Accuracy of macroscopic and microscopic pK_a predictions of small molecules evaluated by the SAMPL6 blind prediction challenge

Mehtap Işık (ORCID: 0000-0002-6789-952X)[1,2*], **Ariën S. Rustenburg (ORCID: 0000-0002-3422-0613)**[1,3], **Andrea Rizzi (ORCID: 0000-0001-7693-2013)**[1,4], **M. R. Gunner**[6], **David L. Mobley (ORCID: 0000-0002-1083-5533)**[5], **John D. Chodera (ORCID: 0000-0003-0542-119X)**[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:**
mehtap.isik@choderalab.org (MI)

Abstract

Complete abstract.

- number of submissions [1]
- summary of analysis
- difficulties observed

0.1 Keywords

SAMPL · blind prediction challenge · acid dissociation constant · pK_a · small molecule · macroscopic pK_a · microscopic pK_a · macroscopic protonation state · microscopic protonation state

0.2 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
RMSE Root mean squared error
MAE Mean absolute error
τ Kendall's rank correlation coefficient (Tau)
R^2 Coefficient of determination (R-Squared)

1 Introduction

Complete introduction section: - Importance of small molecule pKa prediction for pharmaceutical efforts. - Definition of pKa - Acid dissociation equilibrium constant - Add pKa equation - Add free energy of protonation state equation - Definition of microscopic and macroscopic pKas - Introduce linear protonation state free energy diagram [Cite Gunner et al 2019 paper] FIGURE: linear plot of free energy vs pH

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.

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.

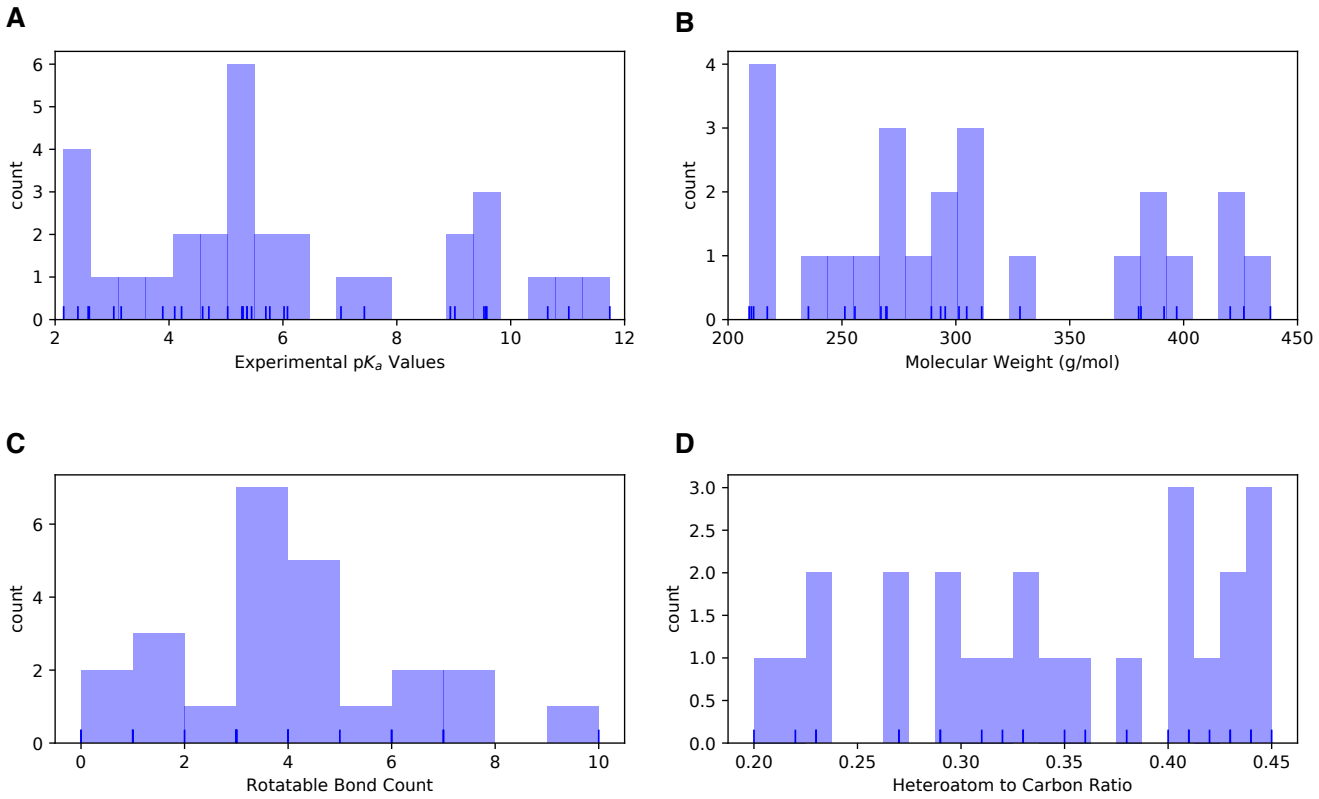

Figure 1. Distribution of molecular properties of 24 compounds in SAMPL6 pK_a Challenge. A Histogram of spectrophotometric pK_a measurements collected with Sirius T3 [1]. Overlayed carpet plot indicates the actual values. Five compounds have multiple measured pK_as in the range of 2-12. **B** Histogram of molecular weights of compounds in SAMPL6 set. Molecular weights were calculated by neglecting counter ions. **C** Histogram of the number of non-terminal rotatable bonds in each molecule. **D** The histogram of the ratio of heteroatom (non-carbon heavy atom) count to the number of carbon atoms.

Explain potantial benefits of these challenge

Improving computational methods...

1.1 Motivation for a blind pKa challenge

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

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: 1. Impact on binding affinity predictions 2. Impact on logD predictions (SAMPL6) 3. Drug-like molecules are especially challenging.

Future challenge direction Challenge path: predict pKas, give people pKas to predict logDs on same molecules, then predict for new set of compounds logDs without provided pKas. Potantial benefits of these challenges: 1. Improving computational methods 2. Detecting hidden contributors to error

1.2 Approaches to predict pKas

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

2 Methods

2.1 Structure and logistics of the SAMPL6 pKa prediction challenge

Describe the structure of SAMPL6 pKa challenge

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.

Communicate concepts behind challenge design and why we made specific choices: 1. Explain why we have types I, II, III 2. Explain why we pre-enumerated 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.

- 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.

Referece Figure **??**. Drug-like molecules are often larger and more complex than the ones used in this study.

2.2 Enumeration of requested prediction microscopic protonation states

1. OpenEye (filter out resonance structures), Epik

2. Participant supplied structures

Microstate pairs: Only +/-1 charge change transitions are allowed. List of allowed transitions. +2 transitions are not considered.

2.3 Evaluation approaches

2.3.1 Statistical 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.

2.3.2 Matching algorithms for pairing predicted and experimental pKas

Explain why it is necessary due to lacking structural information. Cite recommendations from article such as preserving sequence. Experimental data doesn't inform protonation site and overall charge of species. Experimental data doesn't capture the whole picture. We don't know charge and we don't know tautomers. We don't know the charge state of macrostates, this causes a matching problem

Explain Hungarian method for matching experimental and predicted pKas

Explain Closest method for matching experimental and predicted pKas

Explain microstate based matching.

2.4 Reference calculations

Schrodinger Epik Schrodinger Jaguar Chemicalize MoKa

3 Results and Discussion

A paragraph to explain the submission methods. Define method categories: DL, LFER, QSPR/ML, QM, QM+LEC, and QM+MM, Blind predictions, Reference calculations, Null model (pKa prospector lookup)

Submissions spanning different method categories were made to the SAMPL6 pK_a Challenge: database lookup (DL), linear free energy relationship (LFER), quantitative structure property relationship (QSPR), machine learning (ML), quantum mechanics (QM) models with and without linear empirical correction (LEC), and combined quantum mechanics and molecular mechanics (QM+MM). Unique submission IDs were assigned to each submission. Table 1 matches method names with submission IDs. Unique IDs were also assigned when multiple submissions exists for different submission types of the same method such as microscopic pK_a(type I) and macroscopic pK_a (type III).

3.1 Analysis of macroscopic pK_a predictions (Type III)

MI: Methods are indicated by submission IDs.

Refer to SI TABLE: Error statistics for all participants. Refer to SI FIGURE: Error distribution ridge plots for each method (exp-pred macroscopic pKa). Which methods tend to overestimate and which methods tend to undestimate?

MI: SI TABLE: Error statistics for all participants

Describe number of missing and extra pKa for each method. Report in total for all molecules how many predicted pKas are there and how many experimental pKas. Refer to FIGURE: missing and extra pKa counts.

MI: SI TABLE: Missing and extra pKa counts

Describe overall performance comparison of different methods, grouped by methods class.

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

Table 1. Submission IDs, names, category, and type for all the pK_a prediction sets. Reference calculations are labeled as *nb###*. The method name column lists the names provided by each participant in the submission file. The "type" column indicates if submission was or a post-deadline reference calculation, denoted by "Blind" or "Reference" respectively. The table is not ordered by performance.

Method Category	Method	Microscopic pK_a (Type I) Submission ID	Macroscopic pK_a (Type III) Submission ID	Submission Type	Ref.
DL	Substructure matches to experimental data in pKa OpenEye pKa Prospector Database v1.0		*5nm4j*	Null	[2]
DL	OpenEye pKa-Prospector 1.0.0.3 with Analog Search ion identification algorithm		*pwn3m*	Null	[2]
LFER	ACD/pKa GALAS (ACD/Percepta Kernel v1.6)	*v8qph*	*37xm8*	Blind	[3]
LFER	ACD/pKa Classic (ACD/Percepta Kernel, v1.6)		*xmyhm*	Blind	[4]
LFER	Epik Scan (Schrodinger v2017-4)		*nb007*	Reference	[5]
LFER	Epik Microscopic (Schrodinger v2017-4)	*nb008*	*nb010*	Reference	[5]
QSPR/ML	OpenEye Gaussian Process	*6tvf8*	*hytjn*	Blind	[6]
QSPR/ML	OpenEye Gaussian Process Resampled		*q3pfp*	Blind	[6]
QSPR/ML	S+pKa (ADMET Predictor v8.5, Simulations Plus)	*hdiyq*	*gyuhx*	Blind	[7]
QSPR/ML	Chemicalize v18.23 (ChemAxon MarvinSketch v18.23)		*nb015*	Reference	[8]
QSPR/ML	MoKa v3.1.3	*nb016*	*nb017*	Reference	[9, 10]
QM	Adiabatic scheme with single point correction: SMD/M06-2X//6-311++G(d,p)//M06-2X/6-31+G(d) for bases and SMD/M06-2X//6-311++G(d,p)//M06-2X/6-31G(d) for acids + thermal corrections	*ko8yx*	*ryzue*	Blind	[11]
QM	Direct scheme with single point correction: SMD/M06-2X//6-311++G(d,p)//M06-2X/6-31+G(d) for bases and SMD/M06-2X//6-311++G(d,p)//M06-2X/6-31G(d) for acids + thermal corrections	*w4z0e*	*xikp8*	Blind	[11]
QM	Adiabatic scheme: thermodynamic cycle that uses gas phase optimized structures for gas phase free energy and solution phase geometries for solvent phase free energy. SMD/M06-2X/6-31+G(d) for bases and SMD/M06-2X/6-31G(d) for acids + thermal corrections	*wcvnu*	*5byn6*	Blind	[11]
QM	Vertical scheme: thermodynamic cycle that uses only gas phase optimized structures to compute gas hase and solvation free energy. SMD/M06-2X/6-31+G(d) for bases and SMD/M06-2X/6-31G(d) for acids + Thermal corrections	*arcko*	*w4iyd*	Blind	[11]
QM	Direct scheme: solution phase free energy is determined by solution phase geometries without thermodynamic cycle SMD/M06-2X/6-31+G(d) for bases and SMD/M06-2X/6-31G(d) for acids + thermal corrections	*wexjs*	*y75vj*	Blind	[11]
QM + LEC	Jaguar (Schrodinger v2017-4)	*nb011*	*nb013*	Reference	[12]
QM + LEC	CPCM/B3LYP/6–311+G(d,p) and global fitting	*y4wws*	*35bdm*	Blind	[13]
QM + LEC	CPCM/B3LYP/6–311+G(d,p) and separate fitting for neutral to negative and for positive to neutral transformations	*qsicn*	*p0jba*	Blind	[13]
QM + LEC	EC-RISM/MP2/6-311+G(d,p)-P3NI-q-noThiols-2par	*kxztt*	*ds62k*	Blind	[14]
QM + LEC	EC-RISM/MP2/cc-pVTZ-P2-q-noThiols-2par	*ftc8w*	*2ii2g*	Blind	[14]
QM + LEC	EC-RISM/MP2/6-311+G(d,p)-P2-phi-all-2par	*ktpj5*	*nb001*	Blind*	[14]
QM + LEC	EC-RISM/MP2/6-311+G(d,p)-P2-phi-noThiols-2par	*wuuvc*	*nb002*	Blind*	[14]
QM + LEC	EC-RISM/MP2/6-311+G(d,p)-P3NI-phi-all-2par	*2umai*	*nb003*	Blind*	[14]
QM + LEC	EC-RISM/MP2/6-311+G(d,p)-P3NI-phi-noThiols-2par	*cm2yq*	*nb004*	Blind*	[14]
QM + LEC	EC-RISM/MP2/6-311+G(d,p)-P2-phi-all-1par	*z7fhp*	*nb005*	Blind*	[14]
QM + LEC	EC-RISM/MP2/6-311+G(d,p)-P3NI-phi-all-1par	*8toyp*	*nb006*	Blind*	[14]
QM + LEC	EC-RISM/MP2/cc-pVTZ-P2-phi-noThiols-2par	*epvmk*	*ttjd0*	Blind	[14]
QM + LEC	EC-RISM/MP2/cc-pVTZ-P2-phi-all-2par	*xnoe0*	*mkhqa*	Blind	[14]
QM + LEC	EC-RISM/MP2/cc-pVTZ-P3NI-phi-noThiols-2par	*4o0ia*	*mpwiy*	Blind	[14]
QM + LEC	EC-RISM/B3LYP/6-311+G(d,p)-P3NI-q-noThiols-2par	*nxaaw*	*ad5pu*	Blind	[14]
QM + LEC	EC-RISM/B3LYP/6-311+G(d,p)-P3NI-phi-noThiols-2par	*0xi4b*	*f0gew*	Blind	[14]
QM + LEC	EC-RISM/B3LYP/6-311+G(d,p)-P2-phi-noThiols-2par	*cywyk*	*np6b4*	Blind	[14]
QM + LEC	PCM/B3LYP/6-311+G(d,p)	*gdqeg*	*yc70m*	Blind	[14]
QM + LEC	COSMOtherm_FINE17 (COSMOtherm C30_1701, BP/TZVPD/FINE//BP/TZVP/COSMO)	*t8ewk*	*0hxtm*	Blind	[15, 16]
QM + LEC	DSD-BLYP-D3(BJ)/def2-TZVPD//PBEh-3c[DCOSMO-RS] + RRHO(GFN-xTB[GBSA]) + Gsolv(COSMO-RS[TZVPD]) and linear fit		*xvxzd*	Blind	[17]
QM + LEC	ReSCoSS conformations // DSD-BLYP-D3 reranking // COSMOtherm pKa: DSD-BLYP-D3(BJ)/ def2-TZVPD// PBE-D3(BJ)/def2-TZVP/COSMO + RRHO[GFN-xTB + GBSA-water] + Gsolv[COSMO-RS(FINE17/TZVPD)] level and COSMOtherm pKa applied at the single conformer pair level (COSMOthermX17.0.5 release and BP-TZVPD-FINE-C30-1701 parameterization)	*eyetm*	*8xt50*	Blind	[17]
QM + LEC	ReSCoSS conformations // COSMOtherm pKa: DSD-BLYP-D3(BJ)/def2-TZVPD// PBE-D3(BJ)/ def2-TZVP/COSMO + RRHO[GFN-xTB + GBSA-water] + Gsolv[COSMO-RS(FINE17/TZVPD)] level and COSMOtherm pKa was applied directly on the resulting conformer sets with at least 5% Boltzmann weights for each microspecies (COSMOthermX17.0.5 release and BP-TZVPD-FINE-C30-1701 parameterization)	*ccpmw*	*yqkga*	Blind	[17]
QM + MM	M06-2X/6-31G*(for bases) or 6-31+G*(for acids) for gas phase, solvation free energy using TI with explicit solvent and GAFF, solvation free energy of proton -265.6 kcal/mol	*0wfzo*		Blind	[18]
QM + MM	M06-2X/6-31G*(for bases) or 6-31+G*(for acids) for gas phase, solvation free energy using TI with explicit solvent and GAFF, solvation free energy of proton -271.88 kcal/mol	*z3btx*		Blind	
QM + MM	M06-2X/6-31G*(for bases) or 6-31+G*(for acids) + thermal state correction for gas phase, solvation free energy using TI with explicit solvent and GAFF, solvation free energy of proton -265.6 kcal/mol	*758j8*		Blind	
QM + MM	M06-2X/6-31G*(for bases) or 6-31+G*(for acids) + thermal state correction for gas phase, solvation free energy using TI with explicit solvent and GAFF, solvation free energy of proton -271.88 kcal/mol	*hgn83*		Blind	

* Microscopic pK_a submissions were blind, however, participant requested a correction after blind submission deadline for macroscopic pK_a submissions. Therefore, these were assigned submission IDs in the form of *nb###*.

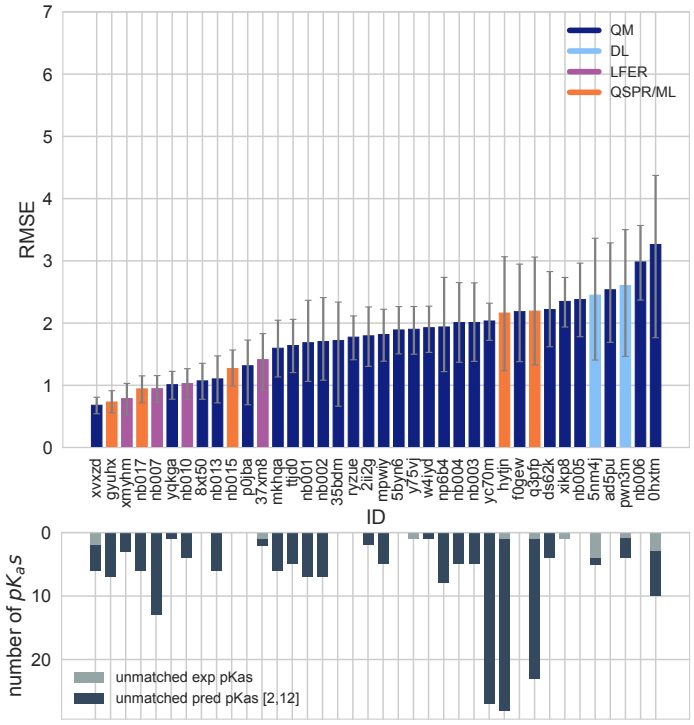

Figure 2. RMSE and unmatched pK_a counts vs. submission ID plots for macroscopic pK_a predictions based on Hungarian matching. Methods are indicated by submission IDs. RMSE is shown with error bars denoting 95% confidence intervals obtained by bootstrapping over challenge molecules. Lower bar plots show the number of unmatched experimental pK_as (light grey, missing predictions) and the number of unmatched pK_a predictions (dark grey, extra predictions) for each method between pH 2 and 12. Submission IDs are summarized in Table 1. Submission IDs of the form *nb###* refer to non-blinded reference methods computed after the blind challenge submission deadline. All others refer to blind, prospective predictions. Submissions are colored by their method categories. Light blue colored database look up methods are utilized as the null prediction method.

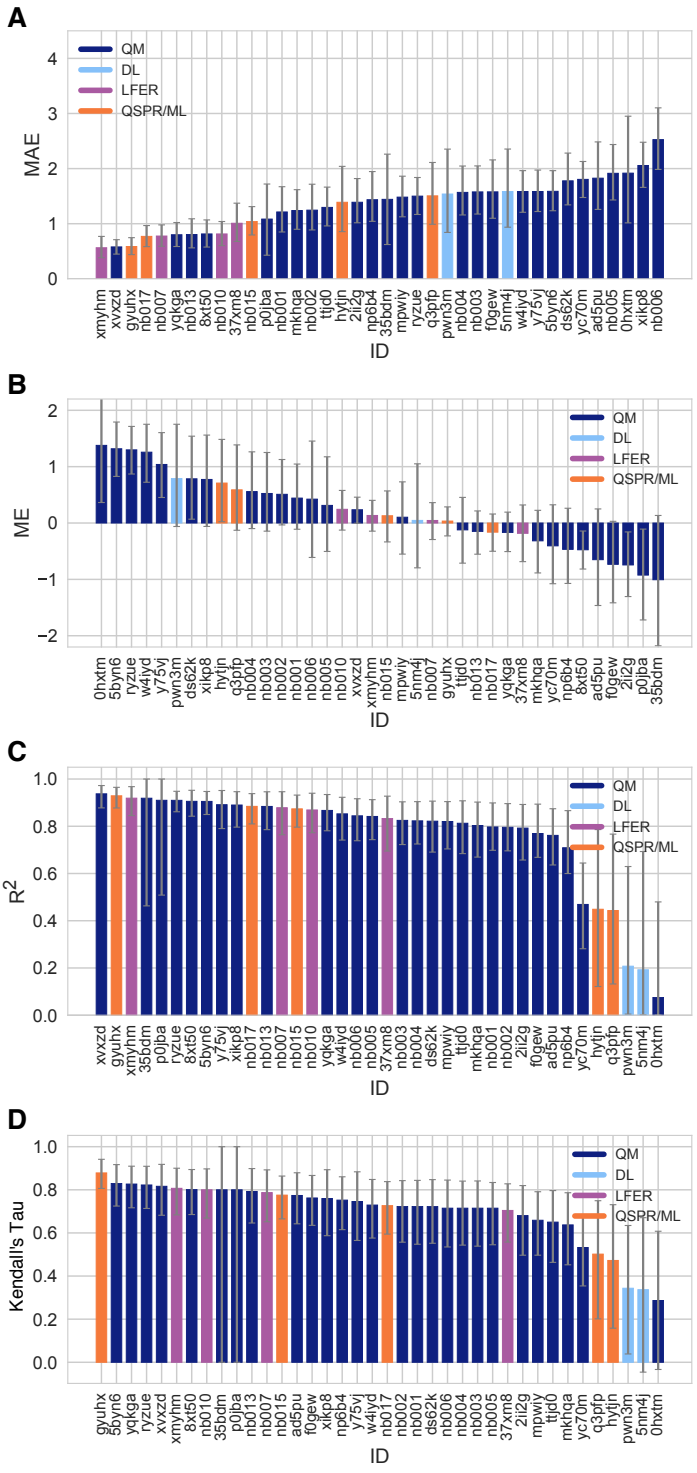

Figure 3. Additional performance statistics for macrocopic pK_a predictions based on Hungarian matching. Methods are indicated by submission IDs. Mean absolute error (MAE), mean error (ME), Pearson's R^2, and Kendall's Rank Correlation Coefficient Tau (τ) are shown, with error bars denoting 95% confidence intervals obtained by bootstrapping over challenge molecules. Refer to Table 1 for submission IDs and method names. Submissions are colored by their method categories. Light blue colored database look up methods are utilized as the null prediction method.

Performance comparison of different methods, grouped by methods class

Method comparison based on statistical metrics. Explain the numerical matching methods used. Explain rationale behind how we analyze the data and determine success/failure. Method comparison according to different statistics: RMSE, MAE, ME, R2, m, Kendall's tau.

3.1.1 Consistently well performing methods for macroscopic pK_a prediction

Table 2. Four consistently well-performing prediction methods for macroscopic pK_a prediction based on consistent ranking within the Top 10 according to various statistical metrics. Submissions were ranked according to RMSE, MAE, R^2, and τ. Consistently well-performing methods were selected as the ones that rank in the Top 10 in each of these statistical metrics. These methods also have less than 2 unmatched experimental pK_as and less than 7 unmatched predicted pK_as according to Hungarian matching. Performance statistics are provided as mean and 95% confidence intervals.

Submission ID	Method Name	RMSE	MAE	R^2	Kendall's Tau (τ)	Unmatched Exp. pK_a Count	Unmatched Pred. pK_a Count [2,12]
xvxzd	Full quantum chemical calculation of free energies and fit to experimental pKa	0.68 [0.54, 0.81]	0.58 [0.45, 0.71]	0.94 [0.88, 0.97]	0.82 [0.68, 0.92]	2	4
gyuhx	S+pKa	0.73 [0.55, 0.91]	0.59 [0.44, 0.74]	0.93 [0.88, 0.96]	0.88 [0.8, 0.94]	0	7
xmyhm	ACD/pKa Classic	0.79 [0.52, 1.03]	0.56 [0.38, 0.77]	0.92 [0.85, 0.97]	0.81 [0.68, 0.9]	0	3
8xt50	ReSCoSS conformations // DSD-BLYP-D3 reranking // COSMOtherm pKa	1.07 [0.78, 1.36]	0.81 [0.58, 1.07]	0.91 [0.84, 0.95]	0.80 [0.68, 0.89]	0	0

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

3.1.2 Which chemicals are harder to predict?

For physical prediction methods sulfur containing heterocycles, amide next to aromatic heterocycles, compounds with iodo and bromo domains have lower pKa prediction accuracy.

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?

Does molecular descriptors explain errors/performance ? We looked for correlation with descriptors, and potential explanation for errors. Keep spurious correlations in mind if we have many descriptors. No correlation observed. Reference the SI Figure of correlations.

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.

MI: Figure SI: correlation between prediction error and molecular descriptors

Are pKa predictions better in middle region? No correlation between pKa value and error was seen. Reference the SI Figure.

Refer to Ridge plots of Delta pKa error to identify compounds that were frequently mispredicted.

Compare ME of molecules across methods. Are there molecules often overestimated or underestimated?

No correlation of macroscopic pKa number to the errors? But we have low representation of multiprotic compounds

3.2 Analysis of microscopic pK_a predictions using microstates determined by NMR (8 molecules)

MI: FIGURE: Assign experimental pKa to microscopic transitions observed by NMR.

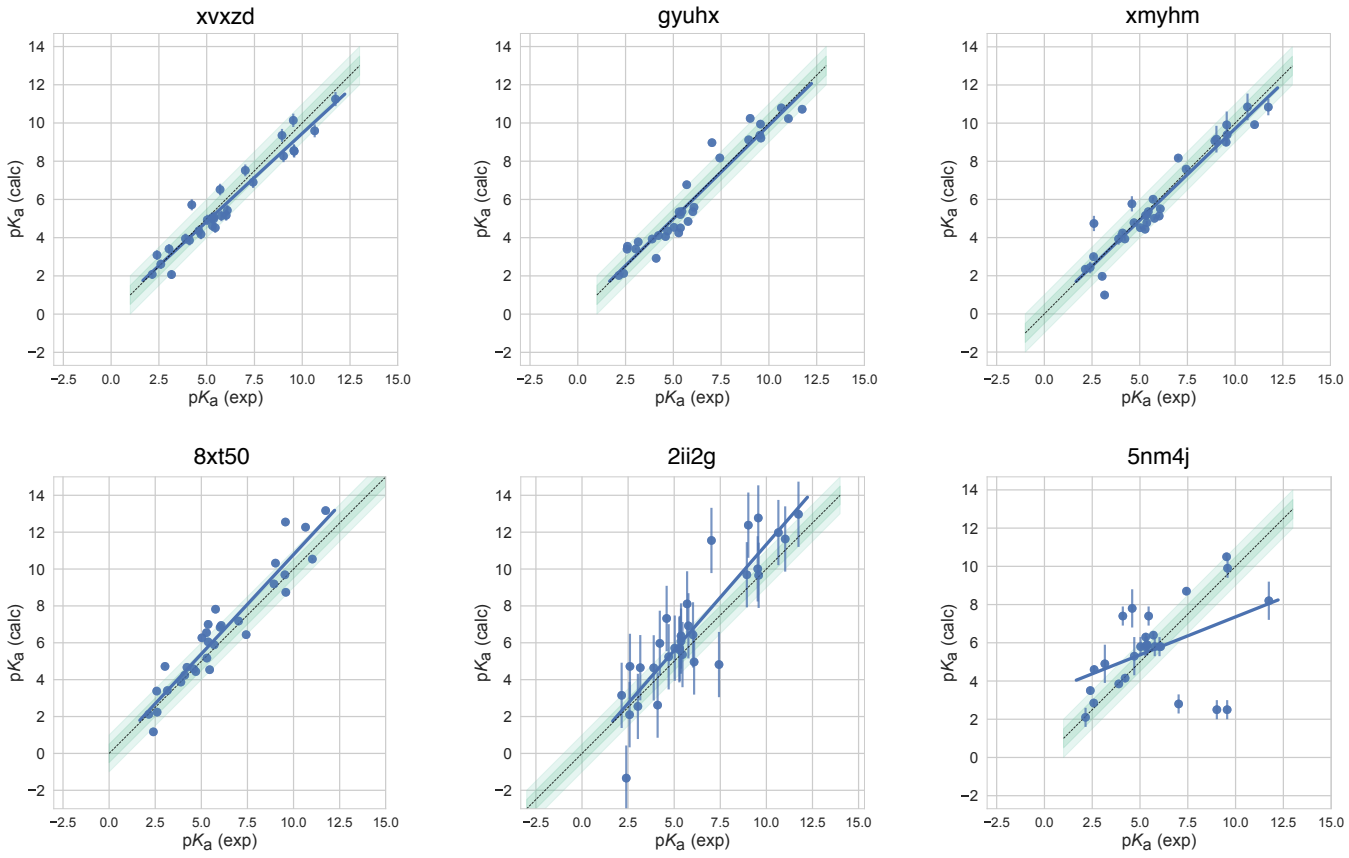

Figure 4. Predicted vs. experimental value correlation plots of 4 consistently well-performing methods, a representative method with average performance (*2ii2g*), and the null method (*5nm4j*). Dark and light green shaded areas indicate 0.5 and 1.0 units of error. Error bars indicate standard error of the mean of predicted and experimental values. Experimental pK_a SEM values are too small to be seen under the data points. EC-RISM/MP2/cc-pVTZ-P2-q-noThiols-2par method (*2ii2g*) was selected as the representative method with average performance because it is the method with the highest RMSE below the median.

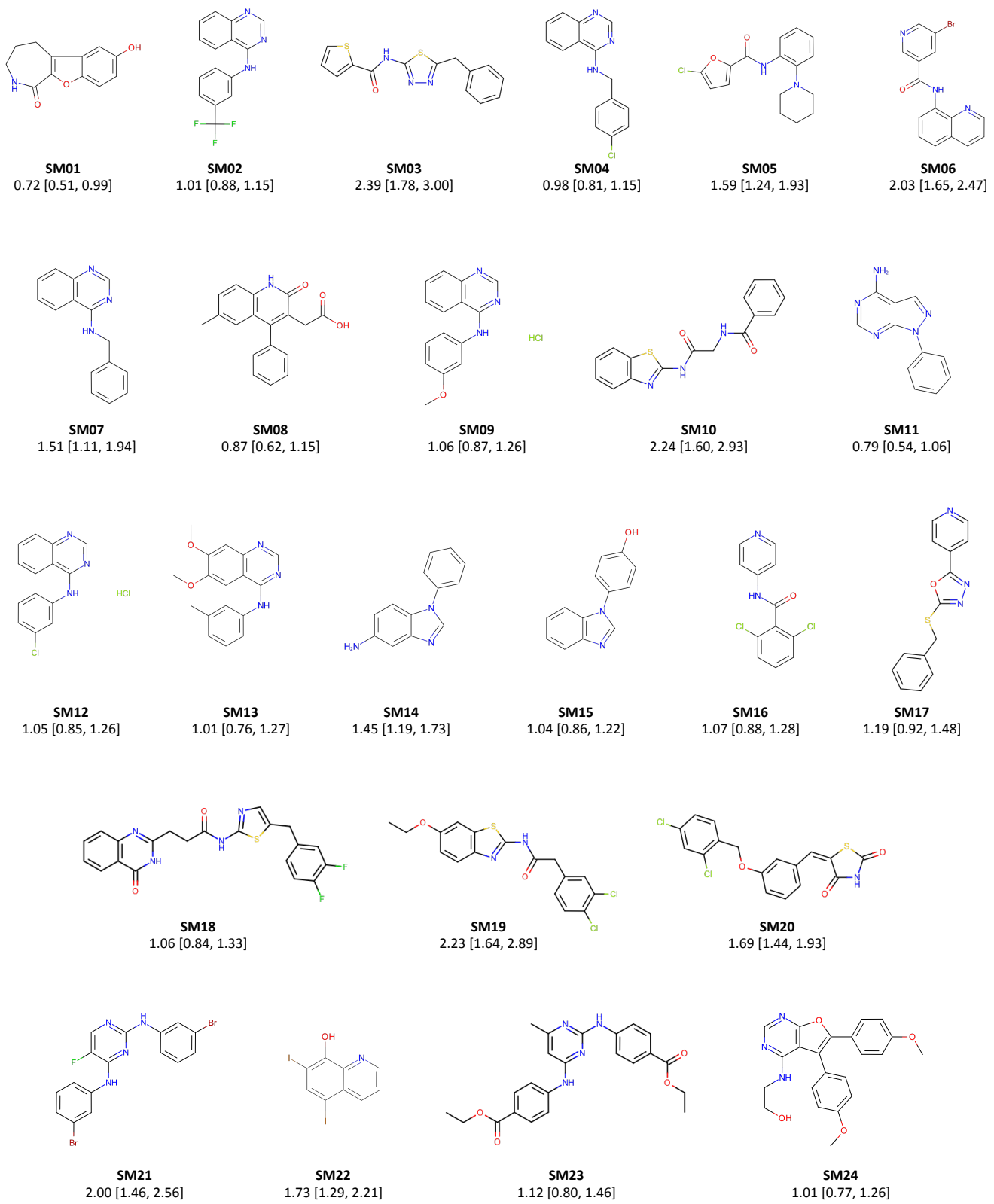

Figure 5. Molecules of SAMPL6 Challenge with MAE calculated for all macroscopic pK_a predictions. MAE calculated considering all prediction methods indicate which molecules had the lowest prediction accuracy in SAMPL6 Challenge. MAE values calculated for each molecule include all the matched pK_a values, which could be more than one per method for multiprotic molecules (SM06, SM14, SM15, SM16, SM18, SM22). Hungarian matching algorithm wasemployed for pairing experimental and predicted pK_a values. MAE values are reported with 95% confidence intervals.

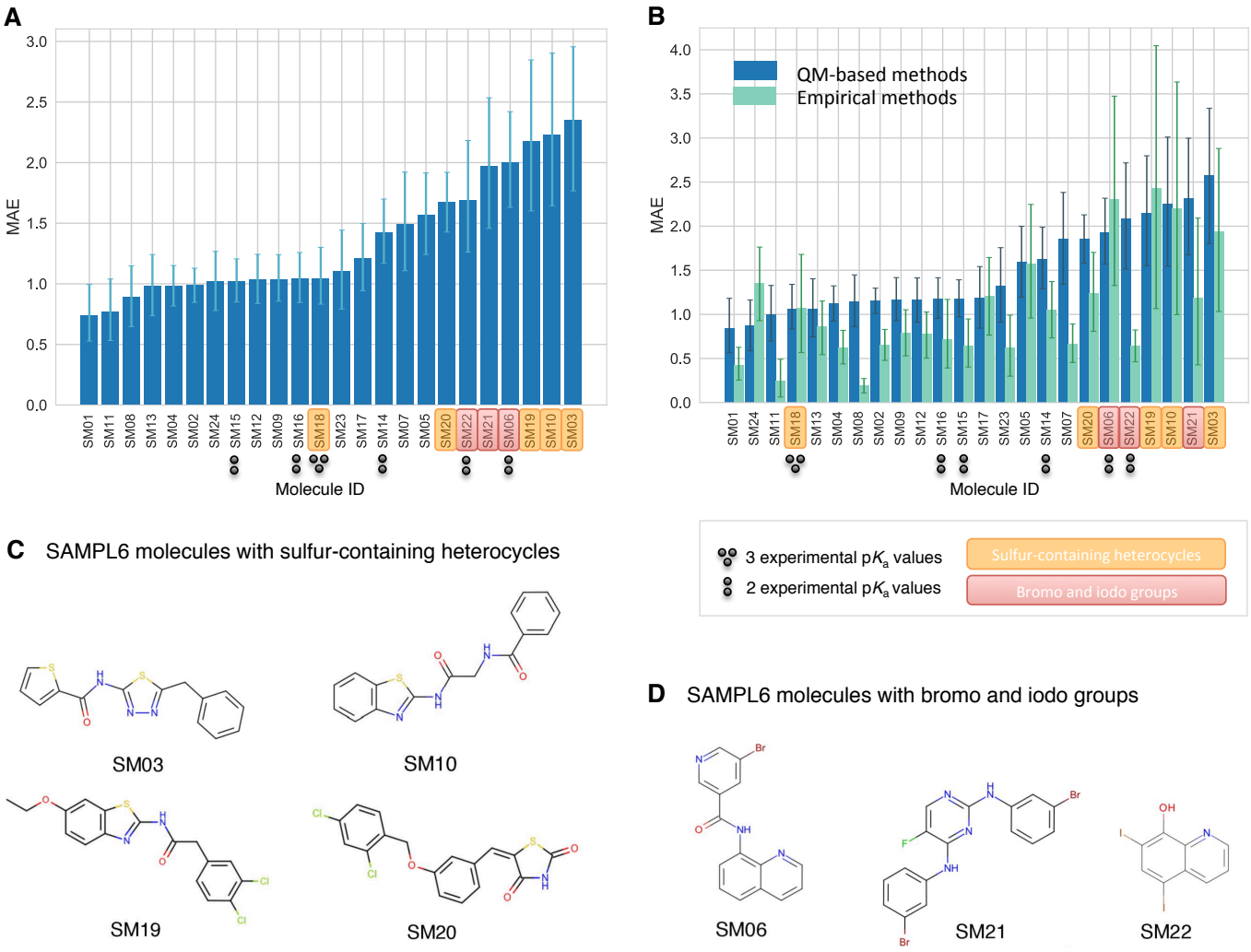

Figure 6. Average prediction accuracy calculated over all prediction methods was lower for molecules with sulfur-containing heterocycles, bromo, and iodo groups. (A) MAE calculated for each molecule as an average of all methods. **(B)** MAE of each molecule broken out by method category. QM-based methods (blue) include QM predictions with or without linear empirical correction. Empirical methods (green) include QSAR, ML, DL, and LFER approaches. **(C)** Depiction of SAMPL6 molecules with sulfur-containing heterocycles. **(D)** Depiction of SAMPL6 molecules with iodo and bromo groups .

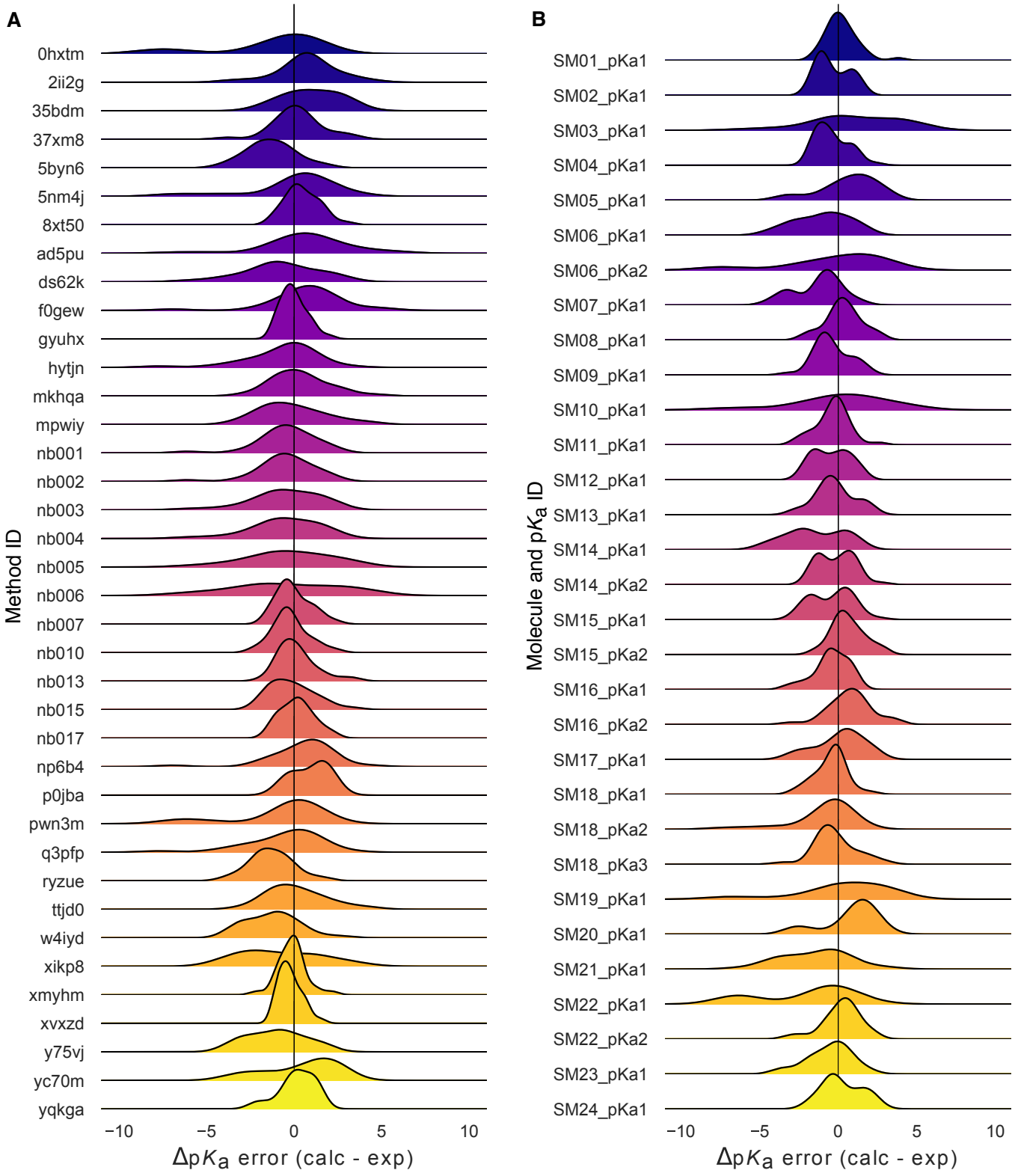

Figure 7. Prediction error distribution plots show how prediction accuracy varies across methods and individual molecules. (A) pK_a prediction error distribution for each submission for all molecules according to Hungarian matching. **(B)** Error distribution for each SAMPL6 molecule for all prediction methods according to Hungarian matching. For multiprotic molecules, pK_a ID numbers (pKa1, pKa2, and pKa3) were assigned in the direction of increasing experimental pK_a value.

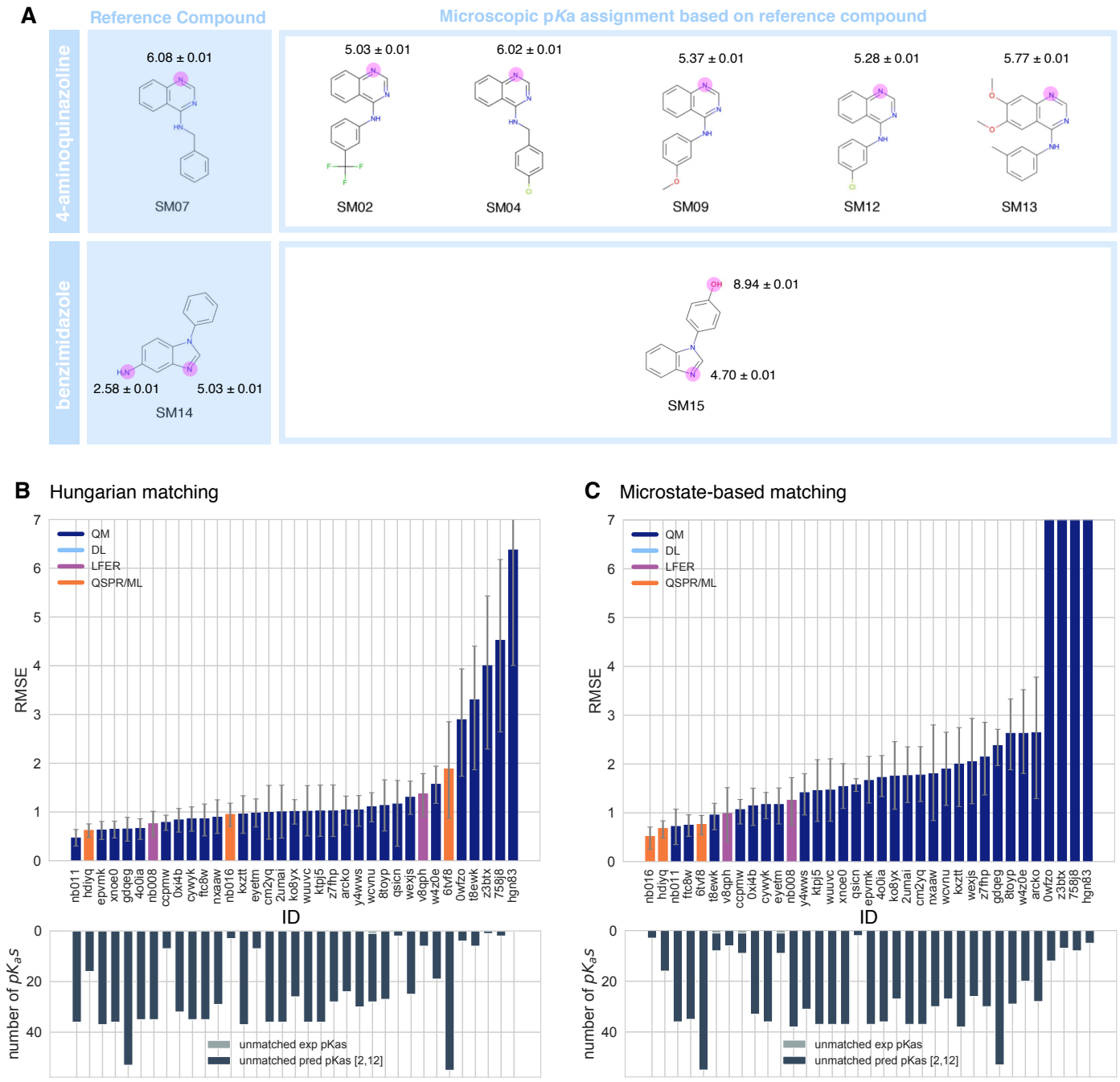

Figure 8. NMR determination of dominant microstates allowed in depth evaluation of microscopic pK_a predictions of 8 compounds. **A** Dominant microstate sequence of two compounds (SM07 and SM14) were determined by NMR [1]. Based on these reference compounds dominant microstates of 6 other derivative compounds were infered and experimental pK_a values were assigned to titratable groups with the assumption that only the dominant microstates have significant contributions to the experimentally observed pK_a. **B** RMSE vs. submission ID and unmatched pK_a vs. submission ID plots for the evaluation of microscopic pK_a predictions of 8 molecules by Hungarian matching to experimental macroscopic pK_as. **C** RMSE vs. submission ID and unmatched pK_a vs. submission ID plots showing the evaluation of microscopic pK_a predictions of 8 molecules by microstate-based matching between predicted microscopic pK_as and experimental macroscopic pK_a values. RMSE is shown with error bars denoting 95% confidence intervals obtained by bootstrapping over challenge molecules. Lower bar plots show the number of unmatched experimental pK_as (light grey, missing predictions) and the number of unmatched pK_a predictions (dark grey, extra predictions) for each method between pH 2 and 12. Submission IDs are summarized in Table 1.

3.2.1 Comparing microscopic pKa predictions directly to macroscopic experimental pKa values leads to underestimation of errors

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

Compare experimental data to microscopic pKa predictions, assuming experimental pKas are titrations of distinguishable sides and therefore equal to microscopic pKas. Molecules with only 1 pKa or well separated multiple pKas (more than 3 pKa units apart) SM14 and SM18 were excluded from this analysis, since their experimental pKa values don't satisfy these criteria.

Errors computed by microstate-based matching are larger compared to numerical matching algorithms. Microscopic pKa analysis with numerical matching algorithms may mask errors due to higher number of guesses made.

Conclusions will only be about 4-aminoquinazoline series and benzimidazole (8 molecules, 10 pKas) Refer to SI figure of dominant microstates.

Choosing molecules with right protonation state is important. Do people predict the correct sequence of dominant microstates? " Even if your pKa prediction is correct, protonation state prediction can be wrong." Analyze which state has lowest free energy for each charge group (The sequence of "experimentally visible states")

3.2.2 Accuracy of predicted pKa values when microstate matching is used

Assessment of individual methods by each of our analysis methods

Performance comparison of different methods, grouped by methods class

MI: FIGURE: Ranking of microscopic pKa prediction error statistsics for all participants (8 mol, microstate match).

MI: FIGURE: Violin plots of Delta pKa error to identify compounds that were frequently mispredicted (microstate match)

3.2.3 Dominant microstate prediction accuracy of methods

Calculate relative free energy of microstates to determine dominant microstate of each charge Compare predicted and experimental dominant microstates and calculate accuracy of each method

MI: FIGURE: Dominant microstate accuracy vs method plot. Charges together and separate.

What percent of the time predictions capture the dominant protonation state correctly? Match by microstate and calculate RMSE and MAE. If you know the microstates, can you predict the value of the pKa right?

Does top 3 methods predict the same dominant microstate sequence? How differently do different methods predict microscopic transitions? (method vs method correlation plot to see if methods predict the same microstate pairs or not)

3.2.4 Which molecules caused lower dominant microstate prediction accuracy?

Which molecule has more errors in predicting the major microstates?

MI: FIGURE: Dominant micorstate Accuracy vs Molecule ID plot, all charges and separate charges. Also think about plotting accuracy of QM and empirical methods separately.

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

3.2.5 Demonstrate how numerical matching often masks the error

Match by Hungarian and calculate accuracy of microstate prediction overall. When matched by pKa value, do people come with the same transition pairs?

MI: FIGURE: [accuracy-of-microstates-based-on-numeric-matching] For most methods the microstate pair of Hungarian predicted pKa does not match experimentally determined microstate pair.

3.3 Analyzing microscopic pKa prediction from the perspective of thermodynamics

Explain linearity relative free energy of protonation states with respect to pH. Free energy perspective simplifies data capturing and analysis. Reference Marilyn's paper.

Thermodynamic cycle closure checking allows evaluation of microsopic pKas without experimental data. Checking for thermodynamic consistency

3.3.1 Cycle closure error

Marilyn observed very good cycle closure results and very bad one that are up to 10 kcal/mol

She suggesting checking the cycle with maximum cycle closure error for each method and reporting that for each method. An historgam of max cycle closure error will help us bin these results into 3 categoris: 1. good agreement 2. moderate 3. severe

"We think thermodyamic cycles of protonation states need to be closed" Message: Methods need to checked for cycle closure errors. There can be information there that can be used to correct pKa predictions. When cycles are not closed it may be used as an indicator of prediction uncertainty.

3.4 How would pKa errors affect protein-ligand binding affinity predictions?

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

MI: FIGURE: a diagram illustrating the ways in which the pKa errors can influence prediction errors for binding affinities (A) When minor aqueous protonation state binds (B) When multiple protonation states can bind the complex

3.5 Lessons learned from SAMPL6 pKa Challenge

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

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

Overall results: 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 problem betwene 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.

Conclusion about prediction performance of individual molecules: 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. Which chemical structures make pKa predictions more difficult?

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

3.6 Suggestions 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

Reporting microscopic pKa predictions with charges, microstate free energies is betetr Experimental dataset with microstate infromation is more helpful.

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

Suggestions about challenge construction Enumeration of protonation states before predictions (which states does one need to consider?) Suggestions about challenge analysis

4 Conclusion

5 Code and data availability

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

6 Overview of supplementary information

Organized in SI document:

- TABLE SI 1: ???

Extra files:

- Any extra files

7 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.

8 Acknowledgments

Complete acknowledgments section. Caitlin Bannan, Thomas Fox

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

9 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.

Table ref: [3, 4, 7, 8, 10] trial: [], +, -, *, #

References

[1] **Işık M**, Levorse D, Rustenburg AS, Ndukwe IE, Wang H, Wang X, Reibarkh M, Martin GE, Makarov AA, Mobley DL, Rhodes T, Chodera JD. pKa Measurements for the SAMPL6 Prediction Challenge for a Set of Kinase Inhibitor-like Fragments. Journal of Computer-Aided Molecular Design. 2018 Oct; 32(10):1117–1138. doi: 10.1007/s10822-018-0168-0.

[2] OpenEye pKa Prospector;. OpenEye Scientific Software, Santa Fe, NM. Accessed on Jan 23, 2018. https://www.eyesopen.com/pka-prospector.

[3] ACD/pKa GALAS (ACD/Percepta Kernel v1.6);. Advanced Chemistry Development, Inc., Toronto, ON, Canada, 2018. https://www.acdlabs.com/products/percepta/predictors/pKa/.

[4] ACD/pKa Classic (ACD/Percepta Kernel v1.6);. Advanced Chemistry Development, Inc., Toronto, ON, Canada, 2018. https://www.acdlabs.com/products/percepta/predictors/pKa/.

[5] **Shelley JC**, Cholleti A, Frye LL, Greenwood JR, Timlin MR, Uchimaya M. Epik: A Software Program for pK a Prediction and Protonation State Generation for Drug-like Molecules. Journal of Computer-Aided Molecular Design. 2007 Dec; 21(12):681–691. doi: 10.1007/s10822-007-9133-z.

[6] **Bannan CC**, Mobley DL, Skillman AG. SAMPL6 Challenge Results from $$pK_a$$ Predictions Based on a General Gaussian Process Model. Journal of Computer-Aided Molecular Design. 2018 Oct; 32(10):1165–1177. doi: 10.1007/s10822-018-0169-z.

[7] Simulations Plus ADMET Predictor v8.5;. Simulations Plus, Lancaster, CA, 2018. https://www.simulations-plus.com/software/admetpredictor/physicochemical-biopharmaceutical/.

[8] Chemicalize v18.23 (ChemAxon MarvinSketch v18.23);. ChemAxon, Budapest, Hungary, 2018. https://docs.chemaxon.com/display/docs/pKa+Plugin.

[9] **Milletti F**, Storchi L, Sforna G, Cruciani G. New and Original p K_a Prediction Method Using Grid Molecular Interaction Fields. Journal of Chemical Information and Modeling. 2007 Nov; 47(6):2172–2181. doi: 10.1021/ci700018y.

[10] MoKa;. Molecular Discovery, Hertfordshire, UK, 2018. https://www.moldiscovery.com/software/moka/.

[11] **Zeng Q**, Jones MR, Brooks BR. Absolute and Relative pKa Predictions via a DFT Approach Applied to the SAMPL6 Blind Challenge. Journal of Computer-Aided Molecular Design. 2018 Oct; 32(10):1179–1189. doi: 10.1007/s10822-018-0150-x.

[12] **Bochevarov AD**, Harder E, Hughes TF, Greenwood JR, Braden DA, Philipp DM, Rinaldo D, Halls MD, Zhang J, Friesner RA. Jaguar: A High-Performance Quantum Chemistry Software Program with Strengths in Life and Materials Sciences. International Journal of Quantum Chemistry. 2013 Sep; 113(18):2110–2142. doi: 10.1002/qua.24481.

[13] **Selwa E**, Kenney IM, Beckstein O, Iorga BI. SAMPL6: Calculation of Macroscopic pKa Values from Ab Initio Quantum Mechanical Free Energies. Journal of Computer-Aided Molecular Design. 2018 Oct; 32(10):1203–1216. doi: 10.1007/s10822-018-0138-6.

[14] **Tielker N**, Eberlein L, Güssregen S, Kast SM. The SAMPL6 Challenge on Predicting Aqueous pKa Values from EC-RISM Theory. Journal of Computer-Aided Molecular Design. 2018 Oct; 32(10):1151–1163. doi: 10.1007/s10822-018-0140-z.

[15] **Klamt A**, Eckert F, Diedenhofen M, Beck ME. First Principles Calculations of Aqueous p K_a Values for Organic and Inorganic Acids Using COSMO-RS Reveal an Inconsistency in the Slope of the p K_a Scale. The Journal of Physical Chemistry A. 2003 Nov; 107(44):9380–9386. doi: 10.1021/jp034688o.

[16] **Eckert F**, Klamt A. Accurate Prediction of Basicity in Aqueous Solution with COSMO-RS. Journal of Computational Chemistry. 2006 Jan; 27(1):11–19. doi: 10.1002/jcc.20309.

[17] **Pracht P**, Wilcken R, Udvarhelyi A, Rodde S, Grimme S. High Accuracy Quantum-Chemistry-Based Calculation and Blind Prediction of Macroscopic pKa Values in the Context of the SAMPL6 Challenge. Journal of Computer-Aided Molecular Design. 2018 Oct; 32(10):1139–1149. doi: 10.1007/s10822-018-0145-7.

[18] **Prasad S**, Huang J, Zeng Q, Brooks BR. An Explicit-Solvent Hybrid QM and MM Approach for Predicting pKa of Small Molecules in SAMPL6 Challenge. Journal of Computer-Aided Molecular Design. 2018 Oct; 32(10):1191–1201. doi: 10.1007/s10822-018-0167-1.

10 Supplementary Information

MI: Figure [typeIII-error-dist-by-method] Distribution of prediction errors for each method in SAMPL6 Challenge. Analyses was performed based on Hungarian matching algorithm. Y-axis labels indicate submission IDs of each method.

MI: [pKa-error-vs-pKa-value]. Error in pKa predictions does not correlate with the true value of pKa. Left figure was constructed using closest match between experimental and predicted pKas. Y-axis is absolute residuals of the pKa prediction.

MI: FIGURE [desc-vs-MAE-correlation]. There is no clear correlation between molecular descriptors and mean absolute error for each molecule when calculated for all methods.

MI: SI Table: Type I collection

MI: SI Table: Type III collection

MI: SI Figure: type I correlation plots of each method

MI: SI Figure: type III correlation plots of each method

MI: TABLE: InChI and SMILES for chemicals

MI: TABLE: Statistics based on hungarian matching

MI: TABLE: Statistics based on microstate matching

MI: TABLE: NMR determined microstates of 8 molecules