Overview of the SAMPL6 small molecule pK_a prediction challenge

Mehtap Işık[1,2], Ariën S. Rustenburg[1,3], Andrea Rizzi[1,4], Caitlin Bannan[5], Marilyn Gunner[6], David L. Mobley[5], John D. Chodera[1*]

[1]Computational and Systems Biology Program, Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, New York, NY 10065, United States; [2]Tri-Institutional PhD Program in Chemical Biology, Weill Cornell Graduate School of Medical Sciences, Cornell University, New York, NY 10065, United States; [3]Graduate Program in Physiology, Biophysics, and Systems Biology, Weill Cornell Medical College, New York, NY 10065, United States; [4]Tri-Institutional PhD Program in Computational Biology and Medicine, Weill Cornell Graduate School of Medical Sciences, Cornell University, New York, NY 10065, United States; [5]Department of Pharmaceutical Sciences and Department of Chemistry, University of California, Irvine, Irvine, California 92697, United States; [6]Department of Physics, City College of New York, New York NY 10031

***For correspondence:**
john.chodera@choderalab.org (JDC)

Abstract

Complete abstact.

- number of submissions
- summary of analysis
- difficulties observed

Keywords

SAMPL · blind prediction · pK_a · small molecule · macroscopic pK_a · microscopic pK_a · macroscopic protonation state · microscopic protonation state

Abbreviations

SAMPL Statistical Assessment of the Modeling of Proteins and Ligands
pK_a $-\log_{10}$ acid dissociation equilibrium constant
SEM Standard error of the mean

Complete abbreviations

Introduction

Complete introduction section.

Importance of small molecule pKa prediction for pharmaceutical efforts.

Explain why we are doing a pKa challenge and connect to past and previous challenges

SAMPL (Statistical Assessment of the Modeling of Proteins and Ligands). About SAMPL challenges: Collectively, these challenges have assessed the effects of force field accuracy, solvation models, pKa and tautomer predictions.

During the SAMPL5 challenge, log D predictions experienced difficulties predicting log D values accurately, unless protonation states and tautomers were taken into account.

For this iteration of the SAMPL challenge, we have taken one step back and isolated just the problem of predicting solvent protonation states.

This is the first time a blind pKa prediction challenge has been fielded as part of SAMPL. In this first iteration of the challenge, we aimed to assess the performance of current pKa prediction methods and isolate potential causes of inaccurate pKa estimates, with the aim of determining how pKa prediction inaccuracies might impact predicted affinities for drug-like molecules. For example, for both logD and binding affinity predictions, any error in predicting the free energy of accessing a minor protonation state in solution that becomes dominant in the complex will directly add to the error in the predicted transfer or binding free energy.

Challenge goal: determining how pKa prediction inaccuracies might impact predicted affinities for drug-like molecules. For example, for both logD and binding affinity predictions, any error in predicting the free energy of accessing a minor protonation state in solution that becomes dominant in the complex will directly add to the error in the predicted transfer or binding free energy.

Reason for blind pKa challenge: - Impact on binding affinity predictions - Impact on logD predictions (SAMPL6) - Drug-like molecules are especially challenging.

Protonation state effects were a dominant accuracy-limiting factor for logD from SAMPL5, and should also be accuracy-limiting in binding free energy predictions. Errors is pK_a predictions can cause modeling the wrong charge, protonation and tautomerization states which affect hydrogen bonding opportunities and overall dipole moment of the ligand.

Explain the physics of the predicted property

EQUATION: pKa equation

EQUATION: free energy of protonation state equation

Introducing linear protonation state free energy diagram

FIGURE: linear plot of free energy vs pH

FIGURE: a diagram illustrating the ways in which the pKa errors can influence prediction errors for binding affinities

Overview of kinds of pKa prediction methods available (ML, QM, empirical methods ...

Explain challenge design.

Experimental macroscopic pKa values were measured using a UV-metric assay performed using a Sirius T3 [cite exp. paper] supported by Merck, MRL, Rahway NJ.

Figure:

Communicate concepts behind challenge design and why we made specific choices: Explain why we have types I, II, III Explain why we preenumerated microstates

Participants had the option to submit predictions in one of 3 categories: Microscopic pKa values (type I), microscopic state populations (type II), or macroscopic pKa values (type III).

The comparison between macroscopic and microscopic pKa values is not always a straightforward one.

Overview of available pKa prediction methods and methods that participated in SAMPL6. [Reminder to cite all papers here.]

Explain future direction for this challenge

Challenge path: predict pKas, give people pKas to predict logDs on same molecules, then predict for new set of compounds logDs without provided pKas.

Explain potantial benefits of these challenge

Improving computational methods...

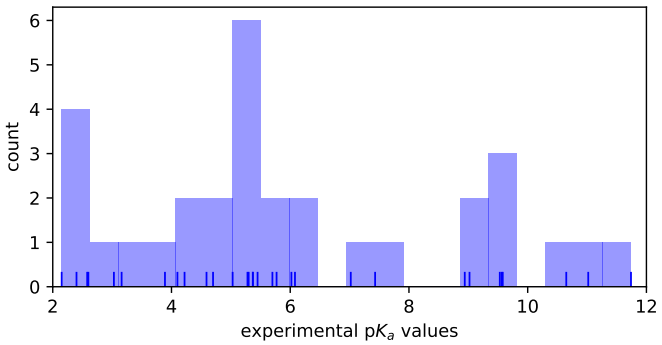

Figure 1. Distribution of experimental pK_a values of 24 compounds in SAMPL6 pKa challenge. Spectrophotometric pK_a measurements were collected with Sirius T3.

Add reference to pKa experiments manuscript.

Five compounds have multiple measured pK_as in the range of 2-12.

Methods

Structure and logistics of the SAMPL6 pKa prediction challenge

Describe the structure of SAMPL6 pKa challenge

- When instructions and input files were made available
- Challenge dates
- Input files
- What to predict? Three type of submissions.
- Multiple submissions allowed
- Predicting the pKa values of the whole set wasn't a requirement.
- 2nd D3R/SAMPL Workshop took place in La Jolla, San Diego on Feb 22-23, 2018.

Analysis metrics for submission performance

- Root mean squared error (RMSE)
- Mean absolute error (MAE)
- Mean Error (ME)
- Square of Pearson Correlation Coefficient (R^2)
- Slope of prediction vs. experimental value linear fit

Uncertainty in each performance statistic was calculated by bootstapping (10,000) to estimate 95% confidence intervals.

Enumeration of requested prediction microscopic protonation states

1. OpenEye (filter out resonance structures), Epik
2. Participant supplied structures

Closest method for matching experimental and predicted pKas

Explain closest method

Hungarian method for matching experimental and predicted pKas

Explain Hungarian method

Matching experimental and predicted pKas based on microstate populations

Figure out how to do this. Bas's network map will be useful.

Comparison of errors/performance against molecular descriptors

Look for correlation with descriptors, and potential explanation for errors. Keep spurious correlations in mind if we have many descriptors.

Results

Analysis of macroscopic pK_a predictions (Type III)

A paragraph to explain the submission methods

Null model 1: pKa prospector lookup

Null model 2: ?

TABLE: Error statistics for all participants

FIGURE: Histograms of macrocopic pKa prediction statistics for all participants

Check if top few performing methods are consistent between error metrics.

FIGURE: Prediction vs experiment scatter plots for top 4-6 methods.)

FIGURE: Violin plots of Delta pKa error to identify compounds that were frequently mispredicted (both closest/hungarian methods)

Explain rationale behind how we analyze the data and determine success/failure

Performance comparison of different methods, grouped by methods class

Analysis of 4-aminoquinazoline series

Assessment of individual methods by each of our analysis methods

Assessment of the groups of methods

Null model for macroscopic pKa predictions

Analysis of microscopic pK_a predictions (Type I & II)

FIGURE: Ranking of microscopic pKa prediction error for all participants

FIGURE: Violin plots of Delta pKa error to identify compounds that were frequently mispredicted (both closest/hungarian methods)

Assessment of individual methods by each of our analysis methods

Assessment of the groups of methods

Does type II predictions capture linear trend in DeltaG vs pH plot

Calculation of predicted macroscopic pKas and comparison to experiment

Performance comparison of different methods, grouped by methods class

Comparison of predicted microstates using consensus set of transitions of high accuracy prediction methods

the consensus of predictions might give some indication here too of whether we're dealing with mono or polyprotic compounds

Analysis of 4-aminoquinazoline series

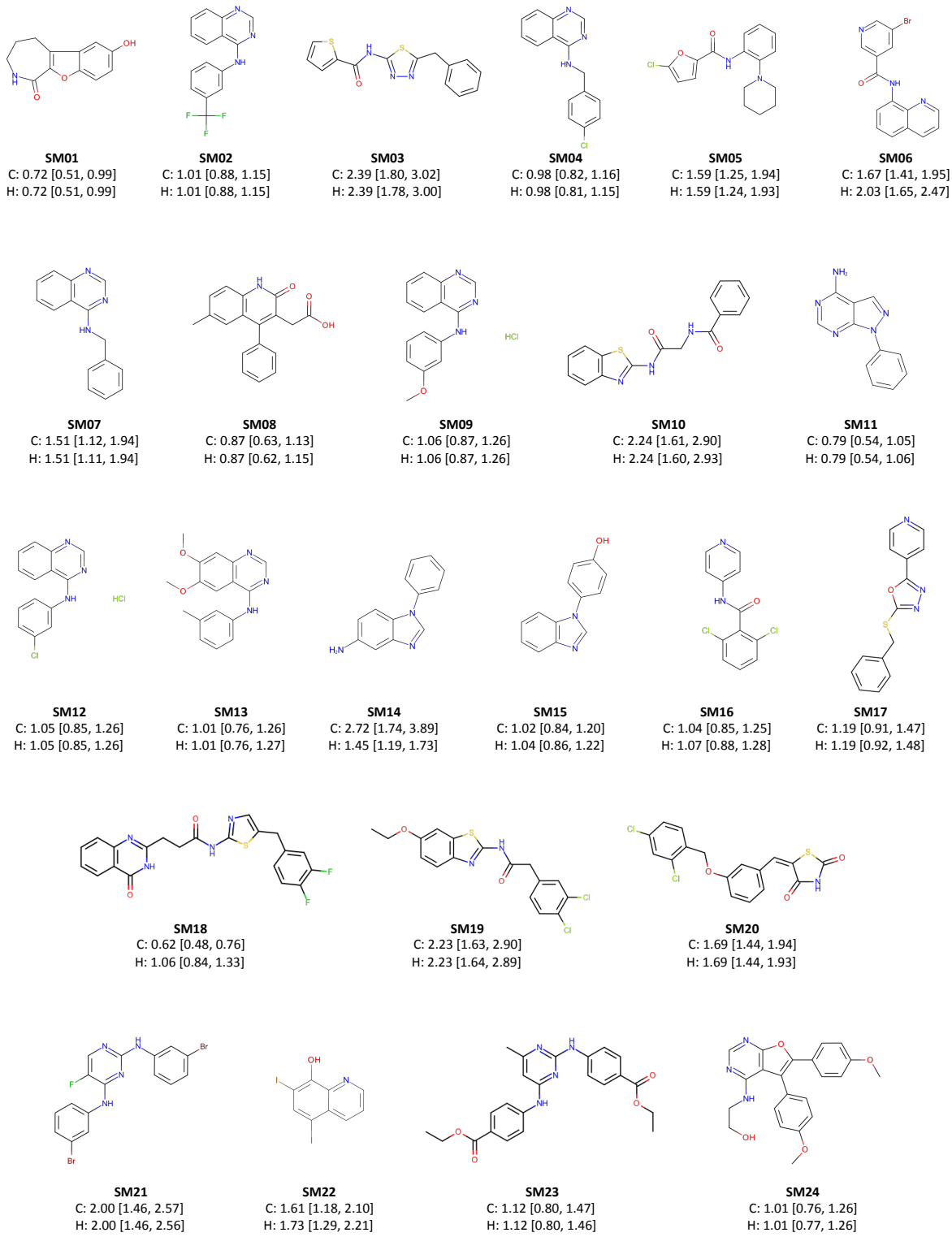

Figure 2. Molecules of SAMPL6 pKa challenge with MAE calculated for all macroscopic (type III) predictions. MAE calculated considering all prediction methods indicate which molecules had the lowest prediction accuracy in SAMPL6 challenge. MAE values calculated for each molecule include all the matched pKa values, which could be more than one per method for multiprotic molecules (SM06, SM14, SM15, SM16, SM18, SM22). "C:" and "H:" indicate results based on Closest and Hungarian matching algorithms employed for pairing experimental and predicted pKa values. Calculated MAE values were observed to differ for molecules with multiple experimental pKas. MAE values are reported with 95% confidence intervals.

How differently do different methods predict microscopic transitions? (method vs method correlation plot to see if methods predict the same microstate pairs or not)

Discussion

Do any methods predict within experimental accuracy (how is the field doing overall)?

Common challenging factors for accurate pKa predictions. Tautomers, Heterocycles etc.

Discussion of matching experimental and predicted values

Difficulty of assessing predicted pKas using experimental data: matching problem

Explain rationale behind how we analyze the data and determine success/failure

Prediction performance of individual molecules

Which chemical structures make pKa predictions more difficult?

SAMPL6 pKa set consisted of only 24 small molecules which limits our ability to do statistical analysis to determine which chemical substructures contribute to greater errors in pKa predictions.

Illustration/explanation of effects where microscopic pKas and macroscopic pKas can differ

Are there any correlations between molecular descriptors and pKa errors?

What can we learn from failures? Which physical effects are driving failures?

How do accuracy limitations in small molecule pKa prediction translate into modeling errors in ligand affinity prediction?

Advice for future challenges

Discuss what can be done to further improve future challenges

How can we maximize what we learn? What should we have people predict? How should we select compounds / measure pKas?

Suggestions about challenge construction

Enumeration of protonation states before predictions (which states does one need to consider?)

Suggestions about challenge analysis

NMR experimental techniques could be used to validate microstate information in future challenges

Conclusion

Code and data availability

- SAMPL6 pK_a challenge instructions, submissions, experimental data and analysis is available at https://github.com/MobleyLab/SAMPL6

Overview of supplementary information

Organized in SI document:

- TABLE SI 1: ???

Extra files:

- Any extra files

Author Contributions

Conceptualization, MI, JDC, CB, DLM ; Methodology, MI, JDC ; Software, MI, AR, ASR ; Formal Analysis, MI, ASR, AR ; Investigation, MI ; Resources, JDC; Data Curation, MI ; Writing-Original Draft, MI, JDC; Writing - Review and Editing, MI, ASR, AR, CB, DLM, JDC; Visualization, MI, AR ; Supervision, JDC, DLM, CB, ASR ; Project Administration, MI ; Funding Acquisition, JDC, DLM.

Acknowledgments

Complete acknowledgments section.

MI, ASR, and JDC acknowledge support from the Sloan Kettering Institute. JDC acknowledges support from NIH grant P30 CA008748. MI acknowledges Doris J. Hutchinson Fellowship. We thank Brad Sherborne for his valuable insights at the conception of the pK_a challenge and connecting us with Timothy Rhodes and Dorothy Levorse who were able to provide resources and expertise for experimental measurements performed at MRL. We acknowledge Paul Czodrowski who provided feedback on multiple stages of this work: challenge construction, purchasable compound selection and manuscript. MI, ASR, AR and JDC are grateful to OpenEye Scientific for providing a free academic software license for use in this work.

Mike Chui

Disclosures

JDC is a member of the Scientific Advisory Board for Schrödinger, LLC. DLM is a member of the Scientific Advisory Board of OpenEye Scientific Software.

References