

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

Mehtap Isik (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. Gunner⁶, David L. Mobley (ORCID: [0000-0002-1083-5533](#))⁵, John D. Chodera (ORCID: [0000-0003-0542-119X](#))¹

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

*For correspondence:
mehtap.isik@choderlab.org (MI)

17

Abstract

Acid dissociation constant (pK_a) prediction is a prerequisite for predicting many other properties of small molecules such as protein-ligand binding affinity, distribution coefficient ($\log D$), membrane permeability, and solubility due to the necessity of predicting relevant protonation states and the free energy penalty of each state. SAMPL6 pK_a Challenge was the first time that a separate challenge was conducted for evaluating pK_a 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_a prediction issues. The goal of the pK_a challenge was to elucidate the performance of contemporary pK_a 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_a prediction methods. Four widely used pK_a prediction methods that were missing from blind predictions were added as reference methods to challenge analysis. Collecting both microscopic and macroscopic pK_a predictions allowed in-depth evaluation of pK_a prediction performance. This article highlights deficiencies of typical pK_a prediction evaluation approaches when the difference between microscopic and macroscopic pK_a s is ignored and suggests more stringent evaluation criteria for microscopic and macroscopic pK_a predictions guided by the available experimental data. Top-performing submissions for macroscopic pK_a 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_a s for the set of 24 molecules. A large number of submissions had RMSE spanning 1-3 pK_a units. Molecules with sulfur-containing heterocycles, iodo, and bromo groups suffered from less accurate pK_a 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_a predictions which was prominent for charged tautomers. SAMPL6 pK_a Challenge demonstrated the need for improving pK_a prediction methods for drug-like molecules, especially for challenging moieties and multiprotic molecules. The level of pK_a 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.

43 0.1 Keywords

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

46 0.2 Abbreviations

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

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

49 **SEM** Standard error of the mean

50 **RMSE** Root mean squared error

51 **MAE** Mean absolute error

52 τ Kendall's rank correlation coefficient (Tau)

53 **R²** Coefficient of determination (R-Squared)

54 1 Introduction

55 The acid dissociation constant (pK_a) describes the protonation state equilibrium of a molecule given pH. Predicting pK_a is a
56 prerequisite for predicting many other properties of small molecules such as protein-ligand binding affinity, distribution coeffi-
57 cient ($\log D$), membrane permeability, and solubility. Computer-aided drug design efforts include assessing properties of virtual
58 molecules to guide synthesis and prioritization decisions. In such cases an experimental pK_a measurement is not possible.
59 Therefore, accurate computational pK_a prediction methods are required.

60 For a monoprotic weak acid (HA) or base (B) dissociation equilibria shown in Equation 1, the acid dissociation constant is
61 expressed as in Equations 2 or its common negative logarithmic form as in Equation 3. The ratio of ionization states can be
62 calculate with HHenderson-Hasselbalch equations shown in Equation 4.



$$K_a = \frac{[A^-][H^+]}{[HA]} \quad K_a = \frac{[B][H^+]}{[BH^+]} \quad (2)$$

$$pK_a = -\log_{10} K_a \quad (3)$$

$$pH = pK_a + \log_{10} \frac{[A^-]}{[HA]} \quad pH = pK_a + \log_{10} \frac{[B]}{[BH^+]} \quad (4)$$

63 Ionizable sites are found often in drug molecules and influence their pharmaceutical properties including target affinity,
64 ADME/Tox, and formulation properties [1]. Drug molecules with titratable groups can exist in many different charge and proto-
65 nation states based on the pH of the environment. We rely on pK_a values to determine in which charge and protonation states
66 the molecules exists and relative populations of these states. The pH of the human gut ranges between 1-8 and 74% of approved
67 drugs can change ionization states withing this physiological pH range [2] and because of this pK_a values of drug molecules pro-
68 vides essential information about their physicochemical and pharmaceutical properties. A wide distribution of acidic and basic
69 pK_a values, ranging from 0 to 12, have been observed in approved drugs [1, 2].

70 Small molecule pK_a predictions influence computational protein-ligand binding affinities in multiple ways. Errors in pK_a pre-
71 dictions can cause modeling the wrong charge, protonation, and tautomerization states which affect hydrogen bonding oppor-
72 tunities and charge distribution of the ligand. The prediction of the dominant protonation state and relative population of minor
73 states in aqueous medium is dictated by the pK_a values. The relative free energy of different protonation states in the aque-
74 ous state is a function of pK_a and pH, it contributes to the overall protein-ligand affinity in the form of a free energy penalty of
75 reaching higher energy protonation states [3].

76 Drug-like molecules present difficulties for pK_a prediction compared to simple monoprotic molecules. Drug-like molecules
77 are frequently multiprotic, have large conjugated systems, heterocycles, tautomerization. In addition that larger molecules
78 with conformational flexibility can have intramolecular hydrogen bonding which shifts pK_a values. These shifts could be real or

79 modeling artifacts due to collapsed conformations caused by deficiencies in solvation models. Yet predicting pK_a s of drug-like
80 molecules accurately is a prerequisite for computational drug discovery and design.

81 The definition of pK_a diverges into two for multiprotic molecules: macroscopic pK_a and microscopic pK_a [4–6]. Macroscopic
82 pK_a describes the equilibrium dissociation constant between different charged states of the molecule. Each charge state can be
83 composed of multiple tautomers. Macroscopic pK_a is about the deprotonation of the molecule, not a particular titratable group.
84 Microscopic pK_a describes the acid dissociation equilibrium between individual tautomeric states of different charges. We refer
85 to collection of all tautomeric states of different macroscopic states (charge states) as microscopic states. Microscopic pK_a value
86 defined between two microstates captures the deprotonation of a single titratable group with a fixed background protonation
87 state of other titratable groups. In molecules with multiple titratable groups, the protonation state of one group can affect the
88 proton dissociation propensity of another functional group, therefore the same titratable group may have different microscopic
89 pK_a values based on the protonation state of the rest of the molecule. Different experimental methods capture different def-
90 initions of pK_a s as explained in more detail in this prior publication [7]. Most common pK_a measurement techniques such as
91 potentiometric and spectrophotometric methods measure macroscopic pK_a s while NMR measurements can determine micro-
92 scopic pK_a s and microstate populations. Therefore, it is important to pay attention to the source and definition of pK_a values
93 to interpret their meaning correctly. Computational methods can predict both microscopic and macroscopic pK_a s. While micro-
94 scopic pK_a predictions are more informative for determining relevant microstates/tautomers of a molecule and their relative
95 free energies, computing predicted macroscopic pK_a s is useful for direct comparison of methods to more common macroscopic
96 experimental measurements. In this paper, we explore approaches to assess the performance of both macroscopic and micro-
97 scopic pK_a predictions, taking advantage of available experimental data.

98 1.1 Motivation for a blind pK_a challenge

99 SAMPL (Statistical Assessment of the Modeling of Proteins and Ligands) is a series of annual computational prediction chal-
100 lenges for the computational chemistry community. The goal of SAMPL is evaluate to current performance of the models and to
101 bring the attention of quantitative biomolecular modeling field on major issues that limit the accuracy of protein-ligand binding
102 models.

103 SAMPL Challenges that focus on different physical properties so far have assessed intermolecular binding models of various
104 protein-ligand and host-guest systems, solvation models to predict hydration free energies and distribution coefficients. Potan-
105 tial benefits of these challenges are motivating improvement computational methods and revealing unexpected contributors to
106 error by focusing on interesting test systems. SAMPL Challenges have demonstrated the effects of force field accuracy, sampling
107 issues, solvation modeling defects, and tautomer/protonation state predictions on protein-ligand binding predictions.

108 During the SAMPL5 log D Challenge, the performance of cyclohexane-water log D predictions were lower than expected and
109 accuracy suffered when protonation states and tautomers were not taken into account [8, 9]. With the motivation of decon-
110 voluting the different sources of error contributing to the large errors observed in the SAMPL5 log D Challenge, we organized
111 separate of pK_a and log P challenges in SAMPL6 [7, 10, 11]. For this iteration of the SAMPL challenge, we have taken one step
112 back and isolated just the problem of predicting aqueous protonation states.

113 This is the first time a blind pK_a prediction challenge has been fielded as part of SAMPL. In this first iteration of the challenge,
114 we aimed to assess the performance of current pK_a prediction methods for drug-like molecules, investigate potential causes
115 of inaccurate pK_a estimates, and determine how much current level of expected accuracy might impact protein binding affinity
116 predictions. In binding free energy predictions, any error in predicting the free energy of accessing a minor aqueous protonation
117 state of ligand that contributes to the complex formation will directly add to the error in the predicted binding free energy.
118 Similarly for log D predictions, inaccurate prediction aqueous protonation state that contribute partitioning between phases or
119 prediction of relative free energy of these states will be detrimental to the accuracy of transfer free energy predictions.

120 1.2 Approaches to predict small molecule pK_a s

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

122 2 Methods

123 2.1 Design and logistics of the SAMPL6 pK_a Challenge

124 The SAMPL6 pK_a Challenge was conducted as a blind prediction challenge focus on predicting aqueous pK_a value of 24 small
125 molecules that resemble fragments of kinase inhibitors. The compound selection process was described in depth in the prior

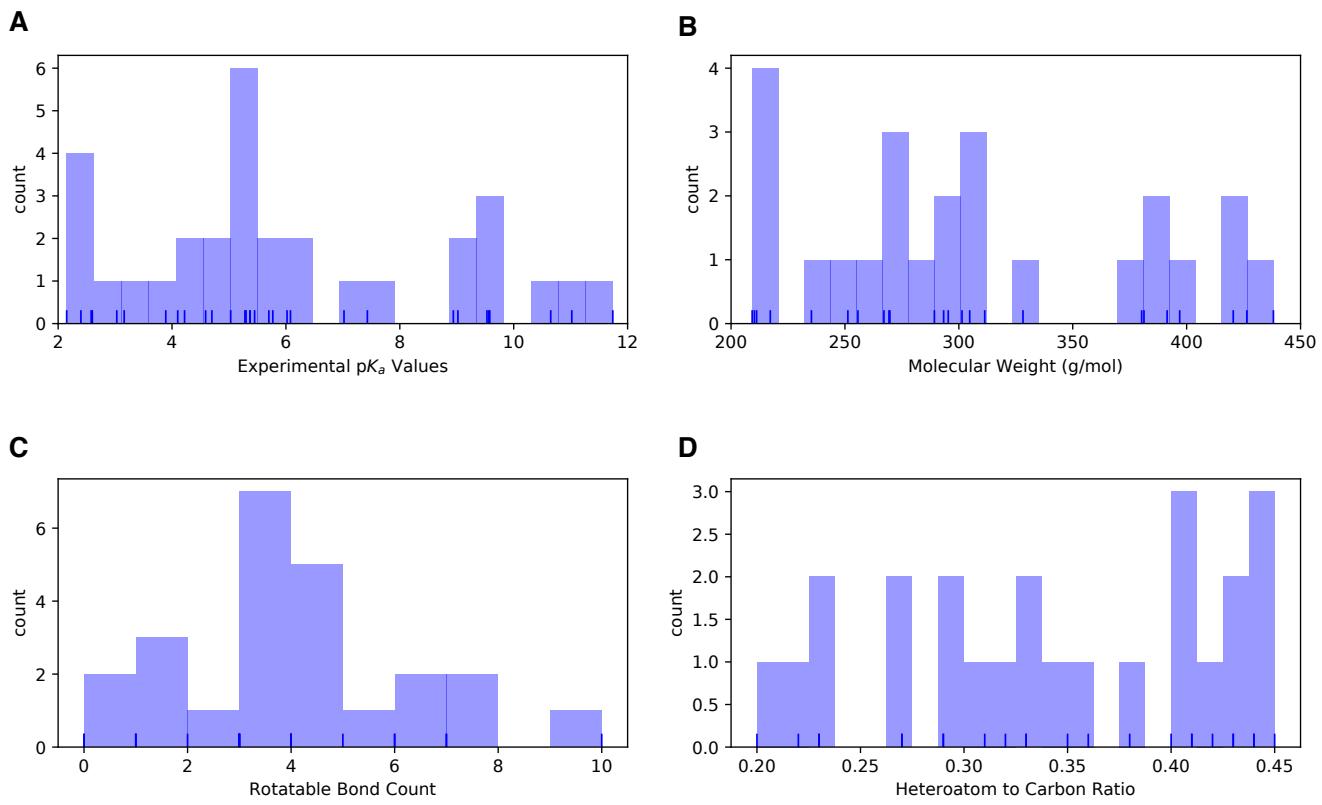


Figure 1. Distribution of molecular properties of 24 compounds in SAMPL6 pK_a Challenge. **A** Histogram of spectrophotometric pK_a measurements collected with Sirius T3 [7]. Overlayed carpet plot indicates the actual values. Five compounds have multiple measured pK_a s in the range of 2-12. **B** Histogram 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.

126 publication reporting SAMPL6 pK_a Challenge experimental data collection [7]. The distribution of molecular weights, experimen-
127 tal pK_a values, number of rotatable bonds, and heteroatom to carbon ratio are depicted in Fig. 1. The challenge molecule set
128 was composed of 17 small molecules with limited flexibility (less than 5 non-terminal rotatable bonds) and 7 molecules with
129 5-10 non-terminal rotatable bonds. The distribution of experimental pK_a values ranged between 2-12 and roughly uniform. 2D
130 representations of all compounds were provided in Fig. 5. Drug-like molecules are often larger and more complex than the ones
131 used in this study, however, aimed for the

132 The dataset composition and details of the pK_a measurement technique, except the identity of the small molecules, were
133 announced about a month before the challenge start time. Experimental macroscopic pK_a measurements were collected with
134 spectrophotometric method of Sirius T3, at room temperature in ionic strength-adjusted water with 0.15 M KCl [7]. The instruc-
135 tions for participation and the identity of the challenge molecules were released at the challenge start date (October 25, 2017).
136 A table of molecule IDs (in the form of SM##) and their canonical isomeric SMILES was provided as input. Blind prediction
137 submissions were accepted until January 22, 2018.

138 Following the conclusion of the blind challenge, the experimental data was made public on January 23, 2018. The SAMPL
139 organizers and participants gathered at the Second Joint D3R/SAMPL Workshop, at UC San Diego, La Jolla, CA on February 22-23,
140 2018 to share results. The workshop aimed to create an opportunity for participants to have discussions, evaluate the results
141 and lessons of the challenge together. The participants reported their results and their own evaluations in the special issue of
142 the Journal of Computer-Aided Molecular Design [12].

143 In this first iteration of pK_a prediction challenge we were not sure what was the best way to capture all necessary informa-
144 tion related to pK_a predictions. Our aim was to directly evaluate macroscopic pK_a predictions comparing them to experimental
145 macroscopic pK_a values and to use collected microscopic pK_a prediction data for more in-depth diagnostics of method perfor-
146 mance. Therefore, we asked participants to submit their predictions in three different submission types:

- 147 • **Type I:** microscopic pK_a values and related microstate pairs
- 148 • **Type II:** fractional microstate populations as a function of pH in 0.1 pH increments
- 149 • **Type III:** macroscopic pK_a values

150 For each submission type, a machine-readable submission file template was specified. For type I submissions, participants
151 were asked to report microstate ID of protonated state, microstate ID of deprotonated state, microscopic pK_a , microscopic
152 pK_a SEM. The reason and method of microstate enumeration is discussed further in Section 2.2 "Enumeration of Microstates".
153 The SEM captures the statistical uncertainty of the predicted method. Microstate IDs were preassigned identifiers for each mi-
154 crostates in the form of SM##_micro##. For type II submission, submission format included a table that started with microstate
155 ID and consecutive columns reporting natural logarithm of fractional microstate population values of each predicted microstate
156 for 0.1 pH increments between pH 2 and 12. For type III submissions participants were asked to report molecule ID, macroscopic
157 pK_a , macroscopic pK_a SEM. It was mandatory to submit predictions for all fields for each prediction, but it was not mandatory to
158 submit predictions for all the molecules or all the submission types. Although we have accepted submissions with partial sets of
159 molecules, it would have been a better choice to require predictions for all the molecules for better comparison of method per-
160 formance. The submission files also included fields for naming the method, listing the software utilized, and a free text method
161 section for the detailed documentation of each method.

162 Participants were allowed to submit predictions with multiple methods as long as they create separate submissions files.
163 Anonymous participation to the challenge was allowed, however all participant opted to make their submissions public. All
164 blind submissions were assigned a 5-digit alphanumeric submission ID, which will be used throughout this paper. These sub-
165 mission IDs were also reported in the evaluation papers of participants and allow cross-referencing. Submission IDs, participant
166 provided method names, and method categories are presented in Table 1. There were many instances that multiple types of
167 submissions of the same method were provided by participants as challenge instructions requested. Although each prediction
168 set was assigned a separate submission ID we have matched the submissions that originated from the same method according
169 to the reports of the participant. Submission ID for both macroscopic (type III) and microscopic (type I) pK_a predictions of each
170 method (when exists) are shown in Table 1.

171 2.2 Enumeration of microstates

172 To capture both the pK_a value and titration position of microscopic pK_a predictions, we needed microscopic pK_a predictions to
173 be reported together with the pair of deprotonated and protonated microstates that describes the transition. String represen-
174 tations of molecules such as canonical SMILES with explicit hydrogens can be written, however, there can be inconsistencies

175 between the interpretation of canonical SMILES written by different softwares and algorithms. In order to avoid complications
176 while reading microstate structure files from different sources, we have decided that the safest route was pre-enumerating all
177 possible microstates of challenge compounds, assigning the microstates IDs to each in the form of SM##_micro##, and require
178 participants to report microstate pairs using the provided microstates IDs.

179 We enumerated an initial list of microstates with Epik and OpenEye QUACPAC and took the union of results. Microstates with
180 Epik were generated using Schrodinger Suite v2016-4, and running Epik to enumerate all tautomers within 20 p K_a units of pH 7.
181 For enumerating microstates with OpenEye QUACPAC, we had to first enumerate formal charges and for each charge enumerate
182 all possible tautomers using the settings of maximum tautomer count 200, level 5, and carbonyl hybridization False. Then we
183 created an union of all enumerated states written as canonical isomeric SMILES. Even though resonance structures correspond
184 to different canonical isomeric SMILES they are not different microstates, therefore it was necessary to remove resonance struc-
185 tures that were replicates of the same tautomer. To detect resonance structures we converted canonical isomeric SMILES to
186 InChI hashes with explicit and fixed hydrogen layer. Structures that describe the same tautomer but different resonance states
187 lead to explicit hydrogen InChI hashes that are identical allowing replicates to be removed. The Jupyter Notebook used for the
188 enumeration of microstates is provided in supplementary documents. Because resonance and geometric isomerism should be
189 ignored when matching predicted structures microstate IDs (except SM20 which should be modelled as E-isomer), we provided
190 microstate ID tables with canonical SMILES and 2D-depictions.

191 Despite pooling together enumerated charge states and tautomers with Epik and OpenEye QUACPAC to our surprise the
192 microstate lists were still incomplete. A better algorithm that can enumerate all possible microstates would be very beneficial.
193 In SAMPL6 Challenge participants came up with new microstates that were not present in the initial list that we provided. Based
194 on participant requests we iteratively had to update the list of microstates and assign new microstate IDs. Every time we received
195 a request, we shared the updated microstate ID lists with all the challenge participants.

196 A working p K_a microstate definition for this challenge was provided in challenge instructions for clarity. Physically meaningful
197 microscopic p K_a s are defined between microstate pairs that can interconvert by single protonation/deprotonation event of only
198 one titrable group. So, microstate pairs should have total charge difference of |1| and only one heavy atom that differs in
199 the number of bound hydrogens, regardless of resonance state or geometric isomerism. All geometric isomer and resonance
200 structure pairs that have the same number of hydrogens bound to equivalent heavy atoms are related to the same microstate.
201 Pairs of resonance structures and geometric isomers (cis/trans, stereo) won't be considered as different microstates, as long as
202 there is no change in the number of hydrogens bound to each heavy atom in these structures. Since we wanted to participants
203 to report only microscopic p K_a s that describe single deprotonation events (in contrast to transitions between microstates
204 that are different in terms of two or more titratable protons), we have also provided a pre-enumerated list of allowed microstate
205 pairs.

206 Provided microstate ID and microstate pair lists were intended to be used for reporting microstate IDs and to aid parsing
207 of submissions. The enumerated lists of microstates were not created with the intent to guide computational predictions. This
208 was clearly stated in the challenge instructions. However, we noticed that some participants still used the microstate lists as
209 an input for their p K_a predictions as we received complaints from participants that due to our updates to microstate lists they
210 needed to repeat their calculations. This would not have been an issue, if participants used p K_a prediction protocols that did
211 not rely on an external pre-enumerated list of microstates as an input. None of the participants have reported this dependency
212 in their method descriptions explicitly, therefore we can not identify which submissions have used the enumerated microstate
213 lists as input and which ones has followed the instructions.

214 2.3 Evaluation approaches

215 Since the experimental data for the challenge was mainly composed of macroscopic p K_a values of both monoprotic and multi-
216 protic compounds, evaluation of macroscopic and microscopic p K_a predictions was not straightforward. For only a subset of 8
217 molecules, dominant microstate sequence could be inferred from NMR. For the rest of the molecules the only experimental in-
218 formation available was the macroscopic p K_a value, while experimental data did not provide any information on which group(s)
219 are being titrated, microscopic p K_a values, identity of associated macrostates (which charge) or microstates (which tautomers). In
220 this comparative performance evaluation of we let the experimental data lead the challenge analysis towards various evaluation
221 routes.

222 2.3.1 Matching algorithms for pairing predicted and experimental pK_a s

223 Macroscopic pK_a predictions can be calculated from microscopic pK_a s for direct comparison to experimental macroscopic pK_a
224 values, although there is still a remaining issue. How to match predicted macroscopic pK_a s to experimental macroscopic pK_a s
225 when there could multiple numbers of each reported for each molecule? Experimental data in this case did not provide any
226 information that would indicate the titration site, the overall charge or the tautomer composition of macrostate pairs that are
227 associated with each measured macroscopic pK_a that can guide the matching.

228 For evaluating predictions taking the experimental data as reference Fraczkiewicz et al. delineated recommendations for fair
229 comparative analysis of computational pK_a predictions [13]. In the absence any experimental information that would aid the
230 match, experimental and computational pK_a s should be matched preserving the order of pK_a values and minimizing sum of
231 absolute errors.

232 We picked Hungarian matching algorithm [14, 15] to assign experimental and predicted macroscopic pK_a s with squared error
233 cost function as suggested by Kiril Lanevskij. The algorithm is available in SciPy package (`scipy.optimize.linear_sum_assignment`) [16].
234 This matching algorithm provides optimum global assignment that minimizes linear sum of squared errors of all pairwise
235 matches. The reason to select squared error cost function instead of absolute error cost function is to avoid misordered matches,
236 For instance, for a molecule with experimental pK_a values of 4 and 6, and predicted pK_a s of 7 and 8, Hungarian matching with
237 absolute error cost function would match 6 to 7 and 4 to 9. Hungarian matching with squared error cost would match 4 to 7
238 and 6 to 9, preserving the increasing pK_a value order between experimental and predicted values. A weakness of this approach
239 would be failing to match experimental value of 6 to predicted value of 7, if that was the correct match based on underlying
240 macrostates. But underlying pair of states were unknown to us both because experimental data of the challenge did not con-
241 tain information about what charge states the transitions were happening between and also because we have not collected the
242 pair of macrostates associated with each pK_a predictions in submissions. There is no perfect solution to numerical pK_a assign-
243 ment problem, but we tried to determine the most fair way to penalize predictions based on their numerical deviation from the
244 experimental values.

245 For the analysis of microscopic pK_a predictions we adopted a different matching approach. Only for the 8 molecules, we uti-
246 lized the dominant microstate sequence inferred from NMR experiments to match computational predictions and experimental
247 pK_a s. We will refer to this assignment method as microstate matching, where experimental pK_a value is matched to the com-
248 putational microscopic pK_a value which was reported for the dominant microstate pair observed for each transition. We have
249 compared the results of Hungarian matching and microstate matching.

250 2.3.2 Statistical metrics for submission performance

251 A variety of accuracy and correlation statistics were considered for analyzing and comparing performance of predictions meth-
252 ods submitted to the SAMPL6 pK_a Challenge. Calculated performance statistics of predictions were provided to participants
253 before the workshop. Details of the analysis and scripts are maintained on the SAMPL6 Github Repository (described in Section
254 5).

255 There are six error metrics reported for the numerical error of the pK_a values: the root-mean-squared error (RMSE), mean ab-
256 solute error (MAE), mean error (ME), coefficient of determination (R^2), linear regression slope (m), and Kendall's Rank Correlation
257 Coefficient (τ). Uncertainty in each performance statistic was calculated as 95% confidence intervals estimated by bootstrapping
258 over predictions with 10000 bootstrap samples. Calculated errors statistics of all methods can be found in Table S2 for macro-
259 scopic pK_a predictions and Tables S4 and S4 for microscopic pK_a predictions.

260 In addition to the numerical error aspect of the pK_a values, we have also evaluated predictions in terms of their ability to cap-
261 ture the correct macrostates (ionization states) and microstates (tautomers of each ionization state) to the extend possible from
262 the available experimental data. For macroscopic pK_a s experiments did not provide any evidence of the identity of the ionization
263 states. However, the number of ionization states indicates the number of macroscopic pK_a s that exists between experimental
264 range of 2.0-12.0. For instance, SM14 has two experimental pK_a s and therefore 3 different charge states were observed between
265 the pH range of 2.0-12.0. If a prediction reported 4 macroscopic pK_a s, it is clear that this method predicted an extra ionization
266 state. With this perspective we reported the number of unmatched experimental pK_a s (the number of missing pK_a predictions,
267 i.e. missing ionization states) and the number of unmatched predicted pK_a s (the number of extra pK_a predictions, i.e. extra
268 ionization states) after Hungarian matching. The later count was restricted to only predictions with pK_a values between 2 and
269 12, because that was the range of the experimental method. Errors in extra or missing pK_a prediction errors highlight failure to
270 predict the correct number of ionization states within a pH range.

271 For the evaluation of microscopic pK_a predictions, taking advantage of the available dominant microstate sequence data for

272 a subset of 8 compounds, we calculated the dominant microstate prediction accuracy. Dominant microstate prediction accuracy
273 is the ratio of correct dominant tautomer predictions for each charge state divided by, calculated over all ionization states of each
274 molecule. In order to extract the sequence of dominant microstates from the microscopic pK_a predictions sets, we calculated
275 the relative free energy of microstates selecting a neutral tautomer and pH 0 as reference following the Equation 5. Calculation
276 of relative free energy of microstates was explained in more detail in a previous publication [17].

277 Relative free energy of state with respect to reference state B at pH 0.0 (arbitrary pH value selected as reference) can be
278 calculated as follows:

$$\Delta G_{AB} = \Delta m_{AB} RT \ln 10 (pH - pK_a) \quad (5)$$

279 Δm_{AB} is equal to the number protons in state A minus state B. R and T indicate molar gas constant and temperature, respectively.
280 By calculating relative free energies of all predicted microstates with respect to the same reference state and pH, we were able
281 to determine the sequence of predicted dominant microstates. The dominant tautomer of each charge state was determined
282 as the the microstate with the lowest free energy in the subset of predicted microstates of each ionization state. This approach
283 is feasible because the relative free energy of tautomers of the same ionization state is independent of pH and therefore the
284 choice of reference pH is arbitrary.

285 We created a shortlist of top-performing methods for macroscopic pK_a predictions based on the following criteria: ranking
286 in the top 10 consistently according to two error (RMSE, MAE) and two correlation metrics (R-Squared, and Kendall's Tau), and
287 also havin a combined count of less than 8 missing or extra macroscopic pK_a s for the entire molecule set (a third of the number
288 of compounds). These methods are presented in Table 2.

289 In addition to comparing the performance comparison of methods, we also wanted to compare pK_a prediction performance
290 on the level of molecules to determine pK_a s of which molecules in the challenge set were harder to predict considering all the
291 methods in the challenge. For this purpose, we plotted prediction error distributions of each molecule considering all prediction
292 methods. We also calculated MAE for each molecule's over all predictions as well as for predictions from each method category.

293 2.4 Reference calculations

294 Including null model as helpful in comparative performance analysis of predictive methods to establish what the performance
295 statistics look like for a baseline method for the specific dataset. Null models or null predictions employ a simple prediction
296 model which is not expected to be particularly successful, but it is useful for providing a simple point of comparison for more
297 sophisticated methods. The expectation is for more sophisticated or costly prediction methods to outperform the predictions
298 from a null model, otherwise the simpler null model would be preferable. In SAMPL6 pK_a Challenge there were two blind submissions
299 that database lookup methods that were suitable to be considered as null predictions. These methods, with submission
300 IDs 5nm4j and 5nm4j both used OpenEye pKa-Porspector database to find the most similar molecule to query molecule and
301 report its pK_a as predicted value.

302 We have also included additional reference calculations in the comparative analysis to provide more perspective. The meth-
303 ods we chose to include as reference calculations were missing from the blind predictions sets although they are widely used
304 methods by academia and industry. representing different methodological approaches: Schrodinger/Epik (nb??), Schrodinger/
305 Jaguar (nb??), Chemaxon/Chemicalize(nb??), and Molecular Discovery/MoKa (nb??). Epik and Jaguar pK_a predictions were col-
306 lected by Bas Rustenburg, Chemicalize predictions by Mehtap Isik, and MoKa predictions by Thomas Fox, after the challenge
307 deadline avoiding any alterations to the respective standard procedures of the methods and guidance of the experimental
308 date.. Reference calculations were not formally blind, as experimental data of the challenge has been made publically available
309 before their collection.

310 All figures and statistics tables in this manuscript include reference calculations. As the reference calculations were not formal
311 submissions, these were omitted from formal ranking in the challenge, but we present plots in this article which show them for
312 easy comparison. These are labeled with submission IDs of the form REF## to allow easy recognition of non-blind reference
313 calculations.

314 3 Results and Discussion

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

317 Submissions spanning different method categories were made to the SAMPL6 pK_a Challenge: database lookup (DL), linear
318 free energy relationship (LFER), quantitative structure property relationship (QSPR), machine learning (ML), quantum mechanics

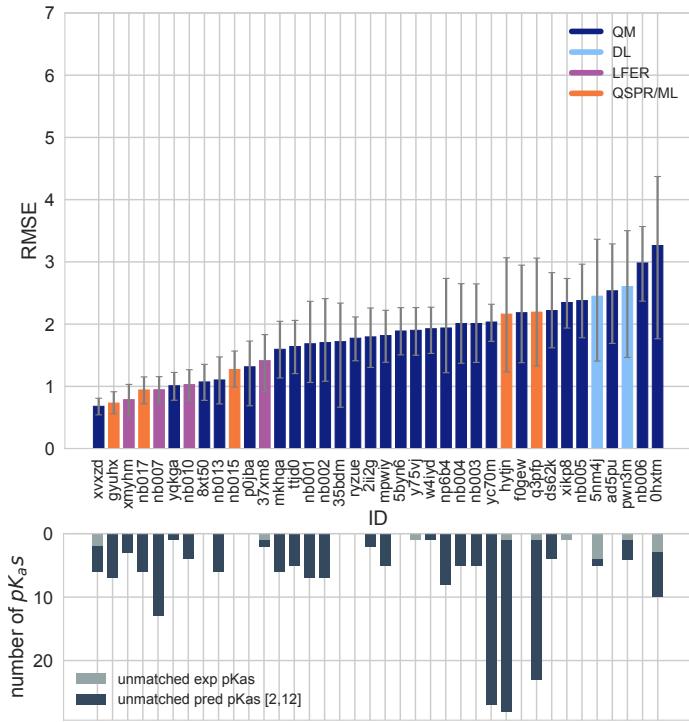


Figure 2. RMSE and unmatched pK_a counts vs. submission ID plots for macroscopic pK_a predictions based on Hungarian matching. Methods are indicated by submission IDs. RMSE is shown with error bars denoting 95% confidence intervals obtained by bootstrapping over challenge molecules. Lower bar plots show 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.

318 (QM) models with and without linear empirical correction (LEC), and combined quantum mechanics and molecular mechanics
319 (QM+MM). Unique submission IDs were assigned to each submission. Table 1 matches method names with submission IDs.
320 Unique IDs were also assigned when multiple submissions exists for different submission types of the same method such as
321 microscopic pK_a (type I) and macroscopic pK_a (type III).

Integral equation-based approaches (e.g. EC-RISM) were also evaluated under the Physical (QM) category. Table 1 indicates the final category assignments in the "Category" column.

3.1 Analysis of macroscopic pK_a predictions (Type III)

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

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

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

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

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.

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	[18]
DL	OpenEye pKa-Prospector 1.0.0.3 with Analog Search ion identification algorithm		pwn3m	Null	[18]
LFER	ACD/pKa GALAS (ACD/Percepta Kernel v1.6)	v8qph	37xm8	Blind	[19]
LFER	ACD/pKa Classic (ACD/Percepta Kernel, v1.6)		xmyhm	Blind	[20]
LFER	Epik Scan (Schrodinger v2017-4)		nb007	Reference	[21]
LFER	Epik Microscopic (Schrodinger v2017-4)	nb008	nb010	Reference	[21]
QSPR/ML	OpenEye Gaussian Process	6tvf8	hytjn	Blind	[9]
QSPR/ML	OpenEye Gaussian Process Resampled		q3pfp	Blind	[9]
QSPR/ML	S+pkA (ADMET Predictor v8.5, Simulations Plus)	hdiyq	gyuhx	Blind	[22]
QSPR/ML	Chemcalize v18.23 (ChemAxon MarvinSketch v18.23)		nb015	Reference	[23]
QSPR/ML	MoKa v3.1.3	nb016	nb017	Reference	[24, 25]
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	k08yx	ryzue	Blind	[26]
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	[26]
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	[26]
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	arcko	w4iyd	Blind	[26]
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	[26]
QM + LEC	Jaguar (Schrodinger v2017-4)	nb011	nb013	Reference	[27]
QM + LEC	CPCM/B3LYP/6-311+G(d,p) and global fitting	y4wws	35bdm	Blind	[28]
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	[28]
QM + LEC	EC-RISM/MP2/6-311+G(d,p)-P3NI-q-noThiols-2par	kxzt	ds62k	Blind	[29]
QM + LEC	EC-RISM/MP2/cc-pVTZ-P2-q-noThiols-2par	ftc8w	2ii2g	Blind	[29]
QM + LEC	EC-RISM/MP2/6-311+G(d,p)-P2-phi-all-2par	ktpj5	nb001	Blind*	[29]
QM + LEC	EC-RISM/MP2/6-311+G(d,p)-P2-phi-noThiols-2par	wuuvc	nb002	Blind*	[29]
QM + LEC	EC-RISM/MP2/6-311+G(d,p)-P3NI-phi-all-2par	2umai	nb003	Blind*	[29]
QM + LEC	EC-RISM/MP2/6-311+G(d,p)-P3NI-phi-noThiols-2par	cm2yq	nb004	Blind*	[29]
QM + LEC	EC-RISM/MP2/6-311+G(d,p)-P2-phi-all-1par	z7fhp	nb005	Blind*	[29]
QM + LEC	EC-RISM/MP2/6-311+G(d,p)-P3NI-phi-all-1par	8toyp	nb006	Blind*	[29]
QM + LEC	EC-RISM/MP2/cc-pVTZ-P2-phi-noThiols-2par	epvmk	tjjd0	Blind	[29]
QM + LEC	EC-RISM/MP2/cc-pVTZ-P2-phi-all-2par	xnoe0	rnkhqa	Blind	[29]
QM + LEC	EC-RISM/MP2/cc-pVTZ-P3NI-phi-noThiols-2par	4o0ia	mpwiy	Blind	[29]
QM + LEC	EC-RISM/B3LYP/6-311+G(d,p)-P3NI-q-noThiols-2par	nxaaw	ad5pu	Blind	[29]
QM + LEC	EC-RISM/B3LYP/6-311+G(d,p)-P3NI-phi-noThiols-2par	0xi4b	f0gew	Blind	[29]
QM + LEC	EC-RISM/B3LYP/6-311+G(d,p)-P2-phi-noThiols-2par	cywyk	np6b4	Blind	[29]
QM + LEC	PCM/B3LYP/6-311+G(d,p)	gdqeg	yc70m	Blind	[29]
QM + LEC	COSMOtherm_FINE17 (COSMOtherm C30_1701, BP/TZVPD/FINE//BP/TZVP/COSMO)	t8ewk	0hxtm	Blind	[30, 31]
QM + LEC	DSD-BLYP-D3(BJ)/def2-TZVPD//PBEh-3c[DCOSMO-RS] + RRHO(GFN-XTB[GBSA]) + Gsolv(COSMO-RS[TZVPD]) and linear fit		xvxd	Blind	[32]
QM + LEC	ReSCoSS conformations // DSD-BLYP-D3 reranking // COSMOtherm pKa: DSD-BLYP-D3(BJ)/def2-TZVPD// PBE-D3(BJ)/def2-TZVPD//PBEh-3c[DCOSMO-RS] + 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	[32]
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	[32]
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	[33]
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	z3btv		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 $nb\#\#\#$.

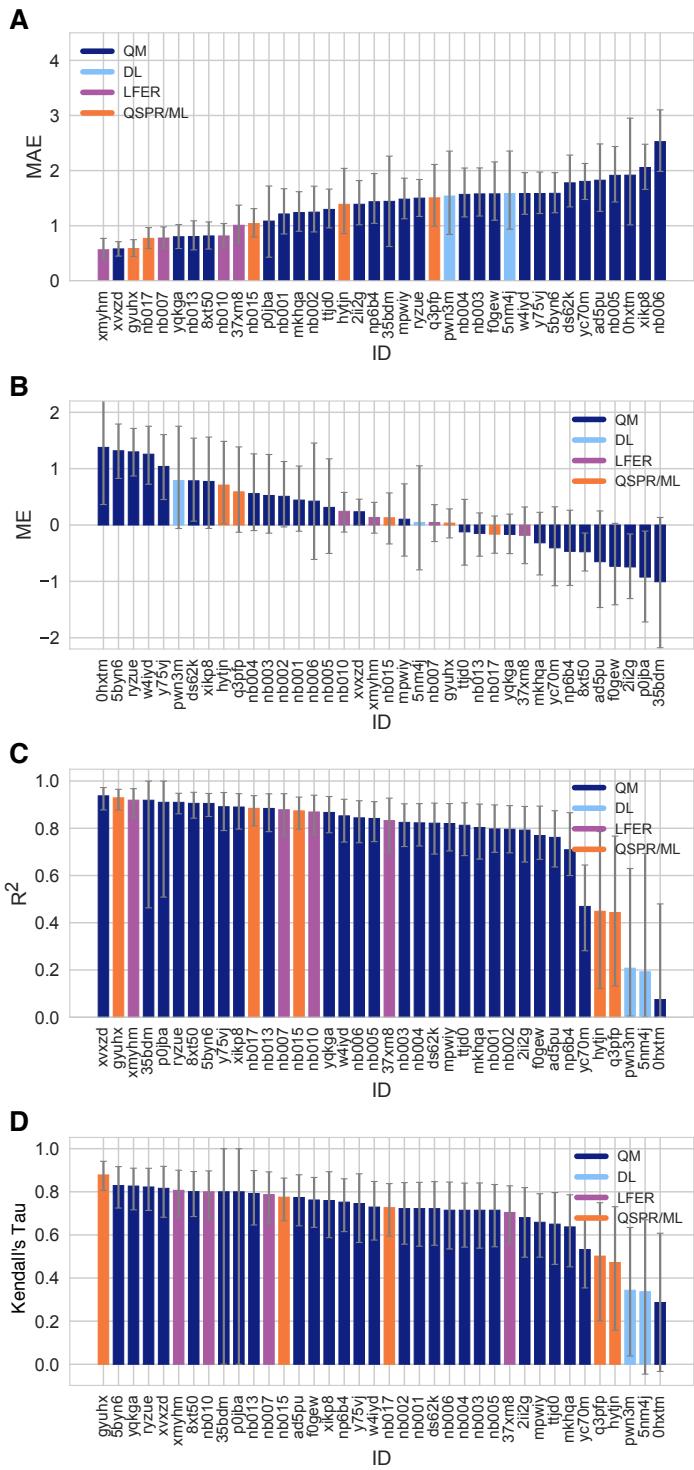


Figure 3. Additional performance statistics for macroscopic pKa predictions based on Hungarian matching. Methods are indicated by submission IDs. Mean absolute error (MAE), mean error (ME), Pearson's R², and Kendall's Rank Correlation Coefficient Tau (τ) are shown, with error bars denoting 95% confidence 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.

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, R², 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

335 3.1.1 Consistently well performing methods for macroscopic pK_a prediction

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

337 3.1.2 Which chemicals are harder to predict?

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

340 Prediction performance of individual molecules

341 Which chemical structures make pKa predictions more difficult?

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

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

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

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

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

350 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.

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

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

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

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

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

358 3.2 Analysis of microscopic pK_a predictions using microstates determined by NMR (8 molecules)

359 3.2.1 Comparing microscopic pKa predictions directly to macroscopic experimental pKa values with numerical
360 matching leads to underestimation of errors

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

363 Reference Figure S4 For most methods the microstate pair of Hungarian predicted pKa does not match experimentally de-
364 termined microstate pair.

365 Discussion of matching experimental and predicted values

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

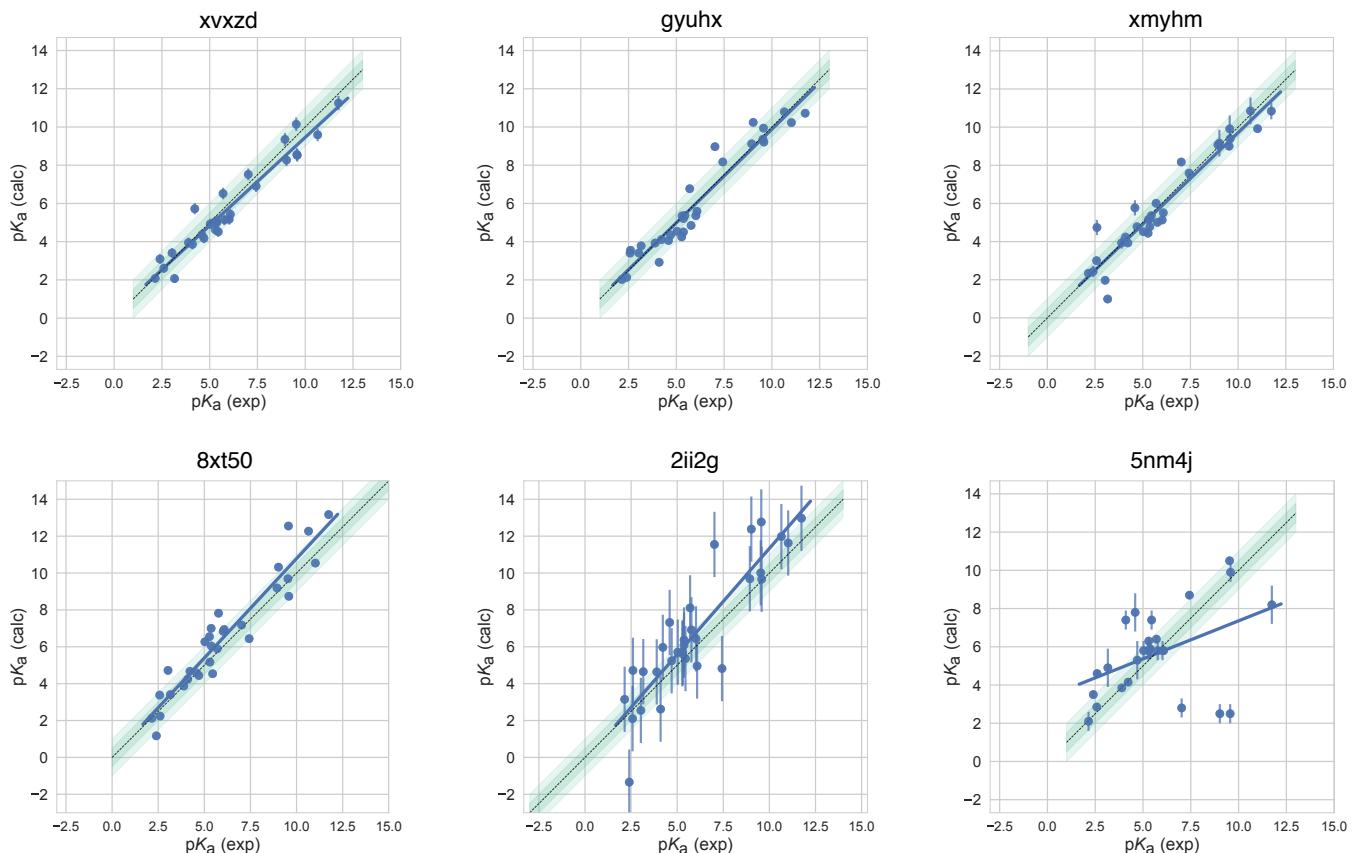


Figure 4. Predicted vs. experimental value correlation plots of 4 consistently well-performing methods, a representative method with average performance (2ii2g), and the null method (5nm4j). Dark and light green shaded areas indicate 0.5 and 1.0 units of error. Error bars indicate standard error of the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

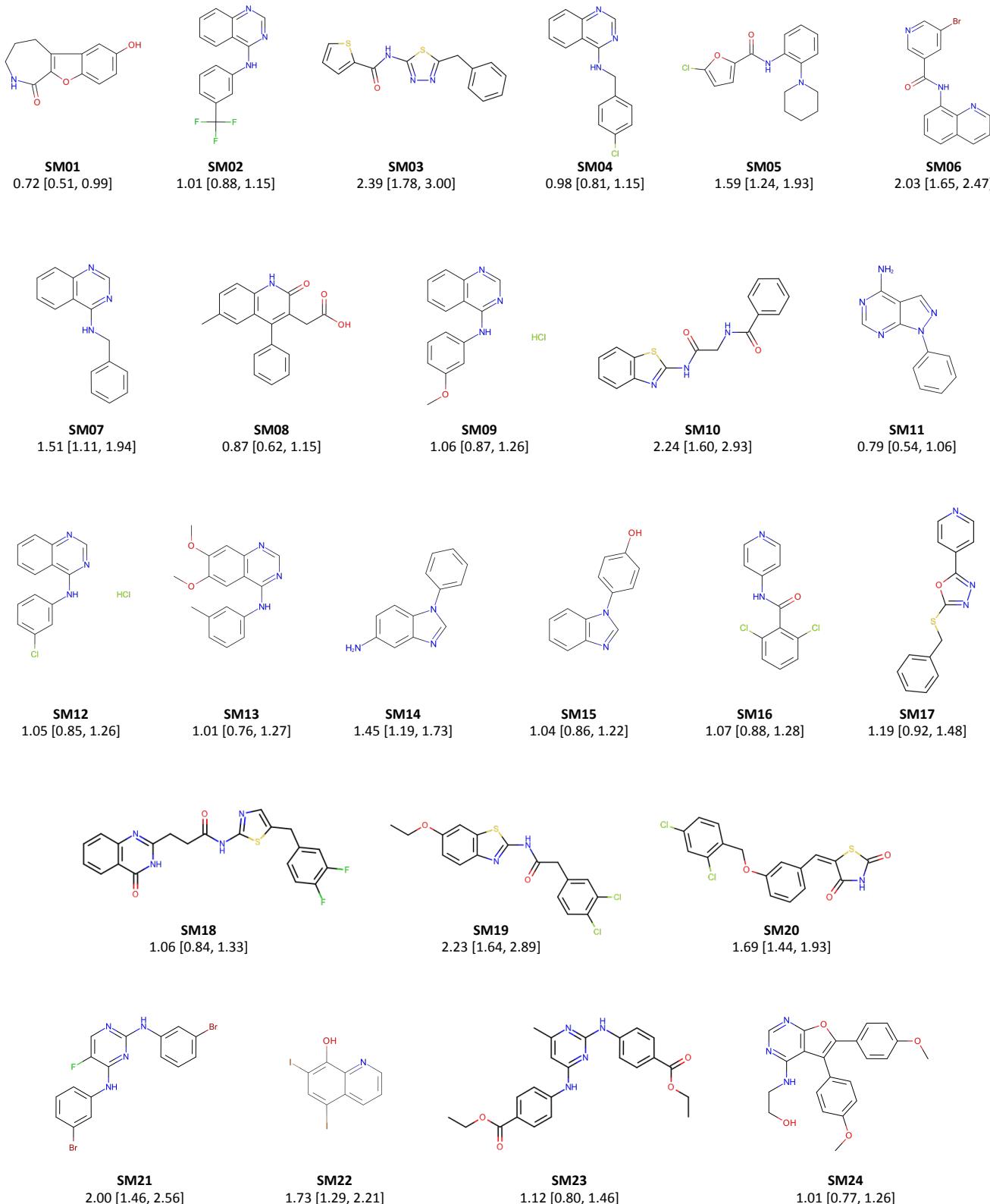
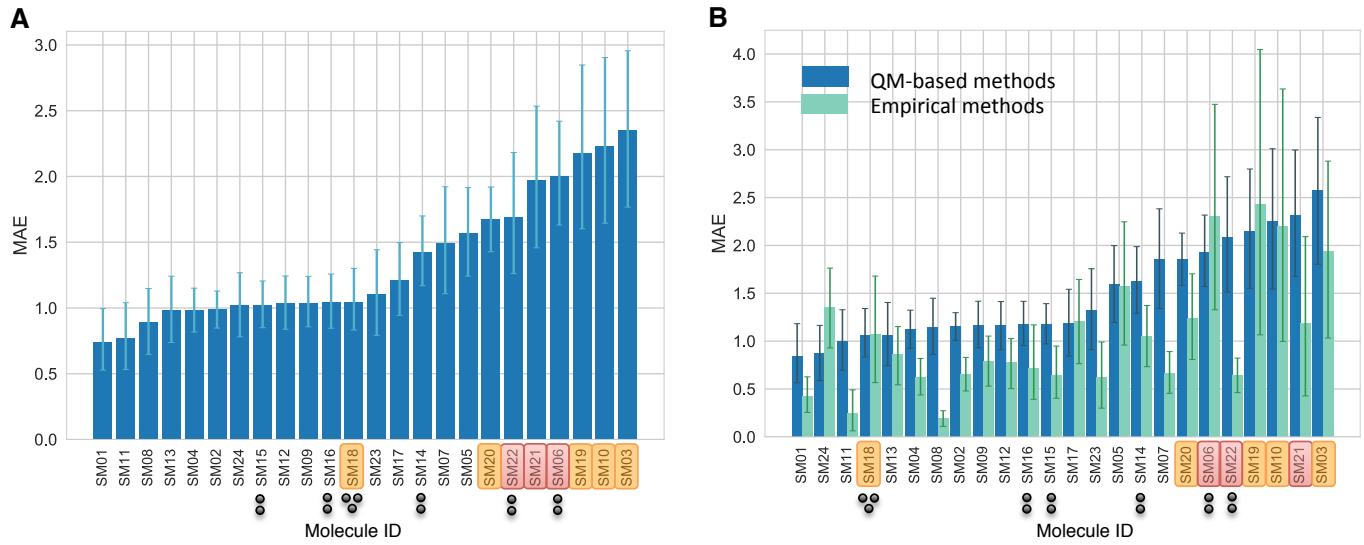
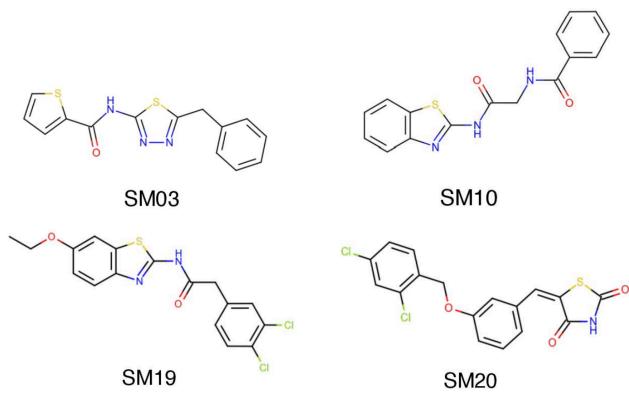


Figure 5. Molecules of SAMPL6 Challenge with MAE calculated for all macroscopic pK_a predictions. MAE calculated considering all prediction methods indicate which molecules had the lowest prediction accuracy in SAMPL6 Challenge. MAE values calculated for each molecule include all the matched pK_a values, 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

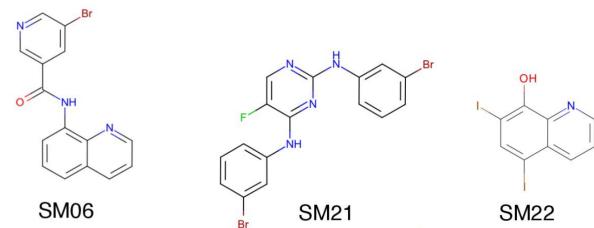


Figure 6. Average prediction accuracy calculated over all prediction methods was lower for molecules with sulfur-containing heterocycles, bromo, and iodo groups. (A) MAE calculated for each molecule as an average of all methods. (B) MAE of each molecule broken out by method category. QM-based methods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

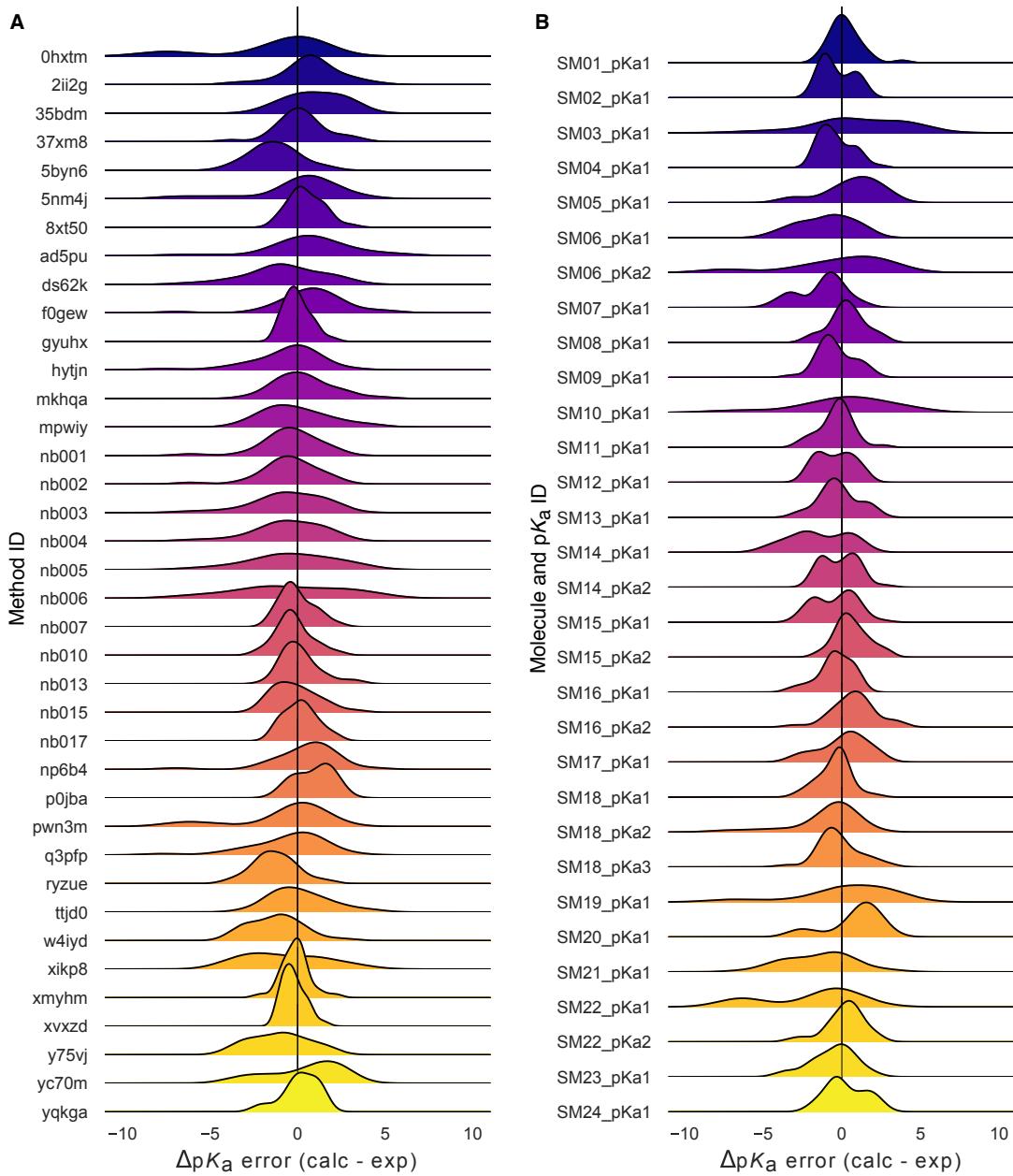
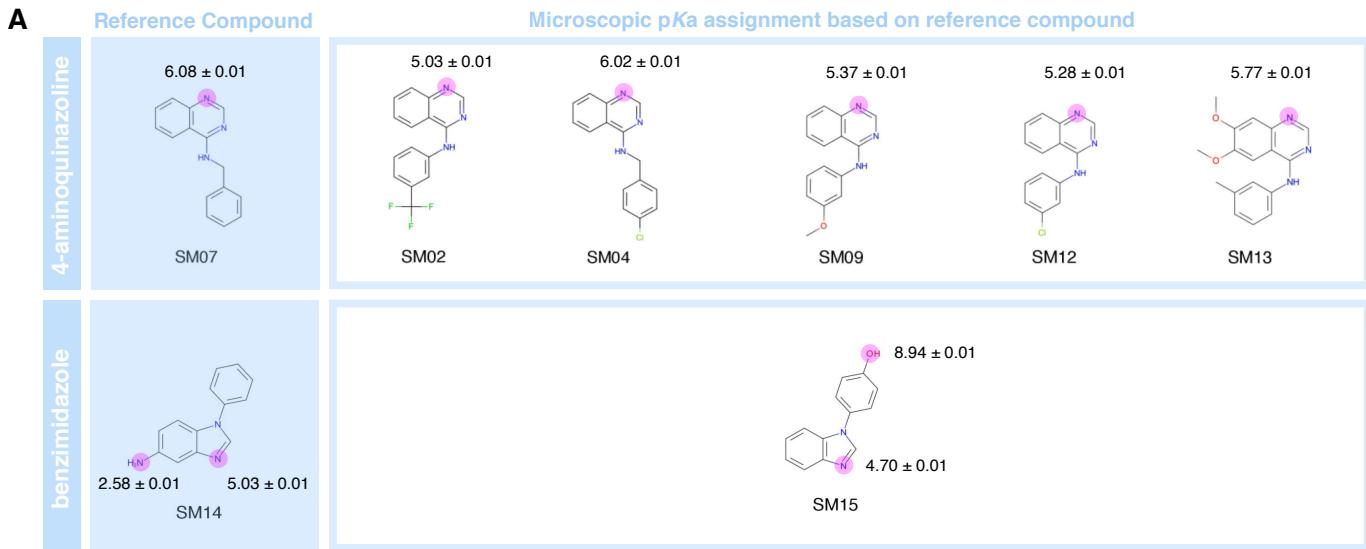
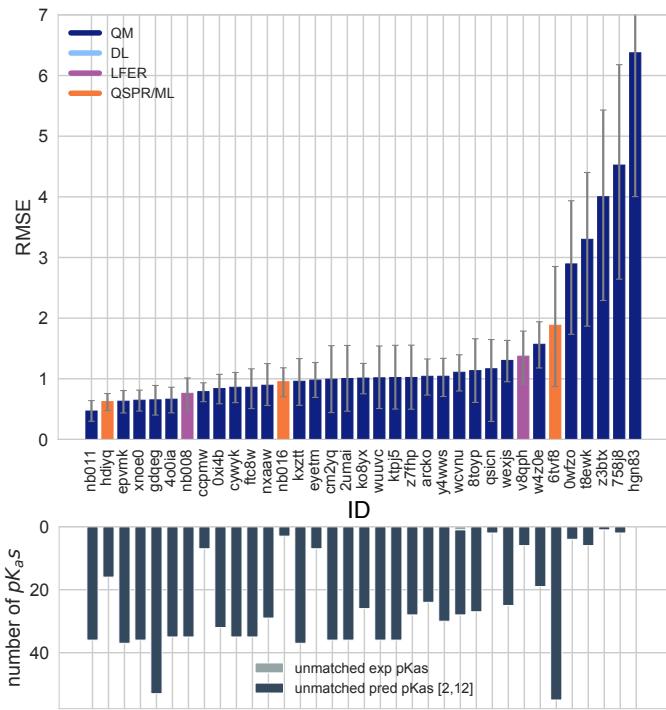


Figure 7. Macroscopic pKa prediction error distribution plots show how prediction accuracy varies across methods and individual molecules. (A) pKa prediction error distribution for each submission for all molecules according to Hungarian matching. (B) Error distribution for each SAMPL6 molecule for all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



B Hungarian matching



C Microstate-based matching

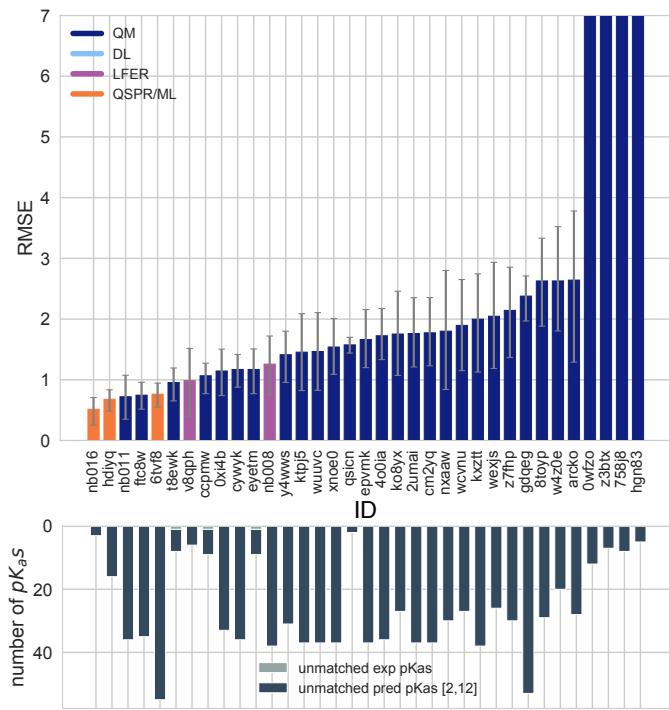


Figure 8. NMR determination of dominant microstates allowed in depth evaluation of microscopic pK_a predictions of 8 compounds.

A Dominant microstate sequence of two compounds (SM07 and SM14) were determined by NMR [7]. 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.

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

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

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

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

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

378 3.2.2 Accuracy of predicted pKa values when microstate matching is used

379 Assessment of individual methods by each of our analysis methods

380 Performance comparison of different methods, grouped by methods class

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

382 3.2.3 Dominant microstate prediction accuracy of methods

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

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

387 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)

388 3.2.4 Which molecules caused lower dominant microstate prediction accuracy?

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

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

391 3.3 Analyzing microscopic pKa prediction from the perspective of thermodynamics

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

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

396 3.3.1 Cycle closure error

397 - Introduce linear protonation state free energy diagram [Cite Gunner et al 2019 paper] FIGURE: linear plot of free energy vs pH

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

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

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

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

404 3.4 How would pKa errors affect protein-ligand binding affinity predictions?

405 Illustrate the ways in which the pKa errors can influence prediction errors for binding affinities

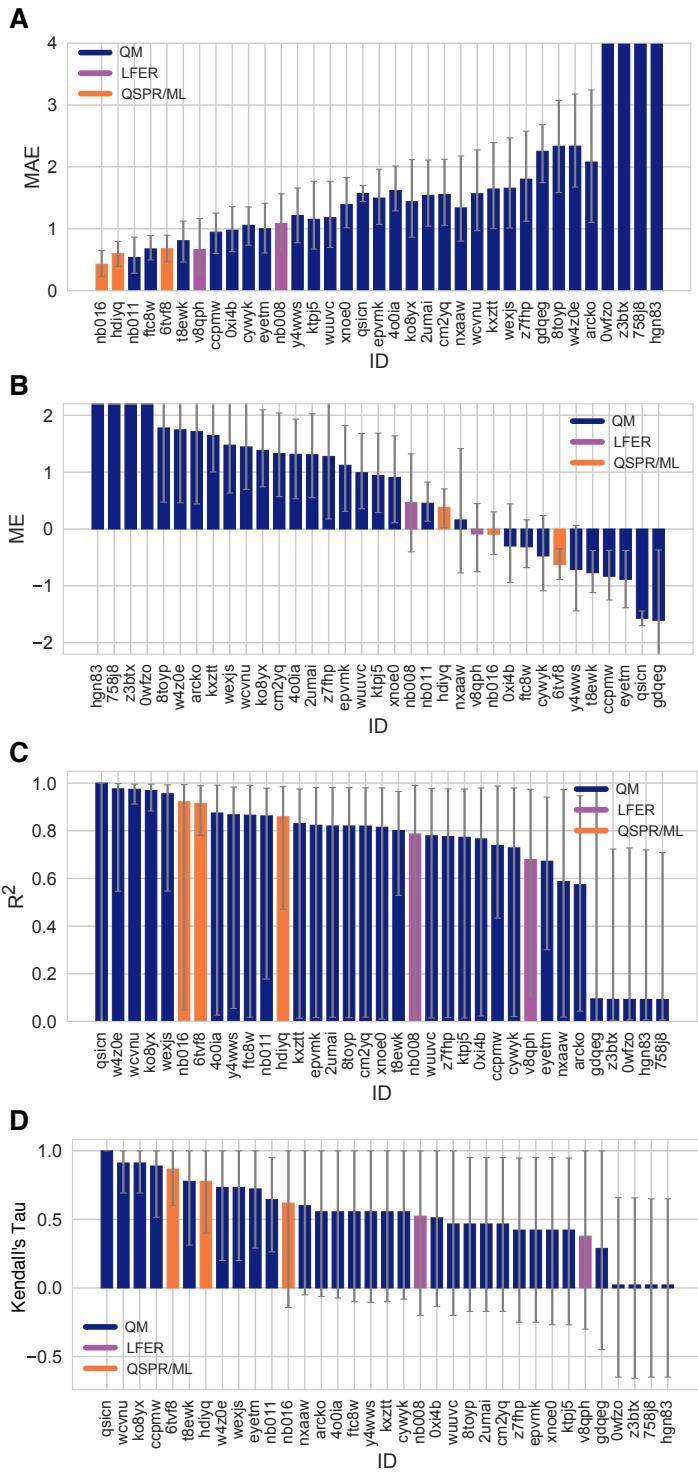


Figure 9. Additional performance statistics for microscopic pK_a predictions for 8 molecules with experimentally determined dominant microstates. Microstate-based matching was performed between experimental pK_a values and predicted microscopic pK_as. Mean absolute error (MAE), mean error (ME), Pearson's R², and Kendall's Rank Correlation Coefficient Tau (τ) 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.

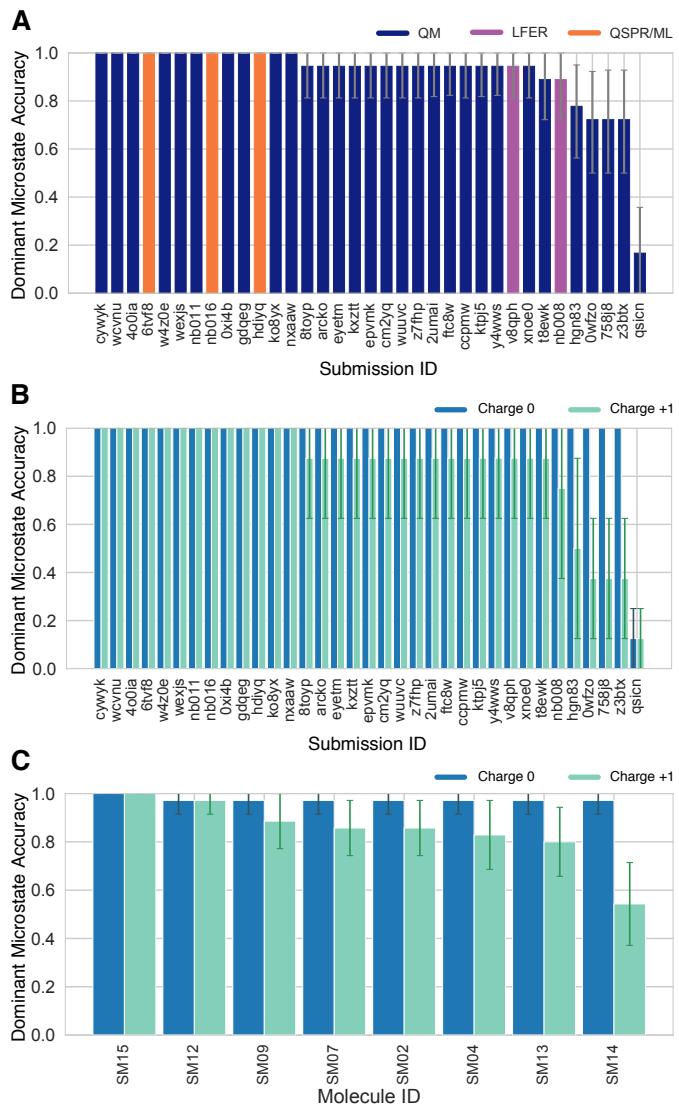
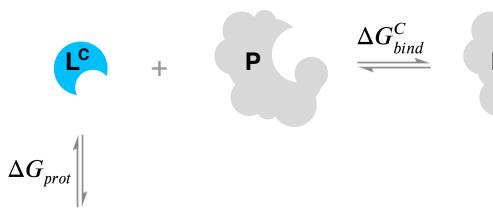


Figure 10. Some methods predicted the sequence of dominant tautomers inaccurately. Prediction accuracy of dominant microstate of each charged state was calculated using the dominant microstate sequence determined by NMR for 8 molecules as reference. **(A)** Dominant microstate accuracy vs. submission 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.

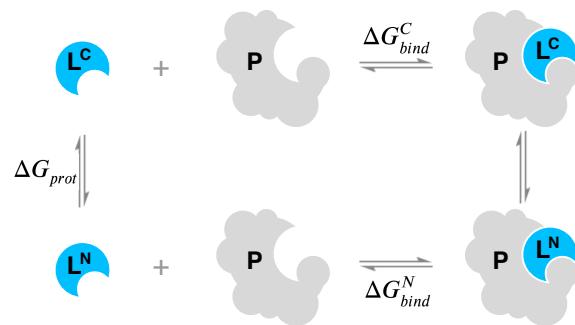
A When only the minor protonation state can bind to the protein



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

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

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 (ΔG_{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 (ΔG_{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 ($\Delta G_{bind}^C - \Delta G_{bind}^N$).

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

406

In addition, determining the free energy penalty of such states [3] also requires knowing the pK_a value.

407

EQUATION: free energy of protonation state equation

408

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

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

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

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

3.5 Lessons learned from SAMPL6 pK_a Challenge

409

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

410

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

411

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.

412

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.

413

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?

414

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

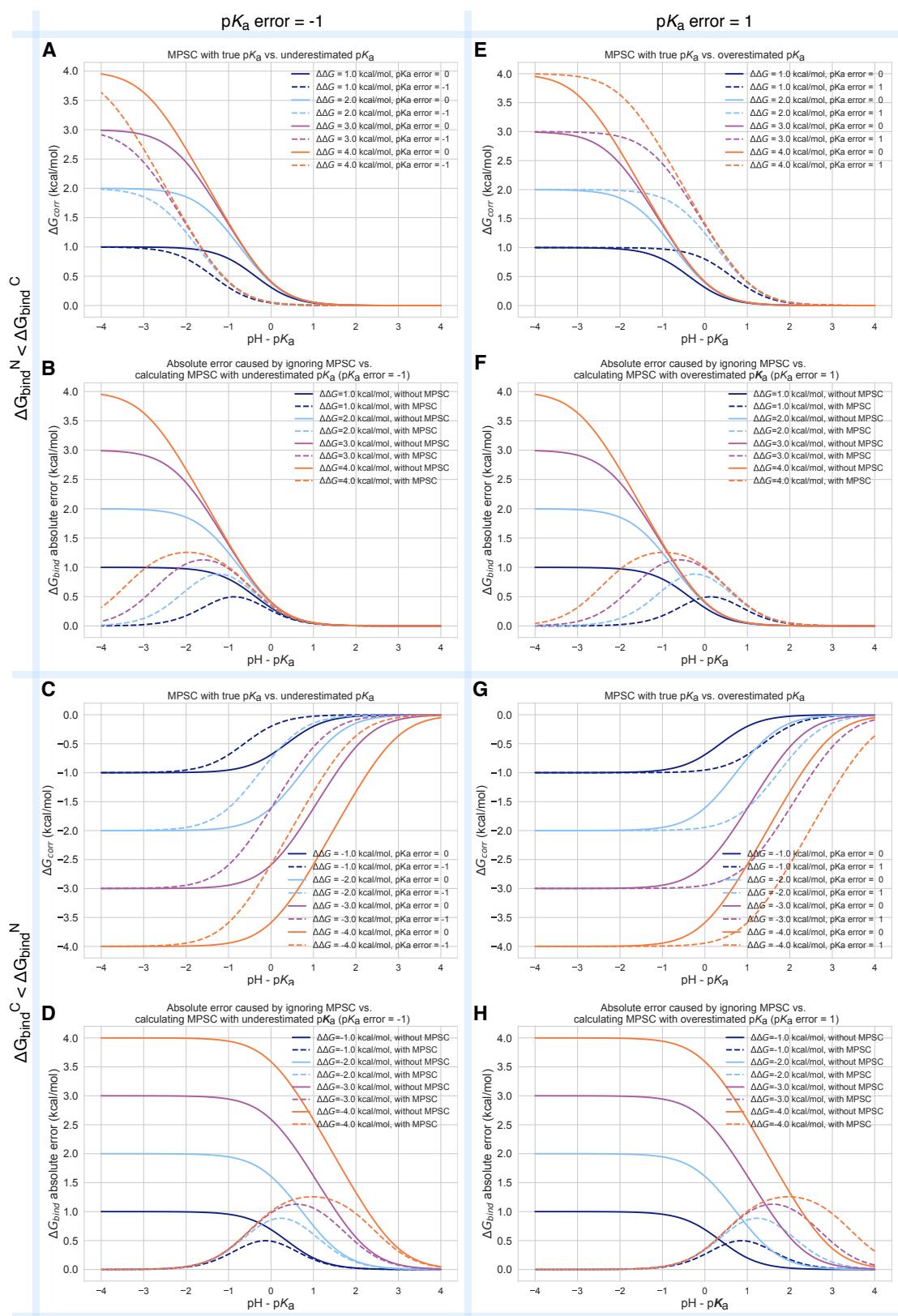


Figure 12. Inaccuracy of pK_a prediction (± 1 unit) affects the accuracy of MPSC and overall protein-ligand binding free energy calculation in varying amounts based on aqueous pK_a value and relative binding affinity of individual protonation states ($\Delta\Delta G = \Delta G_{bind}^C - \Delta G_{bind}^N$). All calculations are made for 25°C, and for a ligand with single basic titratable group. **A, C, E, and G show MPSC (ΔG_{corr}) calculated with true vs. inaccurate pK_a . **B, D, F, and H** show comparison of the absolute error to ΔG_{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 $\Delta\Delta G$.**

420 3.6 Suggestions for future challenges

421 In the SAMPL6 pK_a Challenge there wasn't a requirement that prediction sets should report predictions for all compounds.
422 Some participants reported predictions for only a subset of compounds which may lead these methods to look more accurate
423 than others, due to missing predictions. It would have been a better choice to require submissions for whole sets for better
424 comparison of method performance.

425 Discuss what can be done to further improve future challenges

426 How can we maximize what we learn? What should we have people predict? How should we select compounds / measure
427 pK_a s?

428 Suggestions about challenge construction

429 Future challenge direction Challenge path: predict pK_a s, give people pK_a s to predict logDs on same molecules, then predict
430 for new set of compounds logDs without provided pK_a s.

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

432 Suggestions about challenge analysis

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

434 Reporting microscopic pK_a predictions with charges, microstate free energies is better Experimental dataset with microstate
435 information is more helpful.

436 What can be done to further improve future challenges How can we maximize what we learn? What should we have people
437 predict? How should we select compounds / measure pK_a s? NMR experimental techniques could be used to validate microstate
438 information in future challenges

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

441 4 Conclusion

442 5 Code and data availability

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

444 6 Overview of supplementary information

445 Contents of the Supplementary Information:

- 446 TABLE S1: SMILES and InChI identifiers of SAMPL6 pK_a Challenge molecules.
- 447 TABLE S2: Evaluation statistics calculated for all macroscopic pK_a prediction submissions based on Hungarian match for
24 molecules.
- 448 TABLE S3: Evaluation statistics calculated for all microscopic pK_a prediction submissions based on Hungarian match for 8
molecules with NMR data.
- 449 TABLE S4: Evaluation statistics calculated for all microscopic pK_a prediction submissions based on microstate match for 8
molecules with NMR data.
- 450 FIGURE S1: Dominant microstates of 8 molecules were determined based on NMR measurements.
- 451 FIGURE S2: MAE of macroscopic pK_a predictions of each molecule did not show any significant correlation with any molec-
ular descriptor.
- 452 FIGURE S3: The value of macroscopic pK_a was not a factor affecting prediction error seen in SAMPL6 Challenge according
to the analysis with Hungarian matching.
- 453 FIGURE S4: There was low agreement between experimental dominant microstate pairs and the predicted microstate pairs
selected by Hungarian algorithm for microscopic pK_a predictions.

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

- 461 • SAMPL6-pKa-chemical-identifiers-table.csv
- 462 • macroscopic-pKa-statistics-24mol-hungarian-match.csv
- 463 • microscopic-pKa-statistics-8mol-hungarian-match-table.csv

- 464 • microscopic-pKa-statistics-8mol-microstate-match-table.csv
465 • experimental-microstates-of-8mol-based-on-NMR.csv
466 • enumerate-microstates-with-Epik-and-OpenEye-QUACPAC.ipynb
467 • molecule_ID_and_SMILES.csv

468 7 Author Contributions

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

472 8 Acknowledgments

473 Complete acknowledgments section. Caitlin Bannan for guidance on working microstate definition for the challenge, Thomas Fox for
474 MoKa reference calculations, Kiril Lanevskij for hungarian algorithm

475 MI, ASR, and JDC acknowledge support from the Sloan Kettering Institute. JDC acknowledges support from NIH grant P30
476 CA008748. MI acknowledges Doris J. Hutchinson Fellowship. We thank Brad Sherborne for his valuable insights at the conception
477 of the pK_a challenge and connecting us with Timothy Rhodes and Dorothy Levorse who were able to provide resources and
478 expertise for experimental measurements performed at MRL. We acknowledge Paul Czodrowski who provided feedback on
479 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.

480 Mike Chui

481 9 Disclosures

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

484 Table ref: [19, 20, 22, 23, 25] trial: [], +, -, *, #, \m

485 References

- 486 [1] **Manallack DT**, Prankerd RJ, Yuriev E, Oprea TI, Chalmers DK. The Significance of Acid/Base Properties in Drug Discovery. *Chem Soc Rev.* 2013; 42(2):485–496. doi: [10.1039/C2CS35348B](https://doi.org/10.1039/C2CS35348B).
- 487 [2] **Manallack DT**, Prankerd RJ, Nassta GC, Ursu O, Oprea TI, Chalmers DK. A Chemogenomic Analysis of Ionization Constants-Implications for Drug Discovery. *ChemMedChem*. 2013 Feb; 8(2):242–255. doi: [10.1002/cmdc.201200507](https://doi.org/10.1002/cmdc.201200507).
- 488 [3] **de Oliveira C**, Yu HS, Chen W, Abel R, Wang L. Rigorous Free Energy Perturbation Approach to Estimating Relative Binding Affinities
489 between Ligands with Multiple Protonation and Tautomeric States. *Journal of Chemical Theory and Computation*. 2019 Jan; 15(1):424–435.
doi: [10.1021/acs.jctc.8b00826](https://doi.org/10.1021/acs.jctc.8b00826).
- 490 [4] **Darvey IG**. The Assignment of pKa Values to Functional Groups in Amino Acids. *Biochemical Education*. 1995 Apr; 23(2):80–82. doi:
491 [10.1016/0307-4412\(94\)00150-N](https://doi.org/10.1016/0307-4412(94)00150-N).
- 492 [5] **Bodner GM**. Assigning the pKa's of Polyprotic Acids. *Journal of Chemical Education*. 1986 Mar; 63(3):246. doi: [10.1021/ed063p246](https://doi.org/10.1021/ed063p246).
- 493 [6] **Murray R**. Microscopic Equilibria. *Analytical Chemistry*. 1995 Aug; p. 1.
- 494 [7] **İşik 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).
- 495 [8] **Pickard FC**, König G, Tofoleanu F, Lee J, Simmonett AC, Shao Y, Ponder JW, Brooks BR. Blind Prediction of Distribution in the SAMPL5
496 Challenge with QM Based Protomer and pK_a Corrections. *Journal of Computer-Aided Molecular Design*. 2016 Nov; 30(11):1087–1100. doi:
497 [10.1007/s10822-016-9955-7](https://doi.org/10.1007/s10822-016-9955-7).
- 498 [9] **Bannan CC**, Mobley DL, Skillman AG. SAMPL6 Challenge Results from \$\$pK_a\$\$ Predictions Based on a General Gaussian Process Model.
499 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).

- 505 [10] **Işık M**, Levorse D, Mobley DL, Rhodes T, Chodera JD. Octanol–Water Partition Coefficient Measurements for the SAMPL6 Blind Prediction
506 Challenge. *Journal of Computer-Aided Molecular Design*. 2020 Apr; 34(4):405–420. doi: [10.1007/s10822-019-00271-3](https://doi.org/10.1007/s10822-019-00271-3).
- 507 [11] **Işık M**, Bergazin TD, Fox T, Rizzi A, Chodera JD, Mobley DL. Assessing the Accuracy of Octanol–Water Partition Coefficient Predictions in the
508 SAMPL6 Part II Log P Challenge. *Journal of Computer-Aided Molecular Design*. 2020 Apr; 34(4):335–370. doi: [10.1007/s10822-020-00295-0](https://doi.org/10.1007/s10822-020-00295-0).
- 509 [12] Special Issue: SAMPL6 (Statistical Assessment of the Modeling of Proteins and Ligands); October 2018. Volume 32, Issue 10. *Journal of
510 Computer-Aided Molecular Design*.
- 511 [13] **Fraczkiewicz R.** In Silico Prediction of Ionization. In: *Reference Module in Chemistry, Molecular Sciences and Chemical Engineering* Elsevier;
512 2013. doi: [10.1016/B978-0-12-409547-2.02610-X](https://doi.org/10.1016/B978-0-12-409547-2.02610-X).
- 513 [14] **Kuhn HW.** The Hungarian Method for the Assignment Problem. *Naval Research Logistics Quarterly*. 1955 Mar; 2(1-2):83–97. doi:
514 [10.1002/nav.3800020109](https://doi.org/10.1002/nav.3800020109).
- 515 [15] **Munkres J.** Algorithms for the Assignment and Transportation Problems. *J SIAM*. 1957 Mar; 5(1):32–28.
- 516 [16] SciPy v1.3.1, Linear Sum Assignment Documentation; Sep 27, 2019. The SciPy community. https://docs.scipy.org/doc/scipy-1.3.1/reference/generated/scipy.optimize.linear_sum_assignment.html.
- 517 [17] **Gunner MR**, Murakami T, Rustenburg AS, Işık M, Chodera JD. Standard State Free Energies, Not pKas, Are Ideal for Describing Small
518 Molecule Protonation and Tautomeric States. *Journal of Computer-Aided Molecular Design*. 2020 May; 34(5):561–573. doi: [10.1007/s10822-020-00280-7](https://doi.org/10.1007/s10822-020-00280-7).
- 519 [18] OpenEye pKa Prospector;. OpenEye Scientific Software, Santa Fe, NM. Accessed on Jan 23, 2018. <https://www.eyesopen.com/pka-prospector>.
- 520 [19] ACD/pKa GALAS (ACD/Percepta Kernel v1.6);. Advanced Chemistry Development, Inc., Toronto, ON, Canada, 2018. <https://www.acdlabs.com/products/percepta/predictors/pKa/>.
- 521 [20] ACD/pKa Classic (ACD/Percepta Kernel v1.6);. Advanced Chemistry Development, Inc., Toronto, ON, Canada, 2018. <https://www.acdlabs.com/products/percepta/predictors/pKa/>.
- 522 [21] **Shelley JC**, Cholleti A, Frye LL, Greenwood JR, Timlin MR, Uchimaya M. Epik: A Software Program for pK a Prediction and Protonation State
523 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).
- 524 [22] Simulations Plus ADMET Predictor v8.5;. Simulations Plus, Lancaster, CA, 2018. <https://www.simulations-plus.com/software/admetpredictor/physicochemical-biopharmaceutical/>.
- 525 [23] Chemicalize v18.23 (ChemAxon MarvinSketch v18.23);. ChemAxon, Budapest, Hungary, 2018. <https://docs.chemaxon.com/display/docs/pKa+Plugin>.
- 526 [24] **Milletti F**, Storchi L, Sforza G, Cruciani G. New and Original p K a Prediction Method Using Grid Molecular Interaction Fields. *Journal of
527 Chemical Information and Modeling*. 2007 Nov; 47(6):2172–2181. doi: [10.1021/ci700018y](https://doi.org/10.1021/ci700018y).
- 528 [25] MoKa;. Molecular Discovery, Hertfordshire, UK, 2018. <https://www.moldiscovery.com/software/moka/>.
- 529 [26] **Zeng Q**, Jones MR, Brooks BR. Absolute and Relative pKa Predictions via a DFT Approach Applied to the SAMPL6 Blind Challenge. *Journal
530 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).
- 531 [27] **Bochevarov AD**, Harder E, Hughes TF, Greenwood JR, Braden DA, Philipp DM, Rinaldo D, Halls MD, Zhang J, Friesner RA. Jaguar: A High-
532 Performance Quantum Chemistry Software Program with Strengths in Life and Materials Sciences. *International Journal of Quantum
533 Chemistry*. 2013 Sep; 113(18):2110–2142. doi: [10.1002/qua.24481](https://doi.org/10.1002/qua.24481).
- 534 [28] **Selwa E**, Kenney IM, Beckstein O, Iorga BI. SAMPL6: Calculation of Macroscopic pKa Values from Ab Initio Quantum Mechanical Free
535 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).
- 536 [29] **Tielker N**, Eberlein L, Güssregen S, Kast SM. The SAMPL6 Challenge on Predicting Aqueous pKa Values from EC-RISM Theory. *Journal of
537 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).
- 538 [30] **Klamt A**, Eckert F, Diedenhofen M, Beck ME. First Principles Calculations of Aqueous p K a Values for Organic and Inorganic Acids Using
539 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:
540 [10.1021/jp034688o](https://doi.org/10.1021/jp034688o).
- 541 [31] **Eckert F**, Klamt A. Accurate Prediction of Basicity in Aqueous Solution with COSMO-RS. *Journal of Computational Chemistry*. 2006 Jan;
542 27(1):11–19. doi: [10.1002/jcc.20309](https://doi.org/10.1002/jcc.20309).

- 550 [32] Pracht P, Wilcken R, Udvarhelyi A, Rodde S, Grimme S. High Accuracy Quantum-Chemistry-Based Calculation and Blind Prediction of
551 Macroscopic pKa Values in the Context of the SAMPL6 Challenge. *Journal of Computer-Aided Molecular Design*. 2018 Oct; 32(10):1139–
552 1149. doi: [10.1007/s10822-018-0145-7](https://doi.org/10.1007/s10822-018-0145-7).
- 553 [33] Prasad S, Huang J, Zeng Q, Brooks BR. An Explicit-Solvent Hybrid QM and MM Approach for Predicting pKa of Small Molecules in SAMPL6
554 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).
- 555 [34] OEMolProp Toolkit 2017.Feb.1;. OpenEye Scientific Software, Santa Fe, NM. <http://www.eyesopen.com>.

Table S1. SMILES and InChI identifiers of SAMPL6 pK_a 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)N3CCCC3	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)

556 10 Supplementary Information

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_a values were determined by spectrophotometric pK_a measurements [7]. A CSV version of this table can be found in SAMPL6-supplementary-documents.tar.gz.

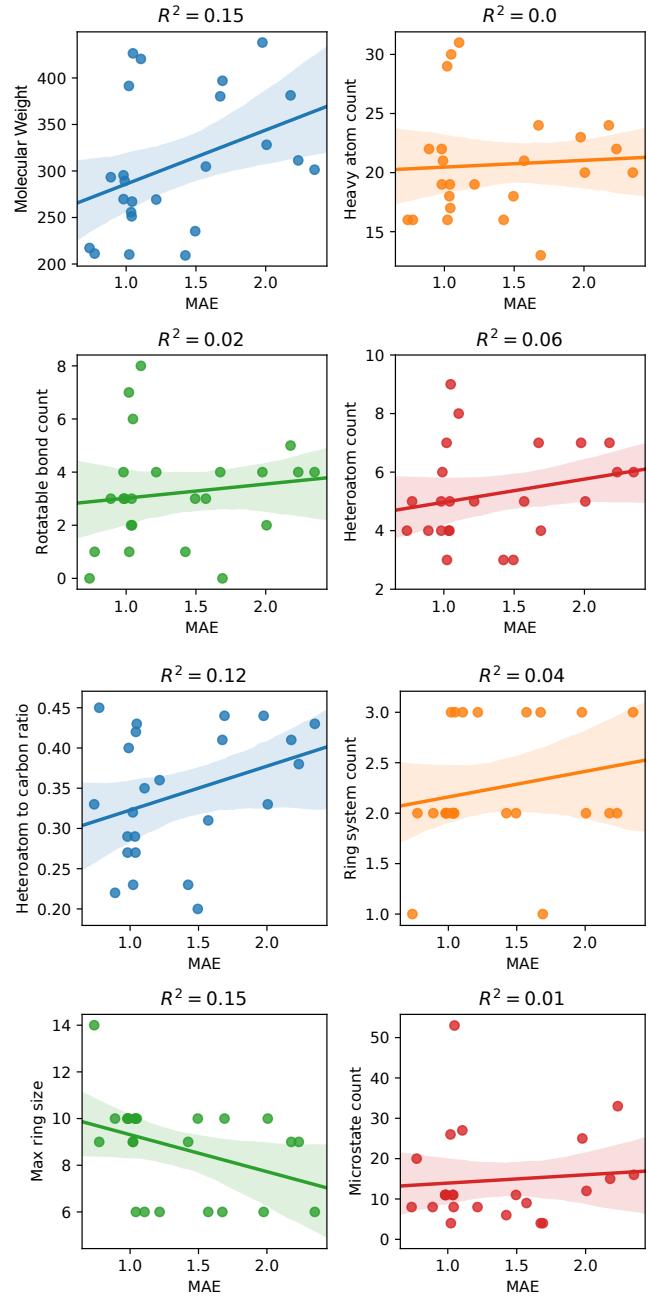


Figure S2. MAE of macroscopic pK_a predictions of each molecule did not show any significant correlation with any molecular descriptor.
 Plots show regression lines, 96% confidence intervals of the regression lines, and R_2 . The following molecular descriptors were calculated using OpenEye OEMolProp Toolkit [34].

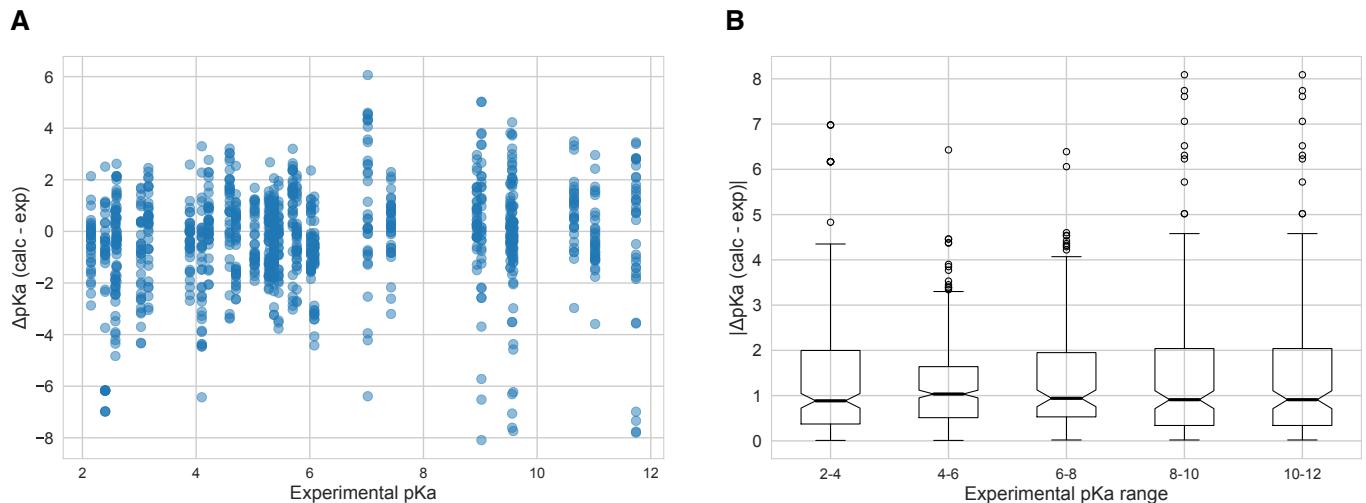


Figure S3. The value of macroscopic pK_a s was not a factor affecting prediction error seen in SAMPL6 Challenge according to the analysis with Hungarian matching. There was not clear trend between pK_a prediction error and the true pK_a error. Very high and very low pK_a values have similar inaccuracy 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 .

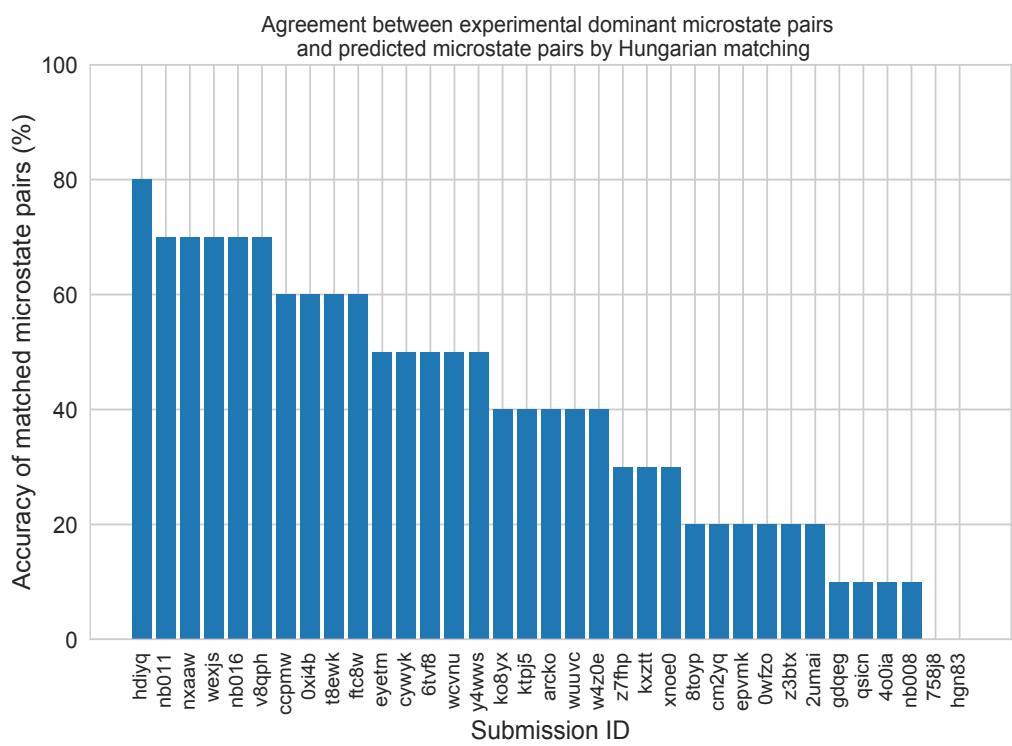


Figure S4. There was low agreement between experimental dominant microstate pairs and the predicted microstate pairs selected by Hungarian algorithm for microscopic pK_a predictions. This analysis could only be performed for 8 molecules with NMR data. Hungarian matching algorithm which matches predicted and experimental values considering only the closeness of the numerical value of pK_a and it often leads to predicted pK_a matches that described a different microstates pair than the experimentally observed dominant microstates..

Table S2. Evaluation statistics calculated for all macroscopic pK_a 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 (τ), unmatched experimental pK_as (number of missing pK_a predictions) and unmatched predicted pK_as (number of extra pK_a 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 ²	m	Kendall's Tau	Unmatched exp. pK _a s	Unmatched pred. pK _a 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>37xm8</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_a 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 (τ), unmatched experimental pK_as (number of missing pK_a predictions) and unmatched predicted pK_as (number of extra pK_a 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 ²	m	Kendall's Tau	Unmatched exp. pK _a s	Unmatched pred. pK _a 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_a 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 (τ), unmatched experimental pK_as (number of missing pK_a predictions) and unmatched predicted pK_as (number of extra pK_a 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 ²	m	Kendall's Tau	Unmatched exp. pK _a s	Unmatched pred. pK _a 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