

# Accuracy of macroscopic and microscopic pK<sub>a</sub> predictions of small molecules evaluated by the SAMPL6 blind prediction challenge

Mehtap Işık (ORCID: [0000-0002-6789-952X](#))<sup>1,2\*</sup>, Ariën S. Rustenburg (ORCID: [0000-0002-3422-0613](#))<sup>1,3</sup>, Andrea Rizzi (ORCID: [0000-0001-7693-2013](#))<sup>1,4</sup>, M. R. Gunner<sup>6</sup>, David L. Mobley (ORCID: [0000-0002-1083-5533](#))<sup>5</sup>, John D. Chodera (ORCID: [0000-0003-0542-119X](#))<sup>1</sup>

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

\*For correspondence:  
[mehtap.isik@choderlab.org](mailto:mehtap.isik@choderlab.org) (MI)

## Abstract

Acid dissociation constant (pK<sub>a</sub>) predictions is a prerequisite for predicting many other properties of small molecules such as protein-ligand binding affinity, distribution coefficient (log D), and solubility due to the necessity of predicting relevant protonation states and the free energy penalty of each state. SAMPL6 pK<sub>a</sub> Challenge was the first time that a separate challenge was conducted for evaluating pK<sub>a</sub> predictions as a part of SAMPL. It was motivated by the inaccuracies observed in prior physical property prediction challenges, such as SAMPL5 log D Challenge, caused by protonation state and pK<sub>a</sub> prediction issues. The goal of the pK<sub>a</sub> challenge was to elucidate the performance of contemporary pK<sub>a</sub> prediction methods for drug-like molecules. The challenge set was composed of 24 kinase inhibitor fragment-like small molecules and some of them were multiprotic. 11 research groups contributed blind prediction sets of 37 pK<sub>a</sub> prediction methods. Four widely used pK<sub>a</sub> prediction methods that were missing from blind predictions were added as reference methods to challenge analysis. Collecting both microscopic and macroscopic pK<sub>a</sub> predictions allowed in-depth evaluation of pK<sub>a</sub> prediction performance. This article highlights deficiencies of typical pK<sub>a</sub> prediction evaluation approaches when the difference between microscopic and macroscopic pK<sub>a</sub>s is ignored and suggests more stringent evaluation criteria for microscopic and macroscopic pK<sub>a</sub> predictions guided by the available experimental data. Top-performing submissions for macroscopic pK<sub>a</sub> predictions achieved RMSE of 0.7-1.0 units and included both quantum-mechanical and empirical approaches. These predictions included less than 8 extra/missing macroscopic pK<sub>a</sub>s for the set of 24 molecules. A large number of submissions had RMSE spanning 1-3 pK<sub>a</sub> units. Molecules with sulfur-containing heterocycles, iodo, and bromo groups suffered from less accurate pK<sub>a</sub> predictions on average considering all methods evaluated. For a subset of molecules, the available NMR-based dominant microstate sequence data was utilized to elucidate dominant tautomer prediction errors of microscopic pK<sub>a</sub> predictions which was prominent for charged tautomers. SAMPL6 pK<sub>a</sub> Challenge demonstrated the need for improving pK<sub>a</sub> prediction methods for drug-like molecules, especially for challenging moieties and multiprotic molecules. The level of pK<sub>a</sub> prediction inaccuracy observed in this challenge has potential to be detrimental to the performance of protein-ligand binding affinity predictions in two ways: (1) errors in predicted dominant charge and tautomeric state and (2) errors in the calculation of free energy correction for minor and multiple protonation states of the ligand.

## 42 0.1 Keywords

43 SAMPL · blind prediction challenge · acid dissociation constant ·  $pK_a$  · small molecule · macroscopic  $pK_a$  · microscopic  $pK_a$  · macro-  
44 scopic protonation state · microscopic protonation state

## 45 0.2 Abbreviations

46 **SAMPL** Statistical Assessment of the Modeling of Proteins and Ligands

47  **$pK_a$**   $-\log_{10}$  acid dissociation equilibrium constant

48 **SEM** Standard error of the mean

49 **RMSE** Root mean squared error

50 **MAE** Mean absolute error

51  $\tau$  Kendall's rank correlation coefficient (Tau)

52 **R<sup>2</sup>** Coefficient of determination (R-Squared)

## 53 1 Introduction

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

57 Importance of small molecule pKa prediction for pharmaceutical efforts.

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

59 Acid dissociation constant ( $pK_a$ ) predictions is a prerequisite for predicting many other properties of small molecules such as  
60 protein-ligand binding affinity, distribution coefficient ( $\log D$ ), and solubility due to the necessity of predicting relevant protonation  
61 states and their free energy penalty of such states. Therefore, accurate computational  $pK_a$  prediction methods are required  
62 for computer-aided drug design.

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

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

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

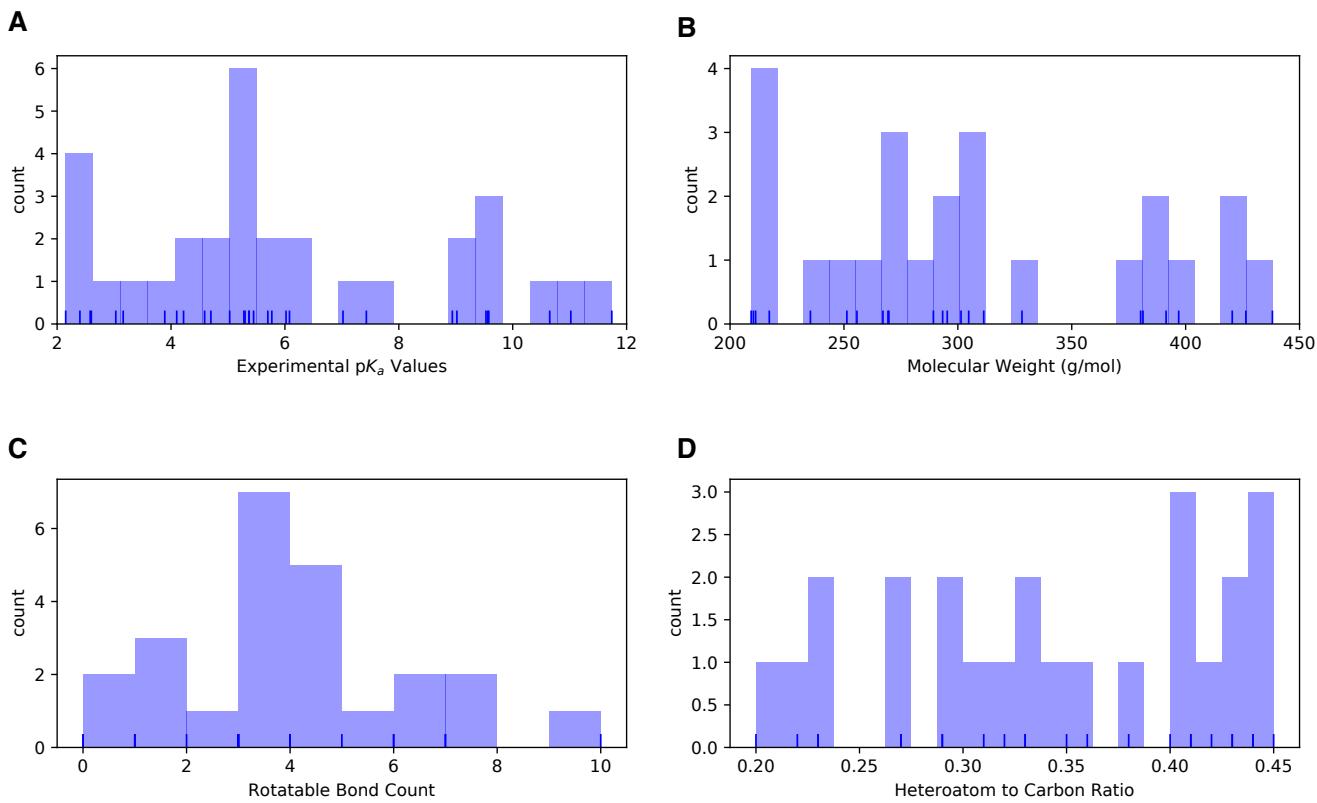
69 This is the first time a blind pKa prediction challenge has been fielded as part of SAMPL. In this first iteration of the challenge,  
70 we aimed to assess the performance of current pKa prediction methods and isolate potential causes of inaccurate pKa estimates,  
71 with the aim of determining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

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

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

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

84 Explain the physics of the predicted property



**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  $pK_a$ s 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.

EQUATION:  $pK_a$  equation

EQUATION: free energy of protonation state equation

Introducing linear protonation state free energy diagram

MI: FIGURE: linear plot of free energy vs pH

Overview of kinds of  $pK_a$  prediction methods available (ML, QM, empirical methods ...)

Explain challenge design.

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

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

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

89 The comparison between macroscopic and microscopic  $pK_a$  values is not always a straightforward one.

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

91 Explain future direction for this challenge

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

100 Explain potential benefits of these challenge

101 Improving computational methods...

## 102 1.1 Motivation for a blind pKa challenge

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

104 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.

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

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

110 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.

116 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.

120 Reason for blind pKa challenge: 1. Impact on binding affinity predictions 2. Impact on logD predictions (SAMPL6) 3. Drug-like molecules are especially challenging.

122 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. Potential benefits of these challenges: 1. Improving computational methods 2. Detecting hidden contributors to error

## 125 1.2 Approaches to predict pKas

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

## 127 2 Methods

### 128 2.1 Structure and logistics of the SAMPL6 pKa prediction challenge

129 Describe the structure of SAMPL6 pKa challenge

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

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

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

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

137 - When instructions and input files were made available

138 - Challenge dates

139 - Input files

140 - What to predict? Three type of submissions.

141 - Multiple submissions allowed

142 - Predicting the pKa values of the whole set wasn't a requirement.

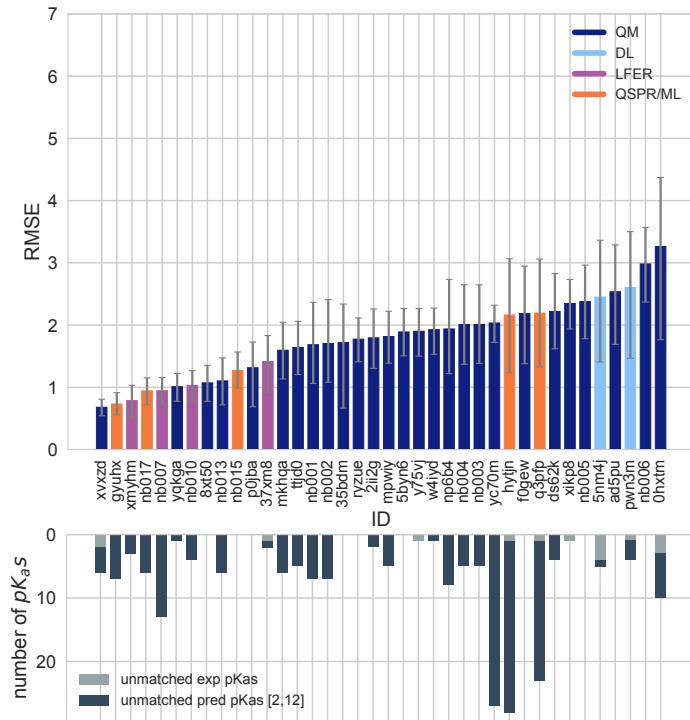
143 - 2nd D3R/SAMPL Workshop took place in La Jolla, San Diego on Feb 22-23, 2018.

- 144 Referce Figure ???. Drug-like molecules are often larger and more complex than the ones used in this study.
- 145 **2.2 Enumeration of requested prediction microscopic protonation states**
- 146 1. OpenEye (filter out resonance structures), Epik
- 147 2. Participant supplied structures
- 148 Microstate pairs: Only +/-1 charge change transitions are allowed. List of allowed transitions. +2 transitions are not consid-
- 149 ered.
- 150 **2.3 Evaluation approaches**
- 151 2.3.1 Statistical metrics for submission performance
- 152 - Root mean squared error (RMSE)
- 153 - Mean absolute error (MAE)
- 154 - Mean Error (ME)
- 155 - Square of Pearson Correlation Coefficient ( $R^2$ )
- 156 - Slope of prediction vs. experimental value linear fit
- 157 Uncertainty in each performance statistic was calculated by bootstrapping (10,000) to estimate 95% confidence intervals.
- 158 2.3.2 Matching algorithms for pairing predicted and experimental pKas
- 159 Explain why it is necessary due to lacking structural information. Cite recommendations from article such as preserving sequence.
- 160 Experimental data doesn't inform protonation site and overall charge of species. Experimental data doesn't capture the whole
- 161 picture. We don't know charge and we don't know tautomers. We don't know the charge state of macrostates, this causes a
- 162 matching problem
- 163 Explain Hungarian method for matching experimental and predicted pKas
- 164 Explain Closest method for matching experimental and predicted pKas
- 165 Explain microstate based matching.
- 166 **2.4 Reference calculations**
- 167 Schrodinger Epik Schrodinger Jaguar Chemicalize MoKa
- 168 **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)
- 169 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).
- 170 **3.1 Analysis of macroscopic  $pK_a$  predictions (Type III)**
- 171 Refer to SI TABLE: Error statistics for all participants. Refer to SI FIGURE: Error distribution ridge plots for each method (exp-pred
- 172 macroscopic pKa). Which methods tend to overestimate and which methods tend to underestimate?
- 173 Describe number of missing and extra pKa for each method. Report in total for all molecules how many predicted pKas are
- 174 there and how many experimental pKas. Refer to FIGURE: missing and extra pKa counts.
- 175 Describe overall performance comparison of different methods, grouped by methods class.
- 176 Explain rationale behind how we analyze the data and determine success/failure
- 177 Performance comparison of different methods, grouped by methods class

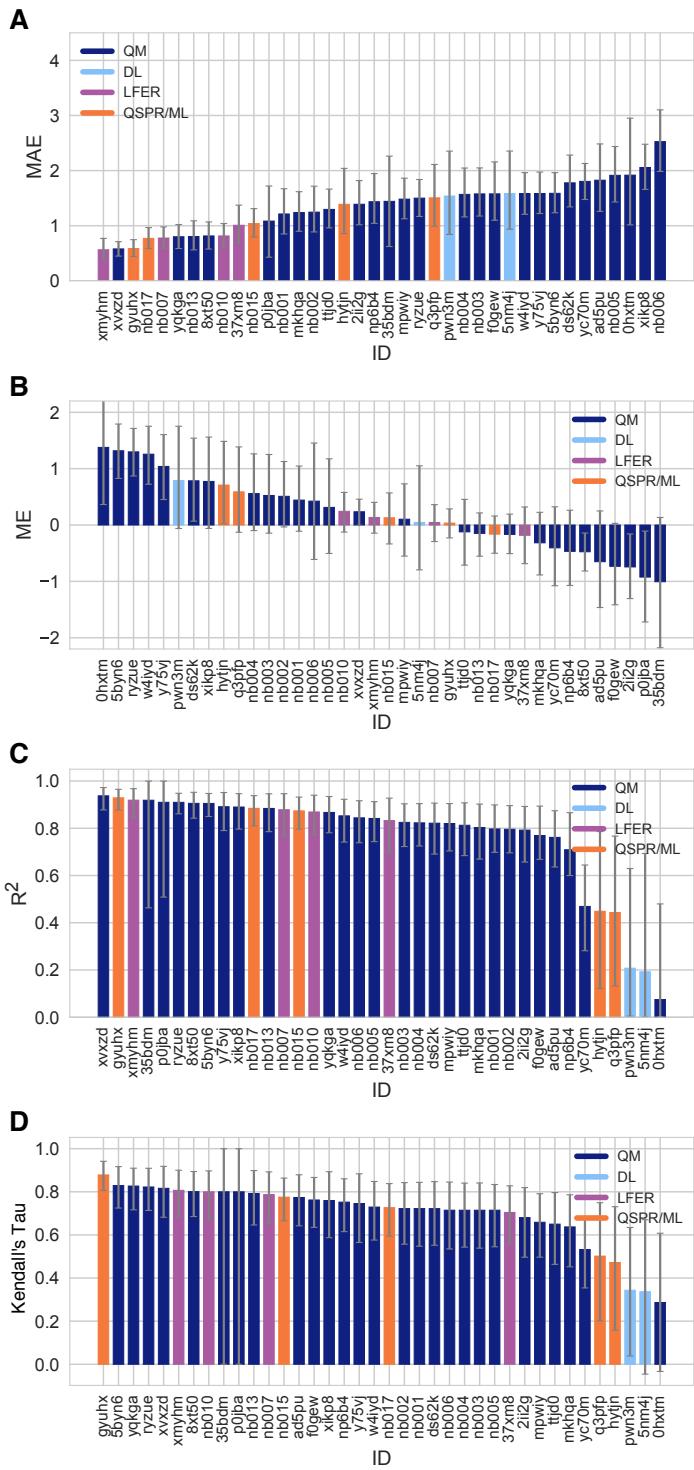
**Table 1. Submission IDs, names, category, and type for all the  $pK_a$  prediction sets.** Reference calculations are labeled as  $nb\#\#\#$ . The method name column lists the names provided by each participant in the submission file. The “type” column indicates if submission was or a post-deadline 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	<i>5nm4j</i>		Null	[2]
DL	OpenEye pKa-Prospector 1.0.0.3 with Analog Search ion identification algorithm	<i>pwn3m</i>		Null	[2]
LFER	ACD/pKa GALAS (ACD/Percepta Kernel v1.6)	<i>v8qph</i>	<i>37xm8</i>	Blind	[3]
LFER	ACD/pKa Classic (ACD/Percepta Kernel, v1.6)		<i>xmyhm</i>	Blind	[4]
LFER	Epik Scan (Schrodinger v2017-4)		<i>nb007</i>	Reference	[5]
LFER	Epik Microscopic (Schrodinger v2017-4)		<i>nb010</i>	Reference	[5]
QSPR/ML	OpenEye Gaussian Process	<i>6tvf8</i>	<i>hytjn</i>	Blind	[6]
QSPR/ML	OpenEye Gaussian Process Resampled		<i>q3pfp</i>	Blind	[6]
QSPR/ML	S+pkA (ADMET Predictor v8.5, Simulations Plus)	<i>hdlyq</i>	<i>gyuhx</i>	Blind	[7]
QSPR/ML	Chemcalize v18.23 (ChemAxon MarvinSketch v18.23)		<i>nb015</i>	Reference	[8]
QSPR/ML	MoKa v3.1.3	<i>nb016</i>	<i>nb017</i>	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	<i>k08yx</i>	<i>ryzue</i>	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	<i>w4z0e</i>	<i>xikp8</i>	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	<i>wcvnu</i>	<i>5byn6</i>	Blind	[11]
QM	Vertical scheme: thermodynamic cycle that uses only gas phase optimized structures to compute gas phase and solvation free energy. SMD/M06-2X/6-31+G(d) for bases and SMD/M06-2X/6-31G(d) for acids + Thermal corrections	<i>arcko</i>	<i>w4iyd</i>	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	<i>wexjs</i>	<i>y75vj</i>	Blind	[11]
QM + LEC	Jaguar (Schrodinger v2017-4)	<i>nb011</i>	<i>nb013</i>	Reference	[12]
QM + LEC	CPCM/B3LYP/6-311+G(d,p) and global fitting	<i>y4wws</i>	<i>35bdm</i>	Blind	[13]
QM + LEC	CPCM/B3LYP/6-311+G(d,p) and separate fitting for neutral to negative and for positive to neutral transformations	<i>qsicn</i>	<i>p0jba</i>	Blind	[13]
QM + LEC	EC-RISM/MP2/6-311+G(d,p)-P3NI-q-noThiols-2par	<i>kxzt</i>	<i>ds62k</i>	Blind	[14]
QM + LEC	EC-RISM/MP2/cc-pVTZ-P2-q-noThiols-2par	<i>ftc8w</i>	<i>2ii2g</i>	Blind	[14]
QM + LEC	EC-RISM/MP2/6-311+G(d,p)-P2-phi-all-2par	<i>ktpj5</i>	<i>nb001</i>	Blind*	[14]
QM + LEC	EC-RISM/MP2/6-311+G(d,p)-P2-phi-noThiols-2par	<i>wuuvc</i>	<i>nb002</i>	Blind*	[14]
QM + LEC	EC-RISM/MP2/6-311+G(d,p)-P3NI-phi-all-2par	<i>2umai</i>	<i>nb003</i>	Blind*	[14]
QM + LEC	EC-RISM/MP2/6-311+G(d,p)-P3NI-phi-noThiols-2par	<i>cm2yq</i>	<i>nb004</i>	Blind*	[14]
QM + LEC	EC-RISM/MP2/6-311+G(d,p)-P2-phi-all-1par	<i>z7fhp</i>	<i>nb005</i>	Blind*	[14]
QM + LEC	EC-RISM/MP2/6-311+G(d,p)-P3NI-phi-all-1par	<i>8toyp</i>	<i>nb006</i>	Blind*	[14]
QM + LEC	EC-RISM/MP2/cc-pVTZ-P2-phi-noThiols-2par	<i>epvmk</i>	<i>tjjd0</i>	Blind	[14]
QM + LEC	EC-RISM/MP2/cc-pVTZ-P2-phi-all-2par	<i>xnoe0</i>	<i>rnkhqa</i>	Blind	[14]
QM + LEC	EC-RISM/MP2/cc-pVTZ-P3NI-phi-noThiols-2par	<i>4o0ia</i>	<i>mpwiy</i>	Blind	[14]
QM + LEC	EC-RISM/B3LYP/6-311+G(d,p)-P3NI-q-noThiols-2par	<i>nxaaw</i>	<i>ad5pu</i>	Blind	[14]
QM + LEC	EC-RISM/B3LYP/6-311+G(d,p)-P3NI-phi-noThiols-2par	<i>0xi4b</i>	<i>f0gew</i>	Blind	[14]
QM + LEC	EC-RISM/B3LYP/6-311+G(d,p)-P2-phi-noThiols-2par	<i>cywyk</i>	<i>np6b4</i>	Blind	[14]
QM + LEC	PCM/B3LYP/6-311+G(d,p)	<i>gdqeg</i>	<i>yc70m</i>	Blind	[14]
QM + LEC	COSMOtherm_FINE17 (COSMOtherm C30_1701, BP/TZVPD/FINE//BP/TZVP/COSMO)	<i>t8ewk</i>	<i>0hxtm</i>	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		<i>xvxd</i>	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[DCOSMO-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)	<i>eyetm</i>	<i>8xt50</i>	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[DCOSMO-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)	<i>ccpmw</i>	<i>yqkga</i>	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	<i>0wfzo</i>		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	<i>z3btv</i>		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	<i>758j8</i>		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	<i>hgn83</i>		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  $pK_a$ s (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 macroscopic pKa predictions based on Hungarian matching.** Methods are indicated by submission IDs. Mean absolute error (MAE), mean error (ME), Pearson's R<sup>2</sup>, and Kendall's Rank Correlation Coefficient Tau ( $\tau$ ) 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.

184 Method comparison based on statistical metrics. Explain the numerical matching methods used. Explain rationale behind  
185 how we analyze the data and determine success/failure. Method comparison according to different statistics: RMSE, MAE, ME,  
186 R<sup>2</sup>, m, Kendall's tau.

187 3.1.1 Consistently well performing methods for macroscopic pK<sub>a</sub> prediction

**Table 2. Four consistently well-performing prediction methods for macroscopic pK<sub>a</sub> prediction based on consistent ranking within the Top 10 according to various statistical metrics.** Submissions were ranked according to RMSE, MAE, R<sup>2</sup>, and  $\tau$ . 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<sub>a</sub>s and less than 7 unmatched predicted pK<sub>a</sub>s according to Hungarian matching. Performance statistics are provided as mean and 95% confidence intervals.

Submission ID	Method Name	RMSE	MAE	R <sup>2</sup>	Kendall's Tau ( $\tau$ )	Unmatched Exp. pK <sub>a</sub> Count	Unmatched Pred. pK <sub>a</sub> 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

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

189 3.1.2 Which chemicals are harder to predict?

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

192 Prediction performance of individual molecules

193 Which chemical structures make pKa predictions more difficult?

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

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

197 Are there any correlations between molecular descriptors and pKa errors?

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

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

202 Comparison of errors/performance against molecular descriptors. Look for correlation with descriptors, and potential explanation for  
203 errors. Keep spurious correlations in mind if we have many descriptors.

204 Refer to Figure S1: correlation between prediction error and molecular descriptors. There is no clear correlation between  
205 molecular descriptors and mean absolute error for each molecule when calculated for all methods.

206 Are pKa predictions better in middle region? Error in pKa predictions does not correlate with the true value of pKa. No  
207 correlation between pKa value and error was seen. Reference the SI Figure.

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

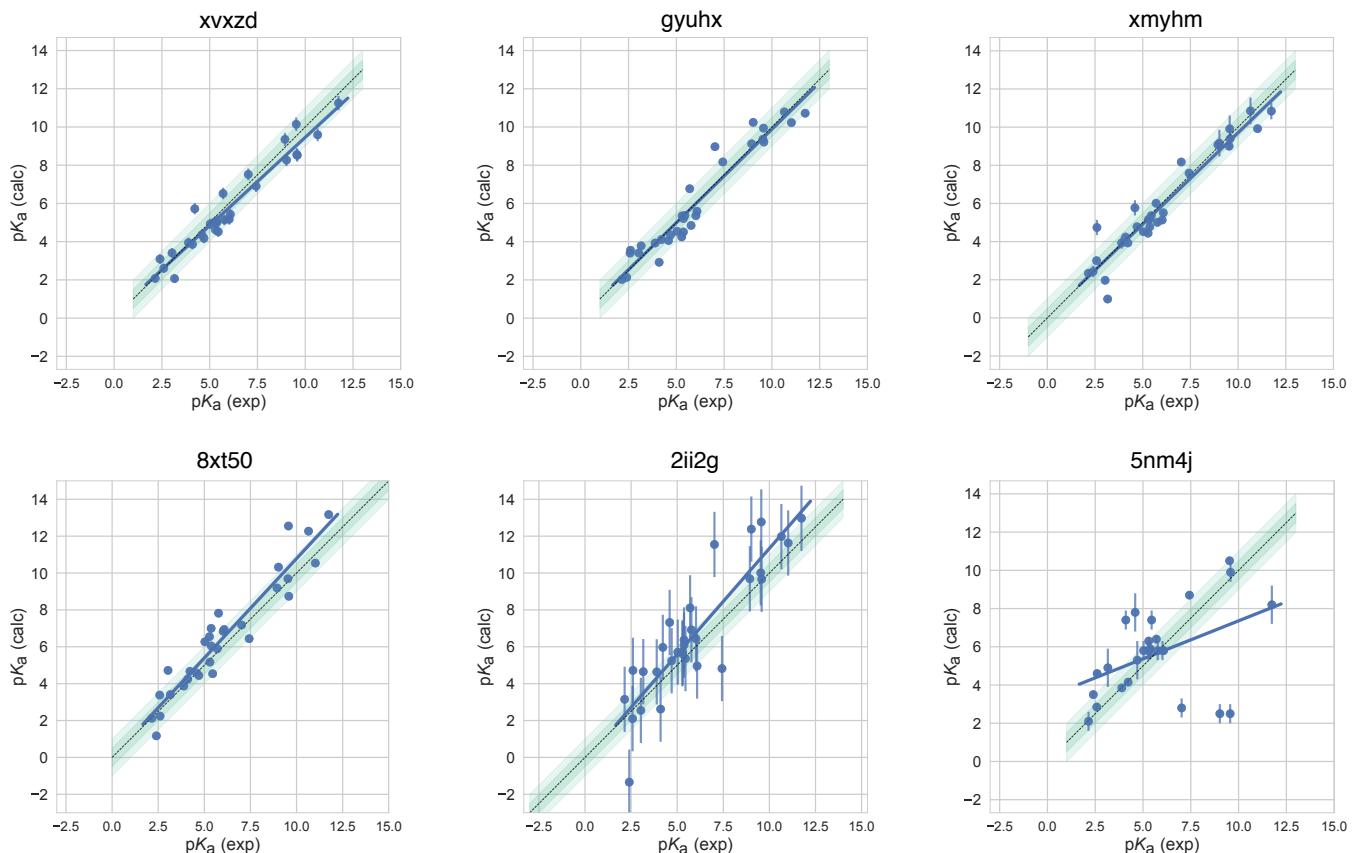
209 Compare ME of molecules across methods. Are there molecules often overestimated or underestimated?

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

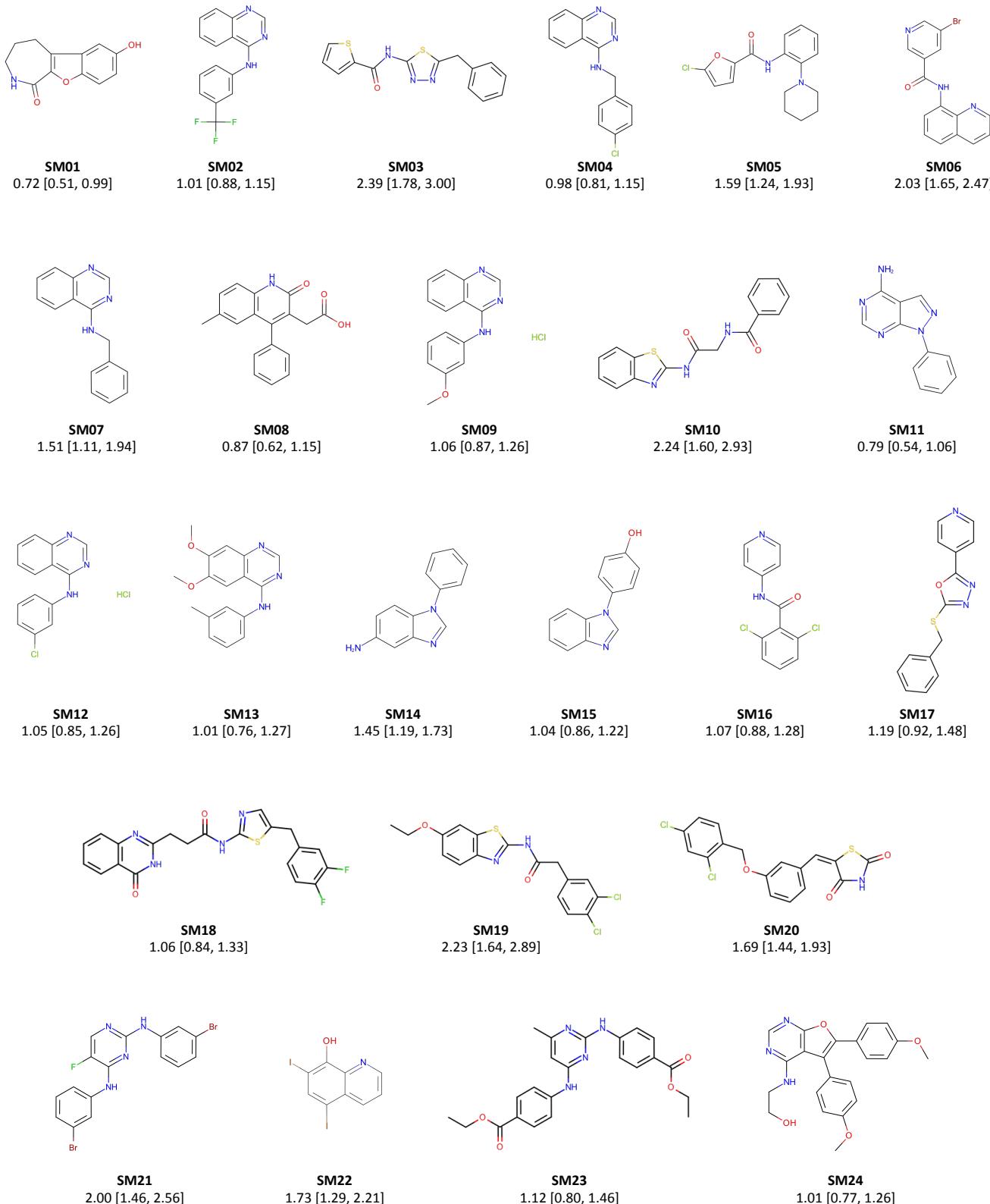
211 3.2 Analysis of microscopic pK<sub>a</sub> predictions using microstates determined by NMR (8 molecules)

212 3.2.1 Comparing microscopic pKa predictions directly to macroscopic experimental pKa values with numerical  
213 matching leads to underestimation of errors

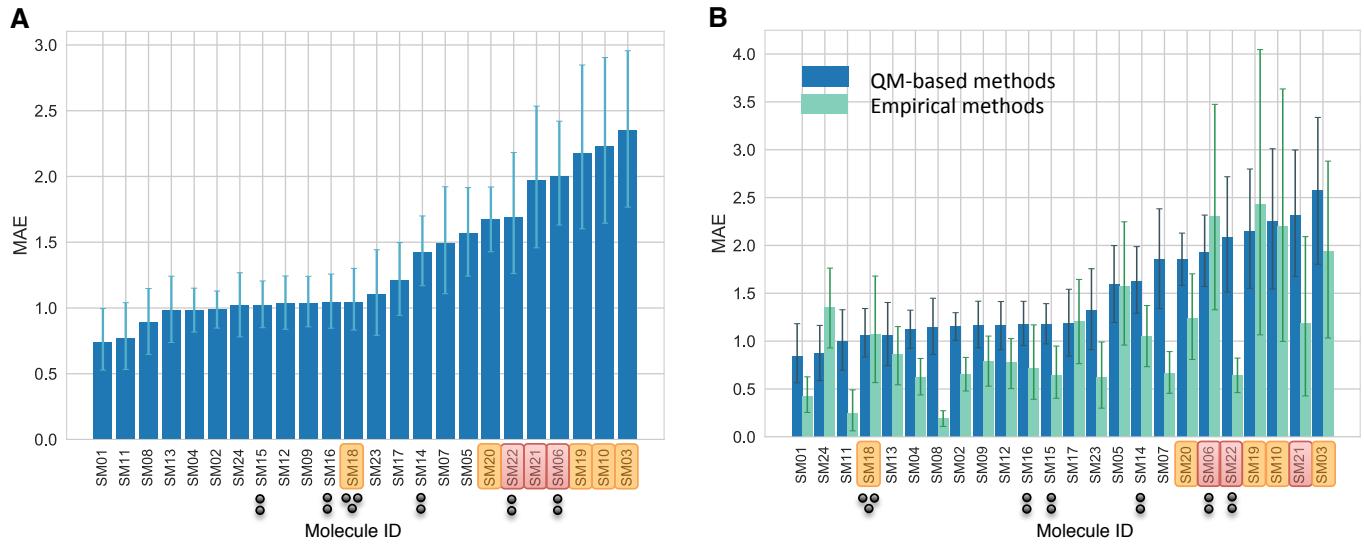
214 Demonstrate how numerical matching often masks the error Match by Hungarian and calculate accuracy of microstate prediction  
overall. When matched by pKa value, do people come with the same transition pairs?



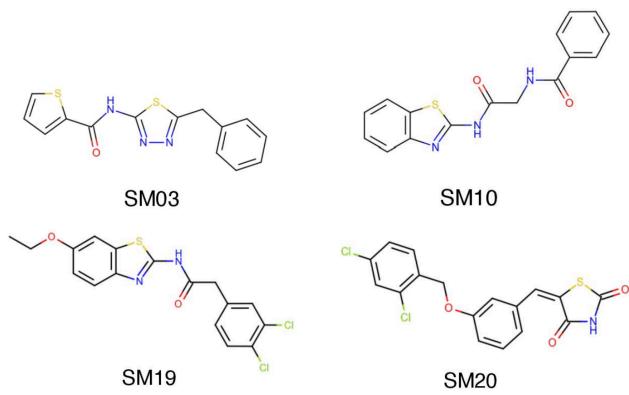
**Figure 4. Predicted vs. experimental value correlation plots of 4 consistently well-performing methods, a representative method with average performance (2ii2g), and the null method (5nm4j).** Dark and light green shaded areas indicate 0.5 and 1.0 units of error. Error bars indicate standard error 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.



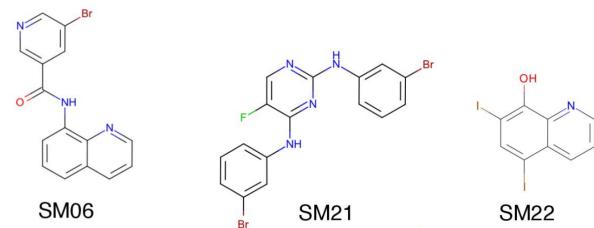
**Figure 5. Molecules of SAMPL6 Challenge with MAE calculated for all macroscopic  $pK_a$  predictions.** MAE calculated considering all prediction methods indicate which molecules had the lowest prediction accuracy in SAMPL6 Challenge. MAE values calculated for each molecule include all the matched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



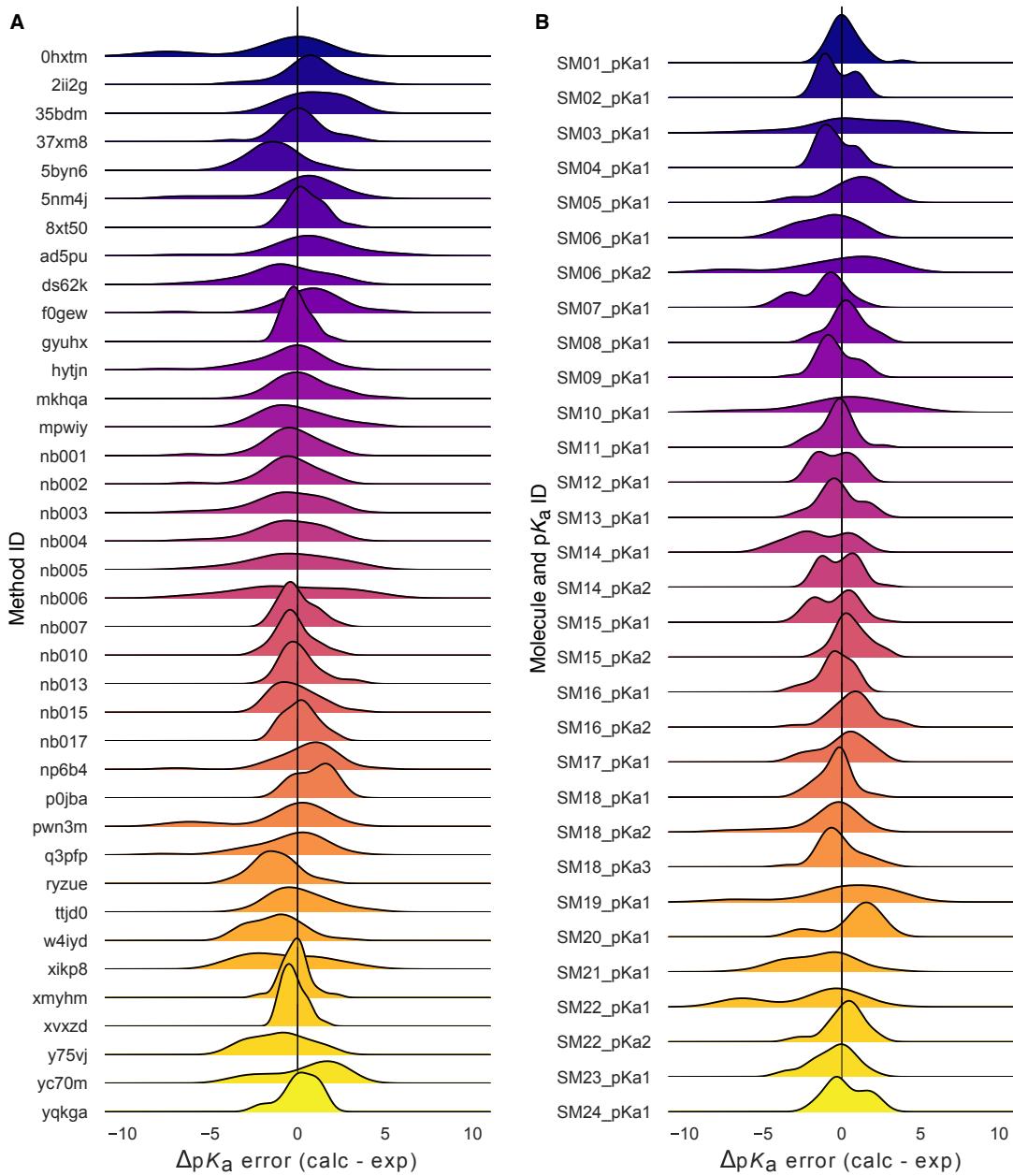
**C SAMPL6 molecules with sulfur-containing heterocycles**



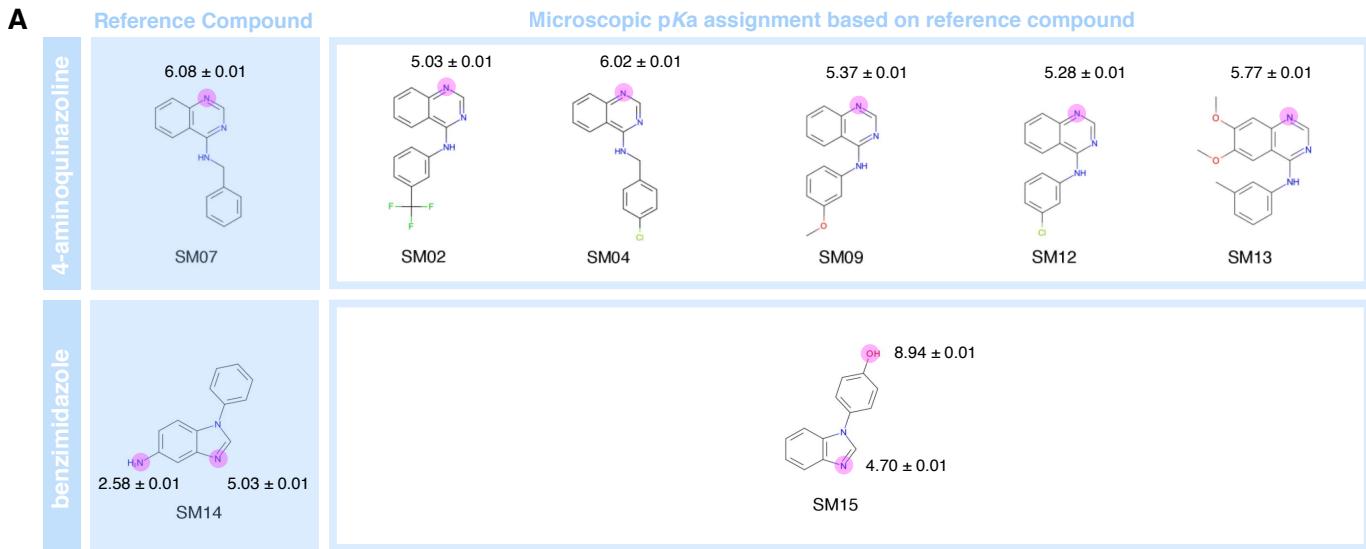
**D SAMPL6 molecules with bromo and iodo groups**



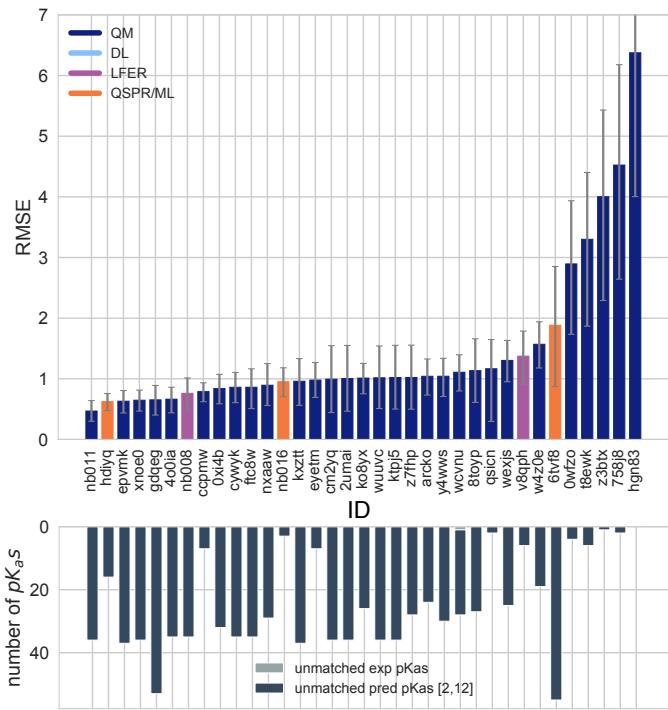
**Figure 6. Average prediction accuracy calculated over all prediction methods was lower for molecules with sulfur-containing heterocycles, bromo, and iodo groups.** (A) MAE calculated for each molecule as an average of all methods. (B) MAE of each molecule broken out by method category. QM-based methods (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.



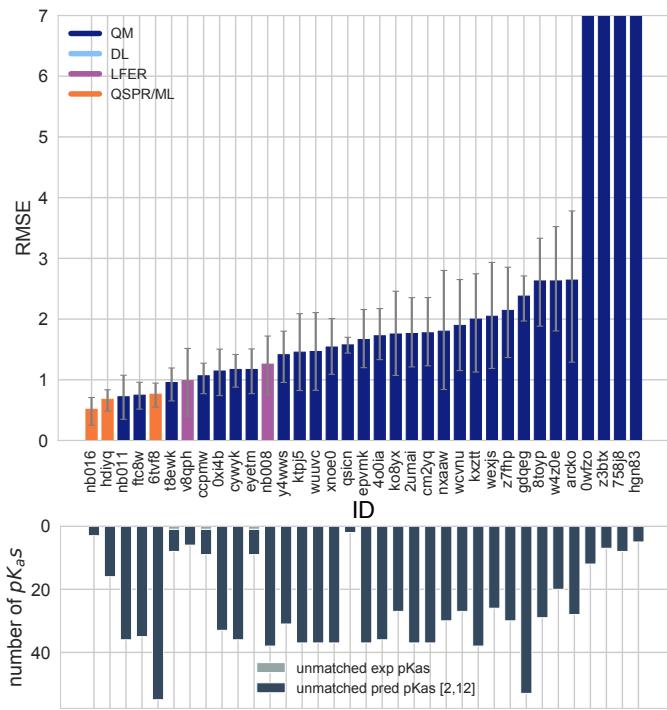
**Figure 7. Macroscopic  $pK_a$  prediction error distribution plots show how prediction accuracy varies across methods and individual molecules.** (A)  $pK_a$  prediction error distribution for each submission for all molecules according to Hungarian matching. (B) Error distribution for each SAMPL6 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.



**B Hungarian matching**



**C Microstate-based matching**



**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 inferred 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  $pK_a$ s. **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. Submissions *Owfzo*, *z3btx*, *758j8*, and *hgn83* have RMSE values bigger than 10  $pK_a$  units which are beyond the y-axis limits of subplot **C** and **B**. 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.

215 MI: SI 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.

216 Discussion of matching experimental and predicted values

217 Difficulty of assessing predicted pKas using experimental data: matching problem

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

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

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

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

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

### 229 3.2.2 Accuracy of predicted pKa values when microstate matching is used

230 Assessment of individual methods by each of our analysis methods

231 Performance comparison of different methods, grouped by methods class

232 Comment on the ranking of microscopic pKa prediction error statistics for all participants (8 mol, microstate match). Refer to Fig. 9

### 233 3.2.3 Dominant microstate prediction accuracy of methods

234 Calculate relative free energy of microstates to determine dominant microstate of each charge Compare predicted and experi-  
235 mental dominant microstates and calculate accuracy of each method

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

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

### 239 3.2.4 Which molecules caused lower dominant microstate prediction accuracy?

240 Which molecule has more errors in predicting the major microstates?

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

## 242 3.3 Analyzing microscopic pKa prediction from the perspective of thermodynamics

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

245 Thermodynamic cycle closure checking allows evaluation of microscopic pKas without experimental data. Checking for ther-  
246 modynamic consistency

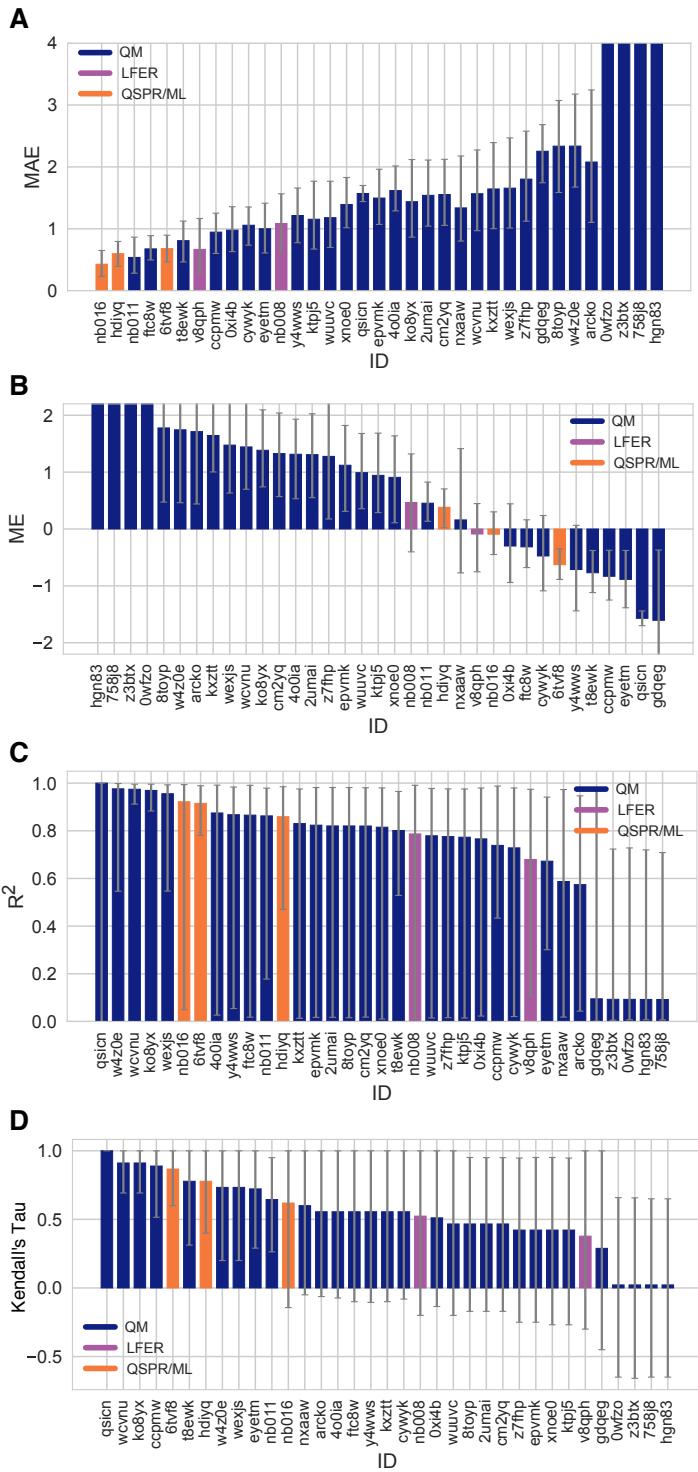
### 247 3.3.1 Cycle closure error

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

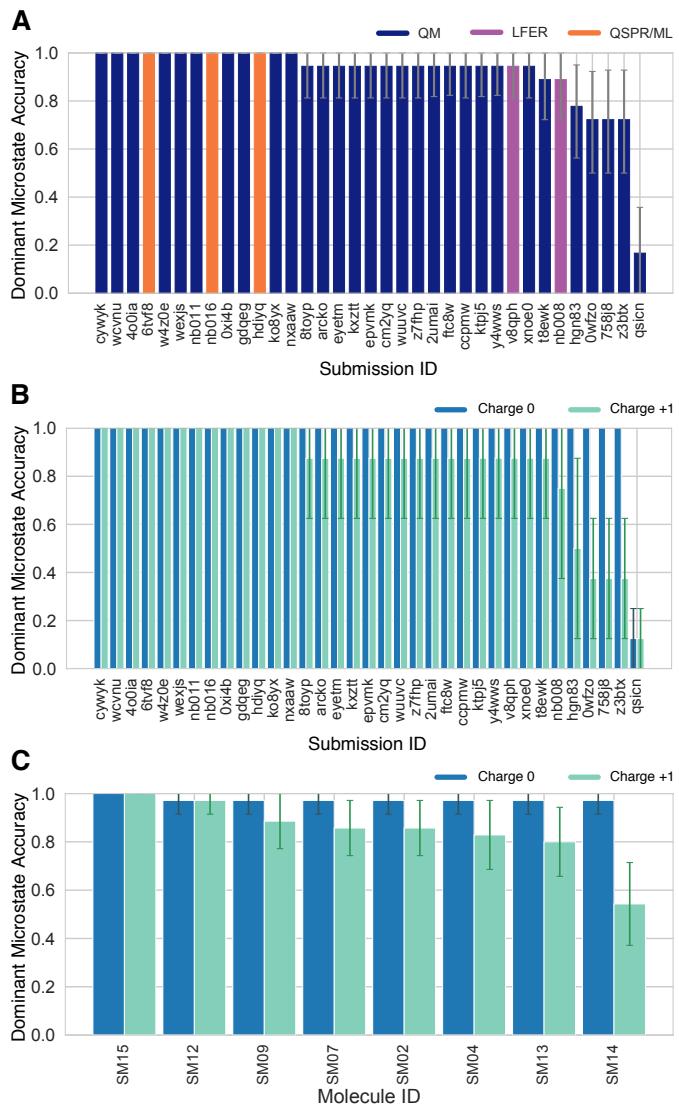
249 She suggesting checking the cycle with maximum cycle closure error for each method and reporting that for each method.

250 An histogram of max cycle closure error will help us bin these results into 3 categories: 1. good agreement 2. moderate 3. severe

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

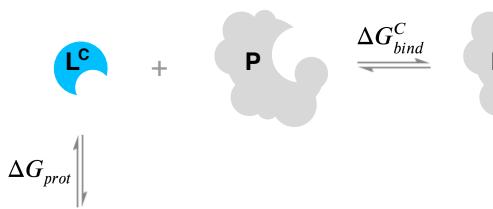


**Figure 9. Additional performance statistics for microscopic  $pK_a$  predictions for 8 molecules with experimentally determined dominant microstates.** Microstate-based matching was performed between experimental  $pK_a$  values and predicted microscopic  $pK_a$  values. Mean absolute error (MAE), mean error (ME), Pearson's  $R^2$ , and Kendall's Rank Correlation Coefficient  $\tau$  are shown, with error bars denoting 95% confidence intervals obtained by bootstrapping over challenge molecules. Methods are indicated by submission IDs. Submissions are colored by their method categories. Refer to Table 1 for submission IDs and method names. Submissions 0wfzo, z3btx, 758j8, and hgn83 have MAE and ME values bigger than 10  $pK_a$  units which are beyond the y-axis limits of subplots A and B. A large number and wide variety of methods have a statistically indistinguishable performance based on correlation based statistic (C and D), in part because of the relatively small dynamic range the small size of the set of 8 molecules.



**Figure 10. Some methods predicted the sequence of dominant tautomers inaccurately.** Prediction accuracy of dominant microstate of each charged state was calculated using the dominant microstate sequence determined by NMR for 8 molecules as reference. **(A)** Dominant microstate accuracy vs. submission ID plot was calculated considering all the dominant microstates seen in the 8 molecule experimental microstate dataset. **(B)** Dominant microstate accuracy vs. submission ID plot was generated considering only the dominant microstates of charge 0 and +1 seen in the 8 molecule experimental microstate dataset. Accuracy of each molecule is broken out by total charge of the microstate. **(C)** Dominant microstate prediction accuracy calculated for each molecule averaged over all methods. In **(B)** and **(C)**, the accuracy of predicting the dominant neutral tautomer is showed in blue and the accuracy of predicting the dominant +1 charged tautomer is showed in green. Error bars denoting 95% confidence intervals obtained by bootstrapping.

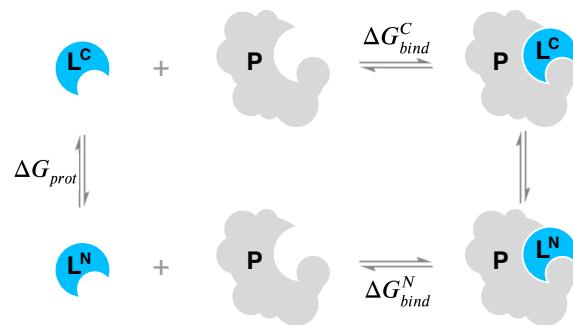
**A** When only the minor protonation state can bind to the protein



$$\Delta G_{bind} = \Delta G_{bind}^C + \Delta G_{prot}$$

$$\Delta G_{bind} = \Delta G_{bind}^C + RT(pH - pK_a) \ln(10)$$

**B** When multiple protonation states can bind to the protein



$$\Delta G_{bind} = \Delta G_{bind}^N + \Delta G_{corr}$$

$$\Delta G_{bind} = \Delta G_{bind}^N - RT \ln \frac{1 + e^{-\frac{\Delta G_{bind}^C - \Delta G_{bind}^N}{RT}} 10^{pK_a - pH}}{1 + 10^{pK_a - pH}}$$

**Figure 11. Aqueous  $pK_a$  of the ligand can influence overall protein-ligand binding affinity.** **A** When only the minor aqueous protonation state contributes to protein-ligand complex formation, overall binding free energy ( $\Delta G_{bind}$ ) needs to be calculated as the sum of binding affinity of the minor state and the protonation penalty of that state. **B** When multiple charge states contribute to complex formation, overall free energy of binding includes a multiple protonation states correction (MPSC) term ( $\Delta G_{corr}$ ). MPSC is a function of pH, aqueous  $pK_a$  of the ligand, and the difference between the binding free energy of charged and neutral species ( $\Delta G_{bind}^C - \Delta G_{bind}^N$ ).

### 3.4 How would $pK_a$ errors affect protein-ligand binding affinity predictions?

Illustrate the ways in which the  $pK_a$  errors can influence prediction errors for binding affinities

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

$$\Delta G_{bind} = \Delta G_{bind}^C + \Delta G_{prot}$$

$$\Delta G_{bind} = \Delta G_{bind}^C + RT(pH - pK_a) \ln(10)$$

$$\Delta G_{bind} = \Delta G_{bind}^N + \Delta G_{corr}$$

$$\Delta G_{bind} = \Delta G_{bind}^N - RT \ln \frac{1 + e^{-\frac{\Delta G_{bind}^C - \Delta G_{bind}^N}{RT}} 10^{pK_a - pH}}{1 + 10^{pK_a - pH}}$$

### 3.5 Lessons learned from SAMPL6 $pK_a$ Challenge

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

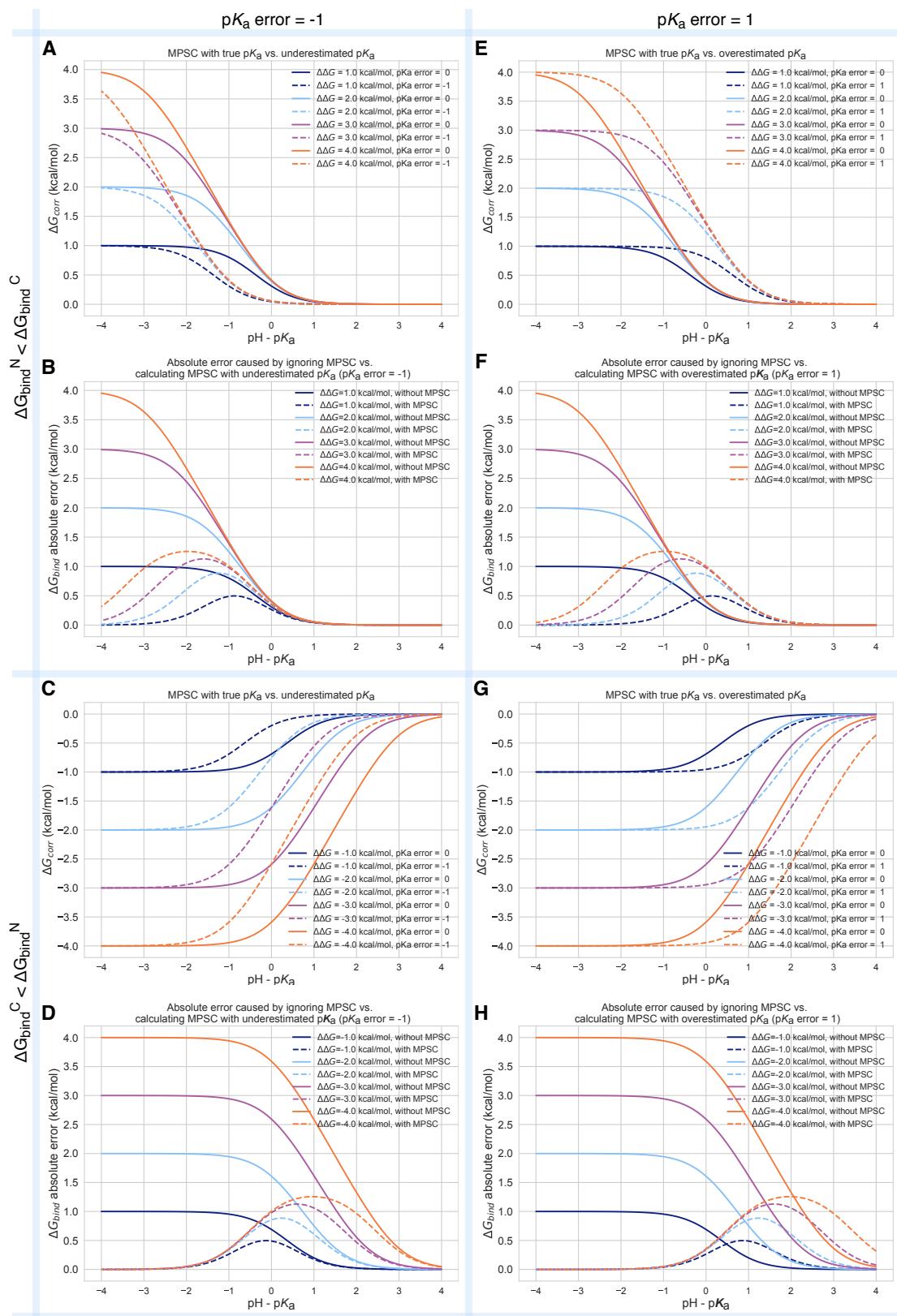
Common challenging factors for accurate  $pK_a$  predictions. Tautomers, Heterocycles etc.

Overall results: Do any methods predict within experimental accuracy (how is the field doing overall)? Common challenging factors for accurate  $pK_a$  predictions. Tautomers, Heterocycles etc.

Discussion of matching problem between experimental and predicted values. Difficulty of assessing predicted  $pK_a$ s 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  $pK_a$  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  $pK_a$  predictions. Which chemical structures make  $pK_a$  predictions more difficult?

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



**Figure 12. Inaccuracy of  $pK_a$  prediction ( $\pm 1$  unit) affects the accuracy of MPSC and overall protein-ligand binding free energy calculation in varying amounts based on aqueous  $pK_a$  value and relative binding affinity of individual protonation states ( $\Delta\Delta G = \Delta G_{bind}^C - \Delta G_{bind}^N$ ). All calculations are made for 25°C, and for a ligand with single basic titratable group. **A, C, E, and G** show MPSC ( $\Delta G_{corr}$ ) calculated with true vs. inaccurate  $pK_a$ . **B, D, F, and H** show comparison of the absolute error to  $\Delta G_{bind}$  caused by ignoring the MPSC completely (solid lines) vs. calculating MPSC based in inaccurate  $pK_a$  value (dashed lines). These plots provide guidance on when it is beneficial to include MPSC correction based on  $pK_a$  error,  $pH - pK_a$ , and  $\Delta\Delta G$ .**

### 268 3.6 Suggestions for future challenges

269 Discuss what can be done to further improve future challenges

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

272 Suggestions about challenge construction

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

274 Suggestions about challenge analysis

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

276 Reporting microscopic pKa predictions with charges, microstate free energies is better Experimental dataset with microstate  
277 information is more helpful.

278 What can be done to further improve future challenges How can we maximize what we learn? What should we have people  
279 predict? How should we select compounds / measure pKas? NMR experimental techniques could be used to validate microstate  
280 information in future challenges

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

## 283 4 Conclusion

## 284 5 Code and data availability

- 285 SAMPL6 pK<sub>a</sub> challenge instructions, submissions, experimental data and analysis is available at  
<https://github.com/samplchallenges/SAMPL6>

## 286 6 Overview of supplementary information

287 Contents of the Supplementary Information:

- 288 TABLE S1: SMILES and InChI identifiers of SAMPL6 pK<sub>a</sub> Challenge molecules.
- 289 TABLE S2: Evaluation statistics calculated for all macroscopic pK<sub>a</sub> prediction submissions based on Hungarian match for  
290 24 molecules.
- 291 TABLE S3: Evaluation statistics calculated for all microscopic pK<sub>a</sub> prediction submissions based on Hungarian match for 8  
292 molecules with NMR data.
- 293 TABLE S4: Evaluation statistics calculated for all microscopic pK<sub>a</sub> prediction submissions based on microstate match for 8  
294 molecules with NMR data.
- 295 FIGURE S1: Dominant microstates of 8 molecules were determined based on NMR measurements.
- 296 FIGURE S2: MAE of macroscopic pK<sub>a</sub> predictions of each molecule did not show any significant correlation with any molecular  
297 descriptor.
- 298 FIGURE S3: The value of macroscopic pK<sub>a</sub> was not a factor affecting prediction error seen in SAMPL6 Challenge according  
299 to the analysis with Hungarian matching.

300 Extra files included in *SAMPL6-supplementary-documents.tar.gz*:

- 301 SAMPL6-pKa-chemical-identifiers-table.csv
- 302 macroscopic-pKa-statistics-24mol-hungarian-match.csv
- 303 microscopic-pKa-statistics-8mol-hungarian-match-table.csv
- 304 microscopic-pKa-statistics-8mol-microstate-match-table.csv
- 305 experimental-microstates-of-8mol-based-on-NMR.csv

## 306 7 Author Contributions

307 Conceptualization, MI, JDC, CB, DLM ; Methodology, MI, JDC ; Software, MI, AR, ASR ; Formal Analysis, MI, ASR, AR ; Investigation,  
308 MI ; Resources, JDC; Data Curation, MI ; Writing-Original Draft, MI, JDC; Writing - Review and Editing, MI, ASR, AR, CB, DLM, JDC;  
309 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: [], +, -, \*, #, \m

## 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](https://doi.org/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](https://doi.org/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](https://doi.org/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, Sforza G, Cruciani G. New and Original  $pK_a$  Prediction Method Using Grid Molecular Interaction Fields. *Journal of Chemical Information and Modeling*. 2007 Nov; 47(6):2172–2181. doi: [10.1021/ci700018y](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/10.1007/s10822-018-0140-z).

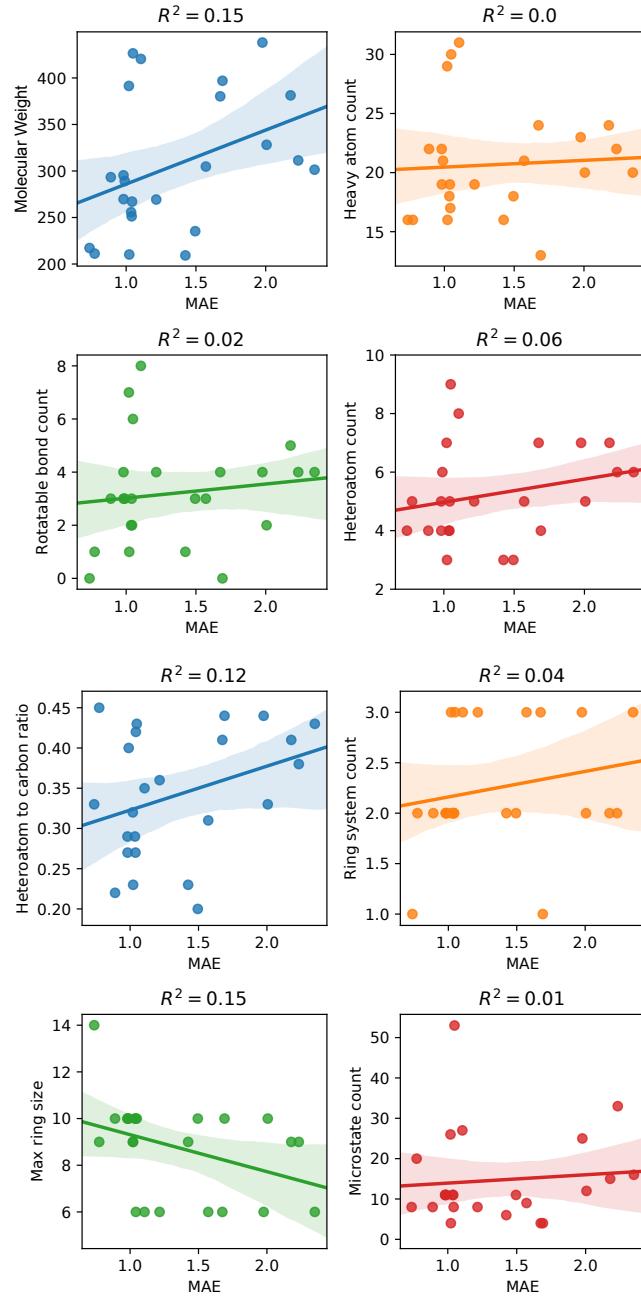
- 353 [15] Klamt A, Eckert F, Diedenhofen M, Beck ME. First Principles Calculations of Aqueous p  $K_a$  Values for Organic and Inorganic Acids Using  
354 COSMO-RS Reveal an Inconsistency in the Slope of the p  $K_a$  Scale. *The Journal of Physical Chemistry A*. 2003 Nov; 107(44):9380–9386. doi:  
355 [10.1021/jp034688o](https://doi.org/10.1021/jp034688o).
- 356 [16] Eckert F, Klamt A. Accurate Prediction of Basicity in Aqueous Solution with COSMO-RS. *Journal of Computational Chemistry*. 2006 Jan;  
357 27(1):11–19. doi: [10.1002/jcc.20309](https://doi.org/10.1002/jcc.20309).
- 358 [17] Pracht P, Wilcken R, Udvarhelyi A, Rodde S, Grimme S. High Accuracy Quantum-Chemistry-Based Calculation and Blind Prediction of  
359 Macroscopic pKa Values in the Context of the SAMPL6 Challenge. *Journal of Computer-Aided Molecular Design*. 2018 Oct; 32(10):1139–  
360 1149. doi: [10.1007/s10822-018-0145-7](https://doi.org/10.1007/s10822-018-0145-7).
- 361 [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  
362 Challenge. *Journal of Computer-Aided Molecular Design*. 2018 Oct; 32(10):1191–1201. doi: [10.1007/s10822-018-0167-1](https://doi.org/10.1007/s10822-018-0167-1).
- 363 [19] OEMolProp Toolkit 2017.Feb.1;. OpenEye Scientific Software, Santa Fe, NM. <http://www.eyesopen.com>.

**Table S1. SMILES and InChI identifiers of SAMPL6 pK<sub>a</sub> Challenge molecules.** A CSV version of this table can be found in *SAMPL6-supplementary-documents.tar.gz*.

SAMPL6 Molecule ID	Isomeric SMILES	InChI
SM01	c1cc2c(cc1O)c3c(o2)C(=O)NCCC3	InChI=1S/C12H11NO3/c14-7-3-4-10-9(6-7)8-2-1-5-13-12(15)11(8)16-10/h3-4,6,14H,1-2,5H2,(H,13,15)
SM02	c1ccc2c(c1)c(ncn2)Nc3cccc(c3)C(F)(F)	InChI=1S/C15H10F3N3/c16-15(17,18)10-4-3-5-11(8-10)21-14-12-6-1-2-7-13(12)19-9-20-14/h1-9H,(H,19,20,21)
SM03	c1ccc(cc1)Cc2nnnc(s2)NC(=O)c3cccs3	InChI=1S/C14H11N3OS2/c18-13(11-7-4-8-19-11)15-14-17-16-12(20-14)9-10-5-2-1-3-6-10/h1-8H,9H2,(H,15,17,18)
SM04	c1ccc2c(c1)c(ncn2)NCc3ccc(cc3)Cl	InChI=1S/C15H12ClN3/c16-12-7-5-11(6-8-12)9-17-15-13-3-1-2-4-14(13)18-10-19-15/h1-8,10H,9H2,(H,17,18,19)
SM05	c1ccc(c(c1)NC(=O)c2ccc(o2)Cl)N3CCCCC3	InChI=1S/C16H17ClN2O2/c17-15-9-8-14(21-15)16(20)18-12-6-2-3-7-13(12)19-10-4-1-5-11-19/h2-3,6-9H,1,4-5,10-11H2,(H,18,20)
SM06	c1cc2ccnc2c(c1)NC(=O)c3cc(cnc3)Br	InChI=1S/C15H10BrN3O/c16-12-7-11(8-17-9-12)15(20)19-13-5-1-3-10-4-2-6-18-14(10)13/h1-9H,(H,19,20)
SM07	c1ccc(cc1)CNc2c3cccc3ncn2	InChI=1S/C15H13N3/c1-2-6-12(7-3-1)10-16-15-13-8-4-5-9-14(13)17-11-18-15/h1-9,11H,10H2,(H,16,17,18)
SM08	Cc1ccc2c(c1)c(c(c(=O)[nH]2)CC(=O)O)c3cccc3	InChI=1S/C18H15NO3/c1-11-7-8-15-13(9-11)17(12-5-3-2-4-6-12)14(10-16(20)21)18(22)19-15/h2-9H,10H2,1H3,(H,19,22)(H,20,21)
SM09	COc1cccc(c1)Nc2c3cccc3ncn2.Cl	InChI=1S/C15H13N3O.CIH/c1-19-12-6-4-5-11(9-12)18-15-13-7-2-3-8-14(13)16-10-17-15;/h2-10H,1H3,(H,16,17,18);1H
SM10	c1ccc(cc1)C(=O)NCC(=O)Nc2nc3cccc3s2	InChI=1S/C16H13N3O2S/c20-14(10-17-15(21)11-6-2-1-3-7-11)19-16-18-1-2-8-4-5-9-13(12)22-16/h1-9H,10H2,(H,17,21)(H,18,19,20)
SM11	c1ccc(cc1)n2c3c(cn2)c(ncn3)N	InChI=1S/C11H9N5/c12-10-9-6-15-16(11(9)14-7-13-10)8-4-2-1-3-5-8/h1-7H,(H,2,12,13,14)
SM12	c1ccc2c(c1)c(ncn2)Nc3cccc(c3)Cl.Cl	InChI=1S/C14H10ClN3.CIH/c15-10-4-3-5-11(8-10)18-14-12-6-1-2-7-13(12)16-9-17-14;/h1-9H,(H,16,17,18);1H
SM13	Cc1cccc(c1)Nc2c3cc(c(c3ncn2)OC)OC	InChI=1S/C17H17N3O2/c1-11-5-4-6-12(7-11)20-17-13-8-15(21-2)16(22-3)9-14(13)18-10-19-17/h4-10H,1-3H3,(H,18,19,20)
SM14	c1ccc(cc1)n2ncn3c2ccc(c3)N	InChI=1S/C13H11N3/c14-10-6-7-13-12(8-10)15-9-16(13)11-4-2-1-3-5-11/h1-9H,14H2
SM15	c1ccc2c(c1)ncn2c3ccc(cc3)O	InChI=1S/C13H10N2O/c16-11-7-5-10(6-8-11)15-9-14-12-3-1-2-4-13(12)15/h1-9,16H
SM16	c1cc(c(c(c1)Cl)C(=O)Nc2ccncc2)Cl	InChI=1S/C12H8Cl2N2O/c13-9-2-1-3-10(14)11(9)12(17)16-8-4-6-15-7-5-8/h1-7H,(H,15,16,17)
SM17	c1ccc(cc1)CSc2nnc(o2)c3ccncc3	InChI=1S/C14H11N3OS/c1-2-4-11(5-3-1)10-19-14-17-16-13(18-14)12-6-8-15-9-7-12/h1-9H,10H2
SM18	c1ccc2c(c1)c(=O)[nH]c(n2)CCC(=O)Nc3ncc(s3)Cc4ccc(c(c4)F)F	InChI=1S/C21H16F2N4O2S/c22-15-6-5-12(10-16(15)23)9-13-11-24-21(30-13)27-19(28)8-7-18-25-17-4-2-1-3-14(17)20(29)26-18/h1-6,10-11H,7-9H2,(H,24,27,28)(H,25,26,29)
SM19	CCOc1ccc2c(c1)sc(n2)NC(=O)Cc3ccc(c(c3)Cl)Cl	InChI=1S/C17H14Cl2N2O2S/c1-2-23-11-4-6-14-15(9-11)24-17(20-14)21-6(22)8-10-3-5-12(18)13(9)7-10/h3-7,9H,2,8H2,1H3,(H,20,21,22)
SM20	c1cc(cc(c1)OCc2ccc(cc2Cl)Cl)/C=C/3\C(=O)NC(=O)S3	InChI=1S/C17H11Cl2NO3S/c18-12-5-4-11(14(19)8-12)9-23-13-3-1-2-10(6-13)7-15-16(21)20-17(22)24-15/h1-8H,9H2,(H,20,21,22)/b15-7+
SM21	c1cc(cc(c1)Br)Nc2c(cnc(n2)Nc3cccc(c3)Br)F	InChI=1S/C16H11Br2FN4/c17-10-3-1-5-12(7-10)21-15-14(19)9-20-16(23-15)22-13-6-2-4-11(18)8-13/h1-9H,(H,20,21,22,23)
SM22	c1cc2c(cc(c(c2nc1)O))l	InChI=1S/C9H5l2NO/c10-6-4-7(11)9(13)8-5(6)2-1-3-12-8/h1-4,13H
SM23	CCOC(=O)c1ccc(cc1)Nc2cc(cnc(n2)Nc3ccc(cc3)C(=O)OCC)C	InChI=1S/C23H24N4O4/c1-4-30-21(28)16-6-10-18(11-7-16)25-20-14-15(3)24-23(27-20)26-19-12-8-17(9-13-19)22(29)31-5-2/h6-14H,4-5H2,1-3H3,(H2,24,25,26,27)
SM24	COc1ccc(cc1)c2c3c(ncn3oc2c4ccc(cc4)OC)NCCO	InChI=1S/C22H21N3O4/c1-27-16-7-3-14(4-8-16)18-19-21(23-11-12-26)24-13-25-22(19)29-20(18)15-5-9-17(28-2)10-6-15/h3-10,13,26H,11-12H2,1-2H3,(H,23,24,25)

Microstate ID of Deprotonated State (A)	Microstate ID of Protonated State (HA)	Molecule ID	pKa (exp)	pKa SEM (exp)	pKa ID	Microstate identification source
		SM07	6.08	0.01	SM07_pKa1	NMR measurement
		SM14	5.3	0.01	SM14_pKa2	NMR measurement
		SM14	2.58	0.01	SM14_pKa1	NMR measurement
		SM02	5.03	0.01	SM02_pKa1	Estimated based on SM07 NMR measurement
		SM04	6.02	0.01	SM04_pKa1	Estimated based on SM07 NMR measurement
		SM09	5.37	0.01	SM09_pKa1	Estimated based on SM07 NMR measurement
		SM12	5.28	0.01	SM12_pKa1	Estimated based on SM07 NMR measurement
		SM13	5.77	0.01	SM13_pKa1	Estimated based on SM07 NMR measurement
		SM15	8.94	0.01	SM15_pKa2	Estimated based on SM14 NMR measurement
		SM15	4.7	0.01	SM15_pKa1	Estimated based on SM14 NMR measurement

**Figure S1. Dominant microstates of 8 molecules were determined based on NMR measurements.** Dominant microstate sequence of 6 derivatives were determined taking SM07 and SM14 as reference. Matched experimental pK<sub>a</sub> values were determined by spectrophotometric pK<sub>a</sub> measurements [1]. A CSV version of this table can be found in SAMPL6-supplementary-documents.tar.gz.



**Figure S2. MAE of macroscopic  $pK_a$  predictions of each molecule did not show any significant correlation with any molecular descriptor.**  
 Plots show regression lines, 96% confidence intervals of the regression lines, and  $R_2$ . The following molecular descriptors were calculated using OpenEye OEMolProp Toolkit [19].

**Table S2. Evaluation statistics calculated for all macroscopic pK<sub>a</sub> prediction submissions based on Hungarian match for 24 molecules.** Methods are represented via their SAMPL6 submission IDs which can be cross referenced with Table 1 for method details. There are eight error metrics reported: the root-mean-squared error (RMSE), mean absolute error (MAE), mean (signed) error (ME), coefficient of determination ( $R^2$ ), linear regression slope (m), Kendall's Rank Correlation Coefficient ( $\tau$ ), unmatched experimental pK<sub>a</sub>s (number of missing pK<sub>a</sub> predictions) and unmatched predicted pK<sub>a</sub>s (number of extra pK<sub>a</sub> predictions between 2 and 12. This table is ranked by increasing RMSE. A CSV version of this table can be found in *SAMPL6-supplementary-documents.tar.gz*.

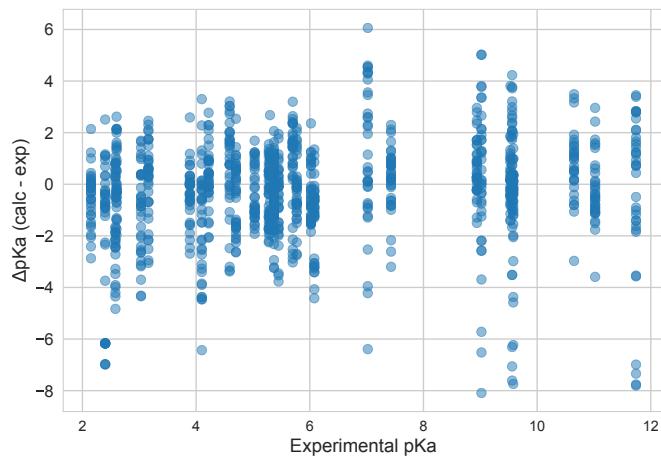
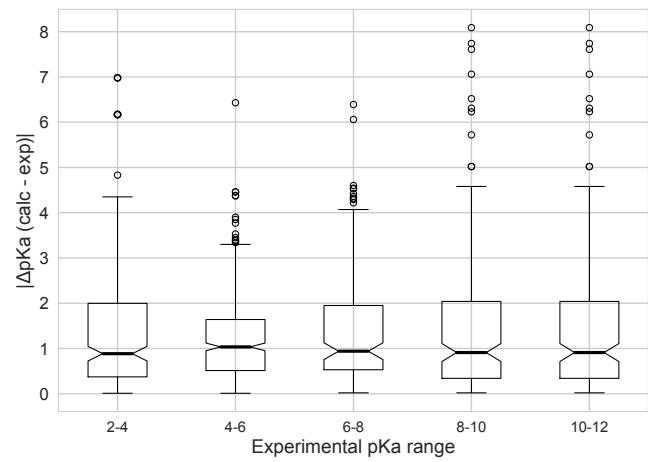
Submission ID	RMSE	MAE	ME	R <sup>2</sup>	m	Kendall's Tau	Unmatched exp. pK <sub>a</sub> s	Unmatched pred. pK <sub>a</sub> s [2,12]
<i>vxxzd</i>	0.68 [0.54, 0.81]	0.58 [0.45, 0.71]	0.24 [-0.01, 0.45]	0.94 [0.88, 0.97]	0.92 [0.84, 1.02]	0.82 [0.68, 0.92]	2	4
<i>gyuhx</i>	0.73 [0.55, 0.91]	0.59 [0.44, 0.74]	0.03 [-0.23, 0.28]	0.93 [0.88, 0.96]	0.98 [0.90, 1.08]	0.88 [0.80, 0.94]	0	7
<i>xmyhm</i>	0.79 [0.52, 1.03]	0.56 [0.38, 0.77]	0.13 [-0.14, 0.41]	0.92 [0.85, 0.97]	0.96 [0.86, 1.08]	0.81 [0.68, 0.90]	0	3
<i>nb017</i>	0.94 [0.72, 1.16]	0.77 [0.58, 0.97]	-0.16 [-0.49, 0.16]	0.88 [0.81, 0.94]	0.94 [0.82, 1.08]	0.73 [0.60, 0.84]	0	6
<i>nb007</i>	0.95 [0.73, 1.15]	0.78 [0.60, 0.97]	0.05 [-0.29, 0.37]	0.88 [0.77, 0.95]	0.84 [0.77, 0.92]	0.79 [0.65, 0.89]	0	13
<i>yqkga</i>	1.01 [0.78, 1.23]	0.80 [0.59, 1.03]	-0.17 [-0.51, 0.19]	0.87 [0.78, 0.93]	0.93 [0.77, 1.08]	0.83 [0.72, 0.91]	0	1
<i>nb010</i>	1.03 [0.77, 1.26]	0.81 [0.61, 1.04]	0.24 [-0.11, 0.59]	0.87 [0.77, 0.94]	0.95 [0.83, 1.08]	0.80 [0.67, 0.90]	0	4
<i>8xt50</i>	1.07 [0.78, 1.36]	0.81 [0.58, 1.07]	-0.47 [-0.82, -0.14]	0.91 [0.84, 0.95]	1.08 [0.94, 1.22]	0.80 [0.68, 0.89]	0	0
<i>nb013</i>	1.10 [0.72, 1.47]	0.80 [0.56, 1.09]	-0.15 [-0.55, 0.22]	0.88 [0.78, 0.95]	1.09 [0.90, 1.25]	0.79 [0.64, 0.90]	0	6
<i>nb015</i>	1.27 [0.98, 1.56]	1.04 [0.80, 1.31]	0.13 [-0.32, 0.56]	0.87 [0.80, 0.93]	1.16 [0.94, 1.34]	0.78 [0.66, 0.86]	0	0
<i>p0jba</i>	1.31 [0.69, 1.73]	1.08 [0.43, 1.72]	-0.92 [-1.72, -0.11]	0.91 [0.51, 1.00]	1.18 [0.36, 1.72]	0.80 [0.00, 1.00]	0	0
<i>37xrn8</i>	1.41 [0.93, 1.84]	1.01 [0.68, 1.38]	-0.18 [-0.69, 0.32]	0.83 [0.70, 0.93]	1.16 [0.98, 1.33]	0.70 [0.56, 0.83]	1	1
<i>mkhqa</i>	1.60 [1.13, 2.05]	1.24 [0.90, 1.62]	-0.32 [-0.89, 0.21]	0.80 [0.67, 0.91]	1.14 [0.98, 1.34]	0.64 [0.44, 0.79]	0	6
<i>ttjd0</i>	1.64 [1.20, 2.06]	1.30 [0.96, 1.67]	-0.12 [-0.70, 0.45]	0.81 [0.69, 0.91]	1.2 [1.03, 1.40]	0.65 [0.47, 0.80]	0	5
<i>nb001</i>	1.68 [1.05, 2.37]	1.21 [0.84, 1.68]	0.44 [-0.10, 1.03]	0.80 [0.70, 0.90]	1.16 [0.95, 1.42]	0.72 [0.55, 0.85]	0	7
<i>nb002</i>	1.70 [1.08, 2.38]	1.25 [0.89, 1.70]	0.51 [-0.04, 1.10]	0.80 [0.70, 0.90]	1.15 [0.95, 1.42]	0.72 [0.56, 0.84]	0	7
<i>35bdm</i>	1.72 [0.66, 2.34]	1.44 [0.62, 2.26]	-1.01 [-2.18, 0.13]	0.92 [0.46, 1.00]	1.45 [0.73, 2.15]	0.80 [0.00, 1.00]	0	0
<i>ryzue</i>	1.77 [1.42, 2.12]	1.50 [1.17, 1.84]	1.30 [0.86, 1.72]	0.91 [0.86, 0.95]	1.23 [1.06, 1.41]	0.82 [0.71, 0.91]	0	0
<i>2ii2g</i>	1.80 [1.31, 2.24]	1.39 [1.01, 1.82]	-0.74 [-1.29, -0.15]	0.79 [0.65, 0.89]	1.15 [0.96, 1.37]	0.68 [0.59, 0.82]	0	2
<i>mpwiy</i>	1.82 [1.39, 2.23]	1.48 [1.14, 1.88]	0.10 [-0.54, 0.73]	0.82 [0.70, 0.91]	1.29 [1.12, 1.51]	0.66 [0.49, 0.80]	0	5
<i>5byn6</i>	1.89 [1.50, 2.27]	1.59 [1.24, 1.97]	1.32 [0.84, 1.80]	0.91 [0.85, 0.95]	1.28 [1.10, 1.48]	0.83 [0.72, 0.92]	0	0
<i>y75vj</i>	1.90 [1.50, 2.26]	1.58 [1.21, 1.97]	1.04 [0.46, 1.60]	0.89 [0.79, 0.95]	1.34 [1.16, 1.53]	0.75 [0.57, 0.88]	1	0
<i>w4iyd</i>	1.93 [1.53, 2.28]	1.58 [1.20, 1.98]	1.26 [0.72, 1.76]	0.85 [0.74, 0.92]	1.21 [1.00, 1.40]	0.73 [0.57, 0.85]	0	1
<i>np6b4</i>	1.94 [1.21, 2.71]	1.44 [1.04, 1.94]	-0.47 [-1.08, 0.24]	0.71 [0.60, 0.87]	1.08 [0.81, 1.43]	0.75 [0.62, 0.86]	0	8
<i>nb004</i>	2.01 [1.38, 2.63]	1.57 [1.16, 2.04]	0.56 [-0.10, 1.27]	0.82 [0.72, 0.90]	1.35 [1.15, 1.60]	0.71 [0.54, 0.84]	0	5
<i>nb003</i>	2.01 [1.39, 2.64]	1.58 [1.18, 2.04]	0.52 [-0.14, 1.22]	0.82 [0.73, 0.91]	1.36 [1.16, 1.61]	0.71 [0.54, 0.84]	0	5
<i>yc70m</i>	2.03 [1.73, 2.33]	1.80 [1.48, 2.13]	-0.41 [-1.09, 0.31]	0.47 [0.28, 0.64]	0.56 [0.35, 0.83]	0.53 [0.35, 0.68]	0	27
<i>hytjn</i>	2.16 [1.24, 3.06]	1.39 [0.86, 2.04]	0.71 [0.03, 1.48]	0.45 [0.13, 0.78]	0.62 [0.26, 1.00]	0.47 [0.16, 0.73]	1	27
<i>f0gew</i>	2.18 [1.38, 2.95]	1.58 [1.09, 2.16]	-0.73 [-1.42, 0.04]	0.77 [0.67, 0.89]	1.29 [1.01, 1.63]	0.76 [0.63, 0.86]	0	0
<i>q3pfp</i>	2.19 [1.33, 3.09]	1.51 [0.99, 2.13]	0.59 [-0.10, 1.37]	0.44 [0.13, 0.77]	0.66 [0.27, 1.07]	0.50 [0.20, 0.75]	1	22
<i>ds62k</i>	2.22 [1.62, 2.81]	1.78 [1.34, 2.27]	0.78 [0.06, 1.52]	0.82 [0.70, 0.90]	1.41 [1.20, 1.63]	0.72 [0.55, 0.85]	0	4
<i>xikp8</i>	2.35 [1.94, 2.73]	2.06 [1.66, 2.47]	0.77 [-0.02, 1.58]	0.89 [0.80, 0.95]	1.59 [1.40, 1.81]	0.76 [0.59, 0.89]	1	0
<i>nb005</i>	2.38 [1.79, 2.95]	1.91 [1.44, 2.43]	0.31 [-0.49, 1.15]	0.84 [0.74, 0.91]	1.56 [1.34, 1.82]	0.71 [0.54, 0.83]	0	0
<i>5nm4j</i>	2.45 [1.42, 3.34]	1.58 [0.94, 2.34]	0.05 [-0.80, 1.07]	0.19 [0.00, 0.70]	0.40 [-0.06, 0.81]	0.34 [-0.04, 0.67]	4	1
<i>ad5pu</i>	2.54 [1.68, 3.30]	1.83 [1.24, 2.49]	-0.65 [-1.48, 0.25]	0.76 [0.64, 0.88]	1.43 [1.12, 1.78]	0.77 [0.63, 0.88]	0	0
<i>pwn3m</i>	2.60 [1.45, 3.53]	1.54 [0.83, 2.37]	0.79 [-0.06, 1.77]	0.21 [0.00, 0.63]	0.37 [0.01, 0.78]	0.34 [0.04, 0.63]	1	3
<i>nb006</i>	2.98 [2.37, 3.56]	2.53 [2.00, 3.10]	0.42 [-0.60, 1.47]	0.84 [0.74, 0.92]	1.78 [1.55, 2.06]	0.71 [0.54, 0.84]	0	0
<i>0hxtm</i>	3.26 [1.81, 4.39]	1.92 [1.03, 2.98]	1.38 [0.37, 2.56]	0.08 [0.00, 0.48]	0.28 [-0.17, 0.83]	0.29 [-0.04, 0.61]	3	7

**Table S3. Evaluation statistics calculated for all microscopic pK<sub>a</sub> prediction submissions based on Hungarian match for 8 molecules with NMR data.** Methods are represented via their SAMPL6 submission IDs which can be cross referenced with Table 1 for method details. There are eight error metrics reported: the root-mean-squared error (RMSE), mean absolute error (MAE), mean (signed) error (ME), coefficient of determination ( $R^2$ ), linear regression slope (m), Kendall's Rank Correlation Coefficient ( $\tau$ ), unmatched experimental pK<sub>a</sub>s (number of missing pK<sub>a</sub> predictions) and unmatched predicted pK<sub>a</sub>s (number of extra pK<sub>a</sub> predictions between 2 and 12. This table is ranked by increasing RMSE. A CSV version of this table can be found in *SAMPL6-supplementary-documents.tar.gz*.

Submission ID	RMSE	MAE	ME	R <sup>2</sup>	m	Kendall's Tau	Unmatched exp. pK <sub>a</sub> s	Unmatched pred. pK <sub>a</sub> s [2,12]
nb011	0.47 [0.30, 0.64]	0.33 [0.22, 0.46]	-0.02 [-0.18, 0.14]	0.97 [0.94, 0.99]	1.01 [0.97, 1.06]	0.90 [0.78, 0.96]	0	36
hdlyq	0.62 [0.47, 0.76]	0.47 [0.33, 0.62]	0.13 [-0.09, 0.34]	0.95 [0.92, 0.97]	0.34 [0.92, 1.09]	0.87 [0.79, 0.93]	0	16
epvmk	0.63 [0.43, 0.81]	0.47 [0.32, 0.63]	-0.02 [-0.25, 0.21]	0.95 [0.89, 0.98]	0.21 [0.91, 1.04]	0.81 [0.68, 0.91]	0	37
xnoe0	0.65 [0.47, 0.82]	0.50 [0.36, 0.66]	-0.1 [-0.32, 0.13]	0.95 [0.89, 0.98]	0.13 [0.92, 1.05]	0.82 [0.69, 0.91]	0	36
gdqeg	0.65 [0.41, 0.89]	0.43 [0.27, 0.62]	0.11 [-0.10, 0.35]	0.94 [0.88, 0.98]	0.35 [0.87, 1.02]	0.83 [0.67, 0.95]	0	53
400ia	0.66 [0.44, 0.86]	0.47 [0.31, 0.64]	0.00 [-0.22, 0.24]	0.94 [0.88, 0.98]	0.24 [0.87, 1.05]	0.85 [0.73, 0.94]	0	35
nb008	0.76 [0.48, 1.02]	0.52 [0.34, 0.73]	-0.08 [-0.37, 0.17]	0.93 [0.85, 0.98]	0.17 [0.79, 0.93]	0.84 [0.73, 0.92]	0	35
ccpmw	0.79 [0.62, 0.94]	0.62 [0.46, 0.80]	-0.17 [-0.44, 0.11]	0.92 [0.86, 0.96]	0.11 [0.82, 1.05]	0.80 [0.67, 0.89]	0	7
0xi4b	0.84 [0.58, 1.07]	0.61 [0.42, 0.83]	0.22 [-0.07, 0.51]	0.92 [0.84, 0.97]	0.51 [0.91, 1.09]	0.81 [0.65, 0.92]	0	32
cwyk	0.86 [0.60, 1.10]	0.62 [0.42, 0.84]	0.13 [-0.16, 0.44]	0.90 [0.82, 0.96]	0.44 [0.86, 1.08]	0.81 [0.64, 0.92]	0	35
ftc8w	0.86 [0.51, 1.17]	0.59 [0.39, 0.83]	0.10 [-0.19, 0.41]	0.90 [0.77, 0.97]	0.41 [0.84, 0.98]	0.75 [0.57, 0.88]	0	35
nxaaw	0.89 [0.56, 1.25]	0.61 [0.41, 0.87]	-0.02 [-0.35, 0.28]	0.89 [0.75, 0.97]	0.28 [0.85, 1.00]	0.79 [0.63, 0.91]	0	29
nb016	0.95 [0.71, 1.18]	0.77 [0.57, 0.98]	-0.23 [-0.56, 0.12]	0.89 [0.83, 0.95]	0.12 [0.82, 1.07]	0.75 [0.62, 0.85]	0	3
kxzt	0.96 [0.56, 1.33]	0.64 [0.41, 0.92]	0.00 [-0.32, 0.36]	0.90 [0.76, 0.97]	0.36 [0.96, 1.13]	0.79 [0.63, 0.91]	0	37
eyetm	0.98 [0.69, 1.27]	0.72 [0.50, 0.97]	-0.32 [-0.65, 0.00]	0.91 [0.86, 0.96]	0.00 [0.94, 1.22]	0.78 [0.64, 0.88]	0	7
cm2yq	0.99 [0.44, 1.54]	0.56 [0.31, 0.90]	0.10 [-0.21, 0.50]	0.91 [0.83, 0.98]	0.50 [0.96, 1.25]	0.89 [0.80, 0.96]	0	36
2umai	1.00 [0.46, 1.54]	0.57 [0.33, 0.91]	0.07 [-0.25, 0.46]	0.91 [0.82, 0.98]	0.46 [0.96, 1.26]	0.87 [0.76, 0.95]	0	36
ko8yx	1.01 [0.76, 1.25]	0.78 [0.56, 1.01]	0.35 [0.01, 0.67]	0.91 [0.82, 0.96]	0.67 [0.96, 1.19]	0.78 [0.64, 0.89]	0	26
wuuvc	1.02 [0.51, 1.53]	0.62 [0.38, 0.93]	0.19 [-0.13, 0.58]	0.88 [0.80, 0.96]	0.58 [0.85, 1.19]	0.90 [0.81, 0.96]	0	36
ktpj5	1.02 [0.51, 1.56]	0.61 [0.37, 0.95]	0.17 [-0.16, 0.57]	0.88 [0.80, 0.96]	0.57 [0.87, 1.22]	0.89 [0.80, 0.96]	0	36
z7fhp	1.02 [0.49, 1.55]	0.61 [0.36, 0.94]	0.08 [-0.24, 0.48]	0.90 [0.82, 0.97]	0.48 [0.97, 1.26]	0.88 [0.80, 0.95]	0	28
arcko	1.04 [0.73, 1.32]	0.77 [0.53, 1.02]	0.37 [0.05, 0.72]	0.89 [0.80, 0.94]	0.72 [0.90, 1.14]	0.78 [0.62, 0.90]	0	24
y4wws	1.04 [0.70, 1.33]	0.74 [0.49, 1.00]	-0.31 [-0.66, 0.05]	0.91 [0.85, 0.96]	0.05 [1.02, 1.26]	0.79 [0.68, 0.88]	0	30
wcvnu	1.11 [0.80, 1.39]	0.84 [0.59, 1.11]	0.28 [-0.10, 0.66]	0.89 [0.77, 0.95]	0.66 [0.98, 1.22]	0.73 [0.54, 0.88]	1	27
8toyp	1.13 [0.61, 1.65]	0.70 [0.42, 1.05]	0.13 [-0.25, 0.56]	0.88 [0.81, 0.96]	0.56 [0.98, 1.29]	0.83 [0.72, 0.92]	0	27
qsicn	1.17 [0.30, 1.65]	0.88 [0.23, 1.54]	-0.76 [-1.54, 0.01]	0.91 [0.46, 1.00]	0.01 [0.52, 1.59]	0.80 [0.00, 1.00]	0	2
wexjs	1.30 [0.95, 1.62]	0.98 [0.68, 1.29]	0.27 [-0.17, 0.74]	0.86 [0.74, 0.93]	0.74 [1.00, 1.29]	0.73 [0.55, 0.86]	0	25
v8qph	1.37 [0.92, 1.79]	0.98 [0.66, 1.34]	-0.15 [-0.64, 0.34]	0.84 [0.70, 0.93]	0.34 [0.97, 1.32]	0.70 [0.55, 0.82]	0	6
w420e	1.57 [1.18, 1.94]	1.23 [0.90, 1.58]	0.09 [-0.48, 0.62]	0.85 [0.76, 0.91]	0.62 [1.08, 1.46]	0.72 [0.60, 0.82]	0	19
6tvf8	1.88 [0.87, 2.85]	1.02 [0.54, 1.66]	0.45 [-0.14, 1.18]	0.51 [0.16, 0.87]	1.18 [0.26, 0.89]	0.61 [0.34, 0.82]	0	55
0wfzo	2.89 [1.73, 3.89]	1.88 [1.17, 2.68]	0.76 [-0.15, 1.77]	0.48 [0.21, 0.75]	1.77 [0.60, 1.37]	0.51 [0.30, 0.70]	0	4
t8ewk	3.30 [1.89, 4.39]	1.98 [1.06, 3.00]	1.32 [0.27, 2.49]	0.07 [0.00, 0.45]	2.49 [-0.17, 0.79]	0.28 [-0.03, 0.6]	0	6
z3btx	4.00 [2.30, 5.45]	2.49 [1.47, 3.65]	1.48 [0.26, 2.86]	0.29 [0.04, 0.60]	2.86 [0.31, 1.44]	0.43 [0.19, 0.63]	0	1
758j8	4.52 [2.64, 6.18]	2.95 [1.85, 4.25]	1.85 [0.48, 3.38]	0.24 [0.02, 0.58]	3.38 [0.20, 1.51]	0.34 [0.08, 0.57]	0	2
hgn83	6.38 [4.04, 8.47]	4.11 [2.52, 5.93]	2.13 [0.07, 4.28]	0.08 [0.00, 0.39]	4.28 [-0.18, 1.43]	0.32 [0.07, 0.56]	0	0

**Table S4. Evaluation statistics calculated for all microscopic pK<sub>a</sub> prediction submissions based on microstate pair match for 8 molecules with NMR data.** Methods are represented via their SAMPL6 submission IDs which can be cross referenced with Table 1 for method details. There are eight error metrics reported: the root-mean-squared error (RMSE), mean absolute error (MAE), mean (signed) error (ME), coefficient of determination ( $R^2$ ), linear regression slope (m), Kendall's Rank Correlation Coefficient ( $\tau$ ), unmatched experimental pK<sub>a</sub>s (number of missing pK<sub>a</sub> predictions) and unmatched predicted pK<sub>a</sub>s (number of extra pK<sub>a</sub> predictions between 2 and 12. This table is ranked by increasing RMSE. A CSV version of this table can be found in *SAMPL6-supplementary-documents.tar.gz*.

Submission ID	RMSE	MAE	ME	R <sup>2</sup>	m	Kendall's Tau	Unmatched exp. pK <sub>a</sub> s	Unmatched pred. pK <sub>a</sub> s [2,12]
nb016	0.52 [0.25, 0.71]	0.43 [0.23, 0.65]	-0.09 [-0.45, 0.30]	0.92 [0.05, 0.99]	0.99 [0.14, 1.16]	0.62 [-0.14, 1.00]	0	3
hdijq	0.68 [0.49, 0.83]	0.60 [0.39, 0.80]	0.38 [0.02, 0.70]	0.86 [0.47, 0.98]	0.91 [0.45, 1.26]	0.78 [0.4, 1.00]	0	16
nb011	0.72 [0.35, 1.07]	0.54 [0.28, 0.86]	0.45 [0.14, 0.83]	0.86 [0.18, 0.98]	0.93 [0.50, 1.21]	0.64 [0.26, 0.95]	0	36
ftc8w	0.75 [0.52, 0.96]	0.68 [0.50, 0.89]	-0.31 [-0.68, 0.16]	0.87 [0.02, 0.99]	1.12 [-0.11, 1.39]	0.56 [-0.10, 1.00]	0	35
6vf8	0.76 [0.55, 0.95]	0.68 [0.46, 0.90]	-0.63 [-0.89, -0.35]	0.92 [0.78, 0.99]	0.94 [0.69, 1.41]	0.87 [0.6, 1.00]	0	55
t8ewk	0.96 [0.65, 1.19]	0.81 [0.46, 1.13]	-0.77 [-1.12, -0.38]	0.80 [0.53, 0.96]	0.96 [0.76, 2.26]	0.78 [0.31, 1.00]	1	7
v8qph	0.99 [0.40, 1.52]	0.67 [0.29, 1.17]	-0.09 [-0.75, 0.45]	0.68 [0.11, 0.97]	0.96 [-1.26, 1.16]	0.38 [-0.3, 1.00]	0	6
ccpmw	1.07 [0.78, 1.27]	0.95 [0.60, 1.25]	-0.83 [-1.25, -0.37]	0.74 [0.43, 0.99]	0.95 [0.70, 2.32]	0.89 [0.52, 1.00]	1	8
Oxi4b	1.15 [0.75, 1.50]	0.98 [0.63, 1.36]	-0.30 [-0.94, 0.44]	0.77 [0.02, 0.98]	1.26 [0.09, 2.10]	0.51 [-0.14, 1.00]	0	33
cywyk	1.17 [0.88, 1.41]	1.06 [0.74, 1.35]	-0.47 [-1.09, 0.24]	0.73 [0.02, 0.98]	1.15 [-0.04, 2.00]	0.56 [-0.08, 1.00]	0	36
eyetm	1.17 [0.77, 1.52]	1.00 [0.61, 1.41]	-0.89 [-1.38, -0.38]	0.67 [0.30, 0.94]	0.93 [0.65, 2.59]	0.72 [0.29, 1.00]	1	8
nb008	1.26 [0.74, 1.71]	1.09 [0.63, 1.57]	0.47 [-0.40, 1.32]	0.79 [0.01, 0.99]	1.21 [-0.59, 1.85]	0.52 [-0.2, 1.00]	0	38
y4wws	1.41 [0.95, 1.80]	1.22 [0.78, 1.66]	-0.71 [-1.44, 0.06]	0.87 [0.05, 0.98]	1.55 [0.41, 2.02]	0.56 [-0.11, 1.00]	0	31
ktpj5	1.46 [0.83, 2.10]	1.15 [0.67, 1.77]	0.94 [0.29, 1.68]	0.77 [0.01, 0.98]	1.28 [-0.26, 1.60]	0.42 [-0.27, 0.95]	0	37
wuuvc	1.47 [0.84, 2.09]	1.18 [0.70, 1.77]	0.99 [0.36, 1.68]	0.78 [0.01, 0.98]	1.27 [-0.24, 1.58]	0.47 [-0.20, 1.00]	0	37
xnoe0	1.54 [1.09, 2.00]	1.39 [1.02, 1.83]	0.91 [0.11, 1.64]	0.82 [0.01, 0.98]	1.47 [-0.30, 1.79]	0.42 [-0.27, 0.95]	0	37
qsicn	1.58 [1.44, 1.70]	1.57 [1.44, 1.70]	-1.57 [-1.7, -1.44]	1.00 [0.00, 1.00]	1.06		0	2
epvmk	1.66 [1.20, 2.15]	1.50 [1.07, 1.96]	1.12 [0.31, 1.82]	0.82 [0.02, 0.98]	1.47 [-0.21, 1.8]	0.42 [-0.25, 0.95]	0	37
400ia	1.73 [1.33, 2.17]	1.62 [1.29, 2.02]	1.31 [0.53, 1.93]	0.87 [0.03, 0.99]	1.50 [0.07, 1.84]	0.56 [-0.07, 1.00]	0	36
ko8yx	1.75 [1.08, 2.45]	1.44 [0.87, 2.12]	1.38 [0.74, 2.10]	0.97 [0.88, 1.00]	1.66 [1.46, 2.28]	0.91 [0.69, 1.00]	0	27
Zumai	1.76 [1.21, 2.35]	1.54 [1.04, 2.11]	1.31 [0.55, 2.03]	0.82 [0.02, 0.98]	1.43 [-0.02, 1.77]	0.47 [-0.17, 0.95]	0	37
cm2yq	1.77 [1.22, 2.36]	1.55 [1.06, 2.12]	1.33 [0.57, 2.04]	0.82 [0.02, 0.98]	1.43 [-0.02, 1.76]	0.47 [-0.17, 0.95]	0	37
nxaaw	1.80 [0.84, 2.80]	1.34 [0.80, 2.18]	0.16 [-0.77, 1.41]	0.59 [0.02, 0.97]	1.37 [-0.08, 2.92]	0.6 [-0.05, 1.00]	0	30
wcvnu	1.90 [1.14, 2.64]	1.57 [0.97, 2.27]	1.44 [0.70, 2.24]	0.97 [0.91, 1.00]	1.78 [1.58, 2.48]	0.91 [0.69, 1.00]	0	27
kxzt	2.00 [1.13, 2.73]	1.64 [1.00, 2.39]	1.64 [1.00, 2.39]	0.83 [0.01, 0.98]	1.42 [-0.21, 1.99]	0.56 [-0.10, 1.00]	0	38
wexjs	2.05 [1.18, 2.93]	1.66 [1.01, 2.47]	1.48 [0.63, 2.39]	0.96 [0.55, 0.99]	1.87 [1.54, 2.29]	0.73 [0.20, 1.00]	0	26
z7fhp	2.14 [1.38, 2.87]	1.80 [1.12, 2.58]	1.28 [0.18, 2.34]	0.78 [0.02, 0.98]	1.71 [-0.41, 2.13]	0.42 [-0.25, 0.95]	0	30
gdqeg	2.38 [1.97, 2.71]	2.25 [1.74, 2.68]	-1.61 [-2.46, -0.37]	0.10 [0.00, 0.98]	0.31 [-0.60, 1.63]	0.29 [-0.45, 1.00]	0	53
8toyp	2.63 [1.89, 3.29]	2.34 [1.59, 3.07]	1.78 [0.47, 2.89]	0.82 [0.02, 0.98]	1.94 [-0.06, 2.39]	0.47 [-0.17, 0.95]	0	29
w4z0e	2.63 [1.81, 3.53]	2.34 [1.67, 3.18]	1.74 [0.46, 2.92]	0.98 [0.55, 1.00]	2.28 [1.52, 2.41]	0.73 [0.20, 1.00]	0	20
arcko	2.64 [1.23, 3.78]	2.08 [1.10, 3.24]	1.71 [0.44, 3.10]	0.57 [0.04, 0.95]	1.42 [0.56, 2.93]	0.56 [-0.06, 1.00]	0	28
0wfzo	18.72 [11.21, 25.03]	15.80 [9.9, 22.35]	15.09 [8.28, 22.12]	0.09 [0.01, 0.73]	2.35 [-10.18, 8.12]	0.02 [-0.65, 0.66]	0	12
z3btv	22.60 [15.03, 29.00]	19.70 [12.97, 26.69]	19.70 [12.97, 26.69]	0.09 [0.01, 0.72]	2.35 [-10.00, 8.28]	0.02 [-0.66, 0.66]	0	7
758j8	23.76 [16.33, 30.24]	21.00 [14.26, 28.00]	21.00 [14.26, 28.00]	0.09 [0.01, 0.71]	2.35 [-10.34, 8.12]	0.02 [-0.65, 0.65]	0	8
hgn83	27.91 [20.54, 34.52]	25.60 [18.9, 32.64]	25.60 [18.9, 32.64]	0.09 [0.01, 0.72]	2.35 [-10.21, 8.00]	0.02 [-0.65, 0.65]	0	5

**A****B**

**Figure S3. The value of macroscopic  $pK_a$ s was not a factor affecting prediction error seen in SAMPL6 Challenge according to the analysis with Hungarian matching.** There was not clear trend between  $pK_a$  prediction error and the true  $pK_a$  error. Very high and very low  $pK_a$  values have similar inaccuracy compared to  $pK_a$  values close to 7. **A** Scatter plot of macroscopic  $pK_a$  prediction error calculated with Hungarian matching vs. experimental  $pK_a$  value **B** Box plot of absolute error of macroscopic  $pK_a$  predictions binned into 2  $pK_a$  unit intervals of experimental  $pK_a$ .