission deadline: Jan 23, 2018

This challenge consists of predicting macroscopic acid dissociation constants (pKas) of 24 small organic molecules. Our aim is to evaluate how well current pKa prediction methods perform with drug fragment-like molecules through blind predictions.

## 24 small molecules are included in the pKa challenge

24 small molecules included in this challenge are fragment-like small molecules, selected for their similarity to kinase inhibitors and for experimental tractability. Each molecule is assigned a molecule ID in the form of SMXX.

Please refer to **pKa\_challenge\_small\_molecules.jpg** for 2D structures of the molecules. Additionally, canonical isomeric SMILES of each molecule are proved in **molecule\_ID\_and\_SMILES.csv** file.

## Instructions and Submission Template

Predict the value of macroscopic pKa(s) of each molecule between values 2 and 12.

Report your best guess of pKa predictions for the chemical structures provided. You are free to use any method but you must report the process you used to reach these predictions. Human expert participants must report their macroscopic pKa predictions using type III submission template.

Empty submission template file: **typeIII\_macroscopic\_pKas.csv**

Example submission file: **typeIII-ExampleSubmissionMehtapIsik-1.csv** file is provided to illustrate expected format when filling submission template.

* You can manipulate this file in Microsoft Excel or a text editor.
* Lines beginning with a hash-tag (#) are included as comments. These and blank lines will be ignored in the analysis.
* Fill one template file for all predicted molecules with one method.
* You may submit predictions from multiple methods, but you should fill a separate template file for each different method and increment the integer at the end of the file name.

**Predictions section**

* For each molecule, report as many macroscopic pKas as your method predicts. For each macroscopic pKa create a new line that starts with molecule ID as identifier.
* Columns of predictions table: Molecule ID, macroscopic pKa, macroscopic pKa SEM
* Report pKa values to two decimal places (e.g. 10.71).
* Reporting the standard error of the mean (SEM) is optional and encouraged. If it is reported, SEM should be reported to two decimal places (e.g. 1.02).
* For values for which you don't have an estimate, leave that cell of the csv table empty.
* Functional groups you associate the pKa values with must not be reported.

**Name section**

Please provide an informal yet informative name of the method used.

**Software section**

Report all major software packages used and their versions. If you haven’t used any software report the word ‘None’.

**Methods section**

* Use free text in this section to report your prediction methodology and computational details. Level of detail should be at least that used in a publication. Minimum of 50 words is required, although typically a detailed description will require a longer text.
* Please report the process you used to reach these predictions: Did you make predictions based on your intuition or knowledge of experimental pKas? Have you used any tools to help your prediction? Please include details of any software(s), database(s), experimental data you used in addition to or as a guide to your chemical expertise and intuition.
* If you have utilized your intuition, chemical reasoning or experimental knowledge for pKa predictions, please use the last paragraph of the methods section to describe your experience and background.

**Renaming the submission file**

Names of the prediction files must begin with the name of the submission type (i.e., typeIII), and your name and must end with an integer indicating set of prediction. For example, if you want to submit one prediction file for type III prediction (macroscopic pKas), you would name it typeIII-myname-1.csv, where myname is arbitrary text of your choice. If you submit two prediction files of the same submission type, you would name them typeIII-myname-1.txt and typeIII-myname-2.txt. The file will be machine parsed, so correct formatting is essential. Files with the wrong format will not be accepted.

## Uploading your predictions

Your predictions must be uploaded on the D3R SAMPL6 webpage by **January 23, 2018**. The experimental results will be released immediately after the challenge closes. You must use the provided templates to upload your predictions to the [SAMPL website](https://drugdesigndata.org/about/sampl6/pka-prediction): https://drugdesigndata.org/about/sampl6/pka-prediction

Please follow the instructions under Submissions tab. You will be directed to register and login and then you can upload your files and declare your decision about anonymity.

Please use the template provided, as the predictions will be parsed and analyzed with automated scripts. Please include all predictions made with same method and same submission type in one file.

We encourage submitting predictions for all 24 molecules when possible. Incomplete submissions - such as for a subset of compounds - will also be accepted, but will not necessarily be evaluated together with the rest of the submissions. However, we would emphasize that omission of SEM estimates will not cause a submission to be regarded as incomplete, though we highly encourage including such estimates.

## Anonymous versus public participation

When you upload your submission, you will have the option of having it treated anonymously. Anonymous submission means that we may report on your predictions and methods, but not your identity. Public participation means we may also include identifying information. Please note that, although we will work to protect the identity of anonymous participants, we cannot make any guarantees. You may use the D3R website to change your submission’s anonymous/public status until the challenge has closed. However, after the challenge has closed, the anonymous/public status can no longer be changed.

## SAMPL6 workshop February 22-23, 2018

Participants are invited to share and discuss their results, as well as the D3R and SAMPL projects more broadly, at the second in-person D3R and SAMPL workshop, which is scheduled for February 22-23, 2018, at UC San Diego, La Jolla, CA. Note that the workshop immediately follows the Biophysical Society National Meeting in San Francisco.

## Experimental details

pKa measurements were collected using spectrophotometric pKa measurements with a Sirius T3 instrument by Mehtap Isik from the Chodera Lab at MSKCC with the support of the Merck Rahway, Preformulation Department, especially Dorothy Levorse, Timothy Rhodes and Brad Sherborne.

Small molecules were purchased in powder form. 10 mg/ml DMSO solutions were prepared and used as stock solutions for the preparation of samples, where 1-5 uL of 10 mg/ml DMSO stock solution is diluted in 1.5 mL ionic-strength adjusted water (0.15 M KCl). pH titrations with acid (0.5 M HCl, 0.15 M KCl) and base (0.5 M KOH, 0.15 M KCl) and cosolvent addition (80% methanol, 0.15 M KCl) were performed in an automated fashion with with a [Sirius T3 instrument (Sirius Analytical)](http://www.sirius-analytical.com/products/t3).

[The UV-metric pKa measurement protocol of the Sirius T3](http://www.sirius-analytical.com/science/pka) measures the change in multiwavelength absorbance in the 250-450 nm UV region of the spectrum while the pH is titrated between pH 1.8 and 12.2 to evaluate pKas[1,2]. A protonation state change of titratable sites near chromophores will modulate the UV absorbance spectra of these chromophores, allowing populations of distinct UV-active species to be resolved as a function of pH. To do this, basis spectra are identified and populations extracted via analysis of the pH-dependent multi-wavelength absorbance. The number of pKas is determined based on the quality of fit between experimental and modeled microstate pH-dependent populations.

This method is capable of measuring pKas between 2 and 12 when protonatable groups are at most 4-5 atoms away from chromophores such that a change in protonation state alters the absorbance spectrum of the chromophore. We have selected compounds where titratable groups are close to potential chromophores (generally aromatic ring systems), but it is possible that our experimental results will not detect the titration of certain cites (and will thereby neglect the pKa of those sites) if a titratable group is not proximal to a UV-chromophore.

pKa measurements of soluble compounds were performed in ionic-strength adjusted water with 0.15 M KCl. Visual inspection of samples and inspection of UV spectra at 500-600 nm was used to verify no detectable precipitation occurred during the course of measurement. For compounds with insufficient solubility to accurately determine a pKa directly in a UV-metric titration, a cosolvent protocol was used in which three UV-metric titrations were performed in different cosolvent:water ratios (typically 30%, 40%, and 50% methanol) and the Yasuda-Shedlovsky extrapolation method[3] was used to estimate the pKa at 0% cosolvent.

Three replicate pKa measurements were made for all compounds at room temperature (25°C). Replicate measurements were set up from the same compound stock solutions (~10 mg/ml in DMSO) and independent aliquotes were taken to prepare samples for Sirius T3 titration. Multiwavelength absorbance analysis on thw Sirius T3 allows for very good resolution of pKas, but it is important to note that this method produces estimates of macroscopic pKas. If multiple microscopic pKas have close pKa values and overlapping changes in UV absorbance spectra associated with protonation/deprotonaton event, the spectral analysis could produce a single macroscopic pKa that represents an aggregation of multiple microscopic pKas.

[1] Tam, K.Y., and Takács-Novák, K. (2001). Multi-wavelength spectrophotometric determination of acid dissociation constants: a validation study. Analytica Chimica Acta 434, 157–167.  
[2] Allen, R.I., Box, K.J., Comer, J.E.A., Peake, C., and Tam, K.Y. (1998). Multiwavelength spectrophotometric determination of acid dissociation constants of ionizable drugs. Journal of Pharmaceutical and Biomedical Analysis 17, 699–712.  
[3] Avdeef, A., Box, K.J., Comer, J.E.A., Gilges, M., Hadley, M., Hibbert, C., Patterson, W., and Tam, K.Y. (1999). PH-metric logP 11. pKa determination of water-insoluble drugs in organic solvent–water mixtures. Journal of Pharmaceutical and Biomedical Analysis 20, 631–641.