

Predicting pKa values and titration curves for the SAMPL6 pKa challenge using Epik and Jaguar

Ariën S. Rustenburg^{1,2}, Mehtap Isik¹, Patrick B. Grinaway^{1,2}, Andrea Rizzi^{1,3}, Marilyn R Gunner⁴, John D. Chodera^{1*}

¹Computational and Systems Biology Program, Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, New York, NY 10065; ²Graduate Program in Physiology, Biophysics, and Systems Biology, Weill Cornell Medical College, New York, NY 10065; ³Tri-Institutional Training Program in Computational Biology and Medicine, New York, NY 10065; ⁴Department of Physics, City College of New York, New York, NY 10031

*For correspondence:

john.chodera@choderalab.org (JDC)

Abstract The goal of the SAMPL6 pKa Challenge was to evaluate the performance of small molecule pKa prediction methods from challenge participants in a blinded fashion on a set of small, drug-like molecules resembling kinase inhibitor fragments. To provide a useful point of comparison for blind participant predictions that may use experimental methods still under development, we performed reference benchmark calculations using a popular empirical model (Epik) and quantum chemical approach (Jaguar) from Schrödinger. Epik predicts microstate populations and pKa values using a Hammett-Taft type model, while Jaguar is a fast DFT quantum chemical method. In this work, we discuss how these reference calculations were performed, provide a broad assessment of the performance of the method, and highlight challenges and considerations in predicting pKa values to benchmark against experiment macroscopic pKa measurements.

Introduction

Titrateable sites are ubiquitous in druglike small molecules. Large-scale computational surveys suggest that 60% of all protein-ligand complexes undergo a change in ionization state upon binding [1], either due to protonation state changes of the small molecule or the protein (where roughly a third of all protein residues are ionizable [2]). More generally, protonation state effects—in which the dominant protonation, charge, or tautomer state shifts upon binding, or a mixture of protonation states are significantly populated in complex or solution—has the potential to cause large modeling errors if these effects are neglected. [3–5] In the SAMPL5 distribution coefficient (logD) challenge, for example, protonation state effects were determined to be a major contributor to loss in accuracy for the otherwise mundane task of predicting a transfer free energy between aqueous and cyclohexane phases [6].

To isolate the question of how well pKa effects could be modeled—and therefore how accurately the community could address these effects—the SAMPL6 challenge featured a blind pKa prediction component, as an intermediate step to logD predictions in which we provide participants with pKa values and later have them predict both pKa and logD [?]. The SAMPL6 pKa challenge consisted of predicting macroscopic pKa values measured by UV-metric titration for a set of small molecules that resembled kinase inhibitors and their fragments [?]. As participants in the SAMPL6 pKa challenge were expected to utilize a wide variety of methods still under development, we endeavored to provide a useful baseline reference set of predictions using well-established, widely-deployed, commercially available methods. We selected both an empirical method (Epik [7]) and quantum chemical method (Jaguar [8]) from the Schrödinger Suite of computational chemistry software, version 2017-4.

Reference calculations—which were not fully blinded—were performed in a manner that attempted to mimic standard use, using recommended settings for each program, without significantly modifying the input parameters. As the computation of UV-metric macroscopic pKa values from the microscopic pKa values predicted by the tools is not necessarily completely straightforward, we considered several alternative possibilities, which we discuss in more detail. We provide an analysis and

broad assessment of the performance of the two methods, and highlight challenges and considerations in predicting pKa values to benchmark against experiment. All analysis tools used to perform this study are available via GitHub at <https://github.com/choderalab/SAMPL6-Reference-pKa-Calculations>

Technology

Epik

Epik[7] is a program for predicting the pKa values of ionizable sites in druglike molecules and for generating their probable protonation states. Epik implements the empirical Hammett-Taft (HT) approach [9] to rapidly estimate the pKa values of Brønsted acidic or basic functional groups identified by SMARTS [10] patterns. The popular HT method describes a linear free energy relationship,

$$pK_{a_i} = pK_{a_i}^0 - \rho_i \sum_j \sigma_j \quad (1)$$

between an ionizable functional group reference $pK_{a_i}^0$ value, how sensitive (i) the group is to substituent perturbation, and how strongly (σ_j) each substituent, j, perturbs the pKa value. Epik's internal $pK_{a_i}^0$, ρ , and σ parameters are fit to experimental data, or, rarely, to high-quality Jaguar calculations when experiments are unavailable.

Given a target pH and pH tolerance, the molecule can be appropriately protonated or deprotonated based on the estimated pKa values of its ionizable groups to generate the probable protonation states. The estimated pKa value of each functional group in a molecule is often exclusive to a specific tautomeric, protonation, and ionization state. Each (de)protonation can give rise to a set of tautomers and subsequent ionization states, all of which are iteratively generated and estimated until either exhausted or the predicted pKa values of the ionizable sites fall outside the tolerance range. Finally, the microscopic populations and corresponding free energy penalties are calculated and reported. Please note that Epik reports only microscopic pKa values, not macroscopic, or apparent pKa values.

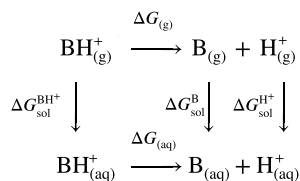
Jaguar

Jaguar uses an *ab initio* DFT method for calculating pKa values.

It attempts to estimate the pKa using the free energy of deprotonation

$$pK_a = \frac{\Delta G_{(aq)}}{RT \ln 10} \quad (2)$$

The value of $\Delta G_{(aq)}$ can be obtained using the following thermodynamic cycle:



, which leads to the following equation.

$$\Delta G_{(aq)} = \Delta G_{(g)} - \Delta G_{\text{sol}}^{\text{BH}^+} + \Delta G_{\text{sol}}^{\text{B}} + \Delta G_{\text{sol}}^{\text{H}^+} \quad (3)$$

From this equation, it follows that when the other 4 terms are known, the pKa value can be estimated.

$\Delta G_{\text{sol}}^{\text{H}^+}$ has previously been established to be -259.5 kcal per mol[11].

Gas phase term $\Delta G_{(g)}$ is calculated using

$$\Delta G_{(g)} = E_{\text{B}_{(g)}} - E_{\text{BH}_{(g)}^+} + \frac{5RT}{2} - T\Delta S \quad (4)$$

The energy terms require first a geometry optimization, which is performed using B3LYP/6-31G*. Afterwards, B3LYP/cc-pVTZ+ is used for atoms involved in the deprotonation reaction, and the remainder are treated using the cc-pVTZ basis set.

ASR: Describe solvation calculation and need for empirical correction

ASR: Describe how multiple conformations are incorporated

ASR: Explain the individual terms in the equations above, citations for method

Analysis

Microstate predictions

Methods like Epik and Jaguar provide microstate pKa predictions. This means that rather than macroscopic pKas observed in the average charge state (potentiometric titration) or number of titration events (UV titration curve), it produces a microequilibrium constant between a single pair of configurations (protonation states or tautomers) is predicted. One way to compare experiment and prediction is by directly comparing pKa values. However, there are likely many microstates, and only a few macrostates. Therefore, we also look at the predicted titration curve, which can incorporate information from every microstate into a macrostate.

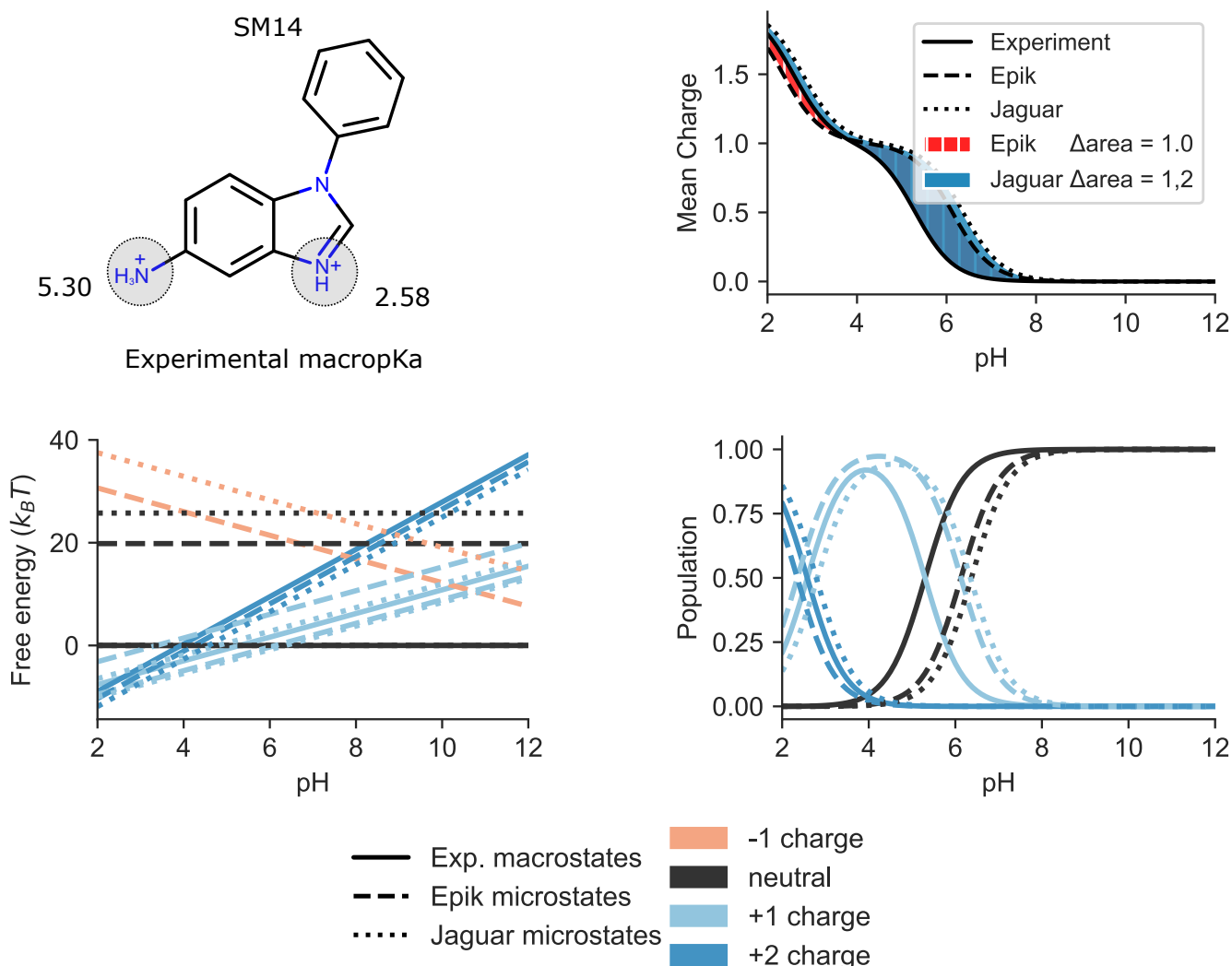


Figure 1. Epik and Jaguar microscopic pKa values can be used to generate free energies (bottom left) or, populations (bottom right), and recapitulate the binding curve (top right) when the charge of each species is added up. Molecule SM14 pKa values (top left) were measured using UV (citation) for the purpose of the SAMPL6 challenge, and based on a subsequent NMR experiment the most likely sites of titration were established. We used Epik and Jaguar to predict microscopic pKa values (denoted as Type I predictions in SAMPL6). The macrostate populations from experiment, and microstate populations from predictions can be compared directly to an experimental titration curve derived from UV. Since NMR data was available for this molecule, the charge state in the experiment is also known.

Predicting macroscopic pKa values

Epik and Jaguar predict microscopic pKa values or microstate energies, which must be translated into macroscopic pKa values for comparison to experiment. Several choices are possible for comparing these microscopic properties into macroscopic pKa values. It is also possible to translate the micropKas into macropKas [12]

$$K_a^{\text{macro}} = \sum_{j=1}^{N_{\text{deprot}}} \frac{1}{\sum_{i=1}^{N_{\text{prot}}} \frac{1}{K_{ij}^{\text{micro}}}} \quad (5)$$

See if I can perform those conversions for my data.

Population curves from pKa

- Description of the population curve generation from pKa

$$g_i(\text{pH}) = \beta \left(n_i * \text{pH} - \sum_j \text{pKa}_j \right) \quad (6)$$

$$\pi_i(\text{pH}) = \frac{e^{-g_i(\text{pH})}}{\sum_i e^{-g_i(\text{pH})}} \quad (7)$$

Virtual electrochemical titration

- Description of the calculation of the mean charge curve

$$\langle q_{\text{total}} \rangle(\text{pH}) = \sum_i q_i \times \pi_i(\text{pH}) \quad (8)$$

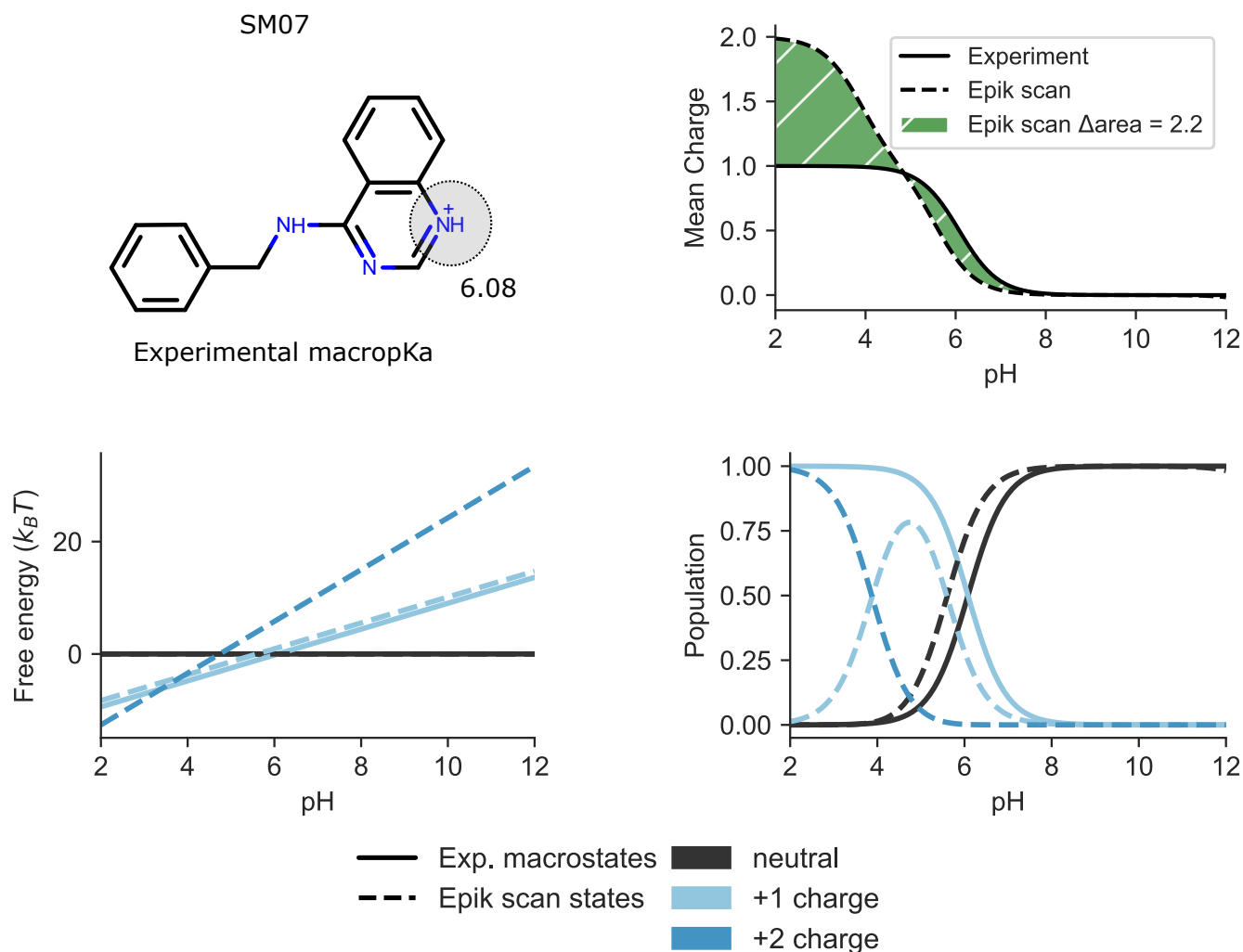


Figure 2. Epik scan provides sequential pKa values for adding single protons, which can approximate a macroscopic titration curve. Molecule SM07 pKa values were measured using UV (citation) for the purpose of the SAMPL6 challenge, and based on a subsequent NMR experiment the identity of the microstates were identified. We used Epik scan mode to predict most prevalent microstates from sequential pKa values, which can approximate macroscopic pKa values. The macrostate populations from experiment, and microstate populations from predictions can be compared directly to an experimental titration curve derived from UV pKa values. Since NMR data was available for this molecule, the charge identity in the experiment is also known.

- Sequential titration avoids the need to compute the energies of *all* protonation states
- Epik can automatically perform a sequential scan (which can also be used in Jaguar with some automation)
- For this experiment, we started from the predicted highest-occupancy microstate at pH 7 and went in both directions, returning pKa values between 2-12
- Average charge vs pH to match electrochemical titrations, look at the mean signed deviation (MSD) or area between curves to see whether the charge curve behavior matches

Results

Table 1. Overall performance of each method using either the closest pKa matching, Hungarian pKa matching, or sequentially aligned pKa values, and the area between the experimental and predicted macroscopic charge titration curve

	pKa		
	RMSE		
	closest	hungarian	align
Epik micropka	0.5; [0.3, 0.7]	0.5; [0.3, 0.7]	—
Epik scan	0.8; [0.6, 1.0]	0.8; [0.6, 1.0]	0.8; [0.6, 1.0]
Jaguar	0.3; [0.2, 0.5]	0.3; [0.2, 0.5]	—
	Pearson ρ		
	closest	hungarian	align
	closest	hungarian	align
Epik micropka	0.96; [0.93, 0.99]	0.96; [0.92, 0.99]	—
Epik scan	0.94; [0.88, 0.97]	0.94; [0.87, 0.97]	0.94; [0.87, 0.97]
Jaguar	0.986; [0.966, 0.995]	0.986; [0.969, 0.995]	—
	Charge vs pH		
	Δ area		
	closest	hungarian	align
Epik micropka	1.8; [1.5, 2.1]		
Epik scan	1.7; [1.2, 2.2]		
Jaguar	1.5; [1.0, 2.1]		

- How much conformation-dependence is there in Jaguar-derived pKa values (or state energies)?
- Does sequential scan or mean molecular charge provide better agreement with experimental macroscopic pKa values? Would it be worthwhile to develop predictive UV-metric models?
- Comparison of observed accuracies to previously reported/expected accuracies; expected accuracy on kinase inhibitors derived from this study
- How does Epik compare to Jaguar in terms of accuracy and computational cost?
- Discussion of outliers

Overall performance

Some low probability structures produced by Epik included questionable protonation states of a heterocyclic moiety, present in SM03 and SM18 (Figure 8).

Overall performance of Epik and Jaguar based on various metrics (figures/tables)

Matching of experimental and calculated pKa values

The experiments do not provide any microscopic information on what atom, or microstate a pKa belongs to. Therefore, it is necessary to perform a matching of pKa values between experiment and predicted pKa values. There are several ways one could go about this, each strategy can prioritize a different aspect to match. We consider three algorithms:

115 Closest pKa matching

Table 2. Mean charge titration curve comparison with respect to experiment (Δ area) with confidence intervals. Confidence intervals were obtained by gaussian bootstrapping over pKa values from the standard errors reported by the software.

Molecule	Epik micropka	Jaguar	Epik scan
SM01	0.5; [0.2, 4.9]	1.1; [0.0, 3.5]	0.5; [0.1, 4.0]
SM02	2.1; [0.2, 3.1]	1.6; [0.3, 5.3]	1.9; [0.7, 4.2]
SM03	1.5; [1.0, 4.9]	0.5; [0.8, 4.8]	0.1; [0.2, 4.7]
SM04	2.6; [0.2, 4.0]	1.1; [0.1, 4.0]	2.2; [0.3, 5.0]
SM05	2.2; [0.6, 4.7]	0.2; [0.1, 4.5]	1.1; [0.5, 4.6]
SM06	2.2; [0.7, 9.5]	1.3; [0.6, 4.7]	2.6; [0.6, 6.8]
SM07	2.6; [0.1, 4.1]	0.8; [0.1, 4.1]	2.2; [0.7, 4.8]
SM08	1.2; [0.5, 6.2]	7.1; [3.3, 9.2]	0.6; [0.1, 5.8]
SM09	2.5; [0.3, 3.4]	1.0; [0.1, 4.3]	2.3; [0.6, 3.3]
SM10	0.6; [0.2, 5.1]	0.3; [0.2, 5.3]	0.6; [0.1, 5.6]
SM11	1.9; [0.2, 4.1]	3.9; [0.5, 6.8]	0.1; [0.2, 5.0]
SM12	2.5; [0.2, 3.4]	0.9; [0.4, 4.7]	2.2; [1.1, 3.6]
SM13	2.6; [0.2, 3.8]	0.5; [0.1, 4.0]	2.8; [0.3, 5.0]
SM14	1.0; [0.8, 5.4]	1.2; [0.1, 4.1]	1.8; [0.6, 5.2]
SM15	1.1; [0.4, 4.6]	0.3; [0.2, 2.4]	1.3; [0.4, 4.2]
SM16	2.9; [0.7, 4.7]	1.5; [0.5, 4.4]	1.4; [0.6, 4.7]
SM17	1.7; [0.1, 3.2]	0.1; [0.0, 1.6]	1.7; [0.2, 3.1]
SM18	0.8; [1.2, 9.7]	2.2; [1.5, 7.9]	1.0; [1.1, 8.2]
SM19	0.7; [0.2, 5.2]	1.8; [0.6, 6.7]	0.7; [0.2, 5.9]
SM20	2.3; [0.1, 4.7]	1.6; [0.2, 3.4]	2.3; [0.1, 4.8]
SM21	2.2; [0.8, 5.5]	1.6; [0.5, 7.0]	0.5; [0.4, 6.7]
SM22	1.6; [0.5, 4.8]	0.2; [0.1, 3.4]	1.6; [0.3, 3.9]
SM23	0.4; [0.2, 3.5]	2.1; [0.7, 5.7]	3.1; [1.9, 4.8]
SM24	3.6; [0.1, 5.0]	2.3; [0.2, 4.9]	5.6; [3.0, 8.6]

116 Hungarian pKa matching

117 This algorithm, also known as linear sum assignment. It finds the combination of pKa that minimizes the overall cost by picking
 118 rows and columns in a matrix, also considering the cost of not mapping certain pKa values in the case of different numbers of
 119 predictions and experimental values. The results do not seem to differ from closest-pka matching.

120 Sequential pKa matching

121 For macroscopic only, because it doesn't make sense to sequentially align microscopic pKa values to a set of macroscopic pKa
 122 values algorithm 2

123 Macroscopic titration curves

- 124 • Compare macroscopic titration results
- 125 • different types of mistakes
 - 126 1. wrong pKa value
 - 127 2. too many pKa values (wrong total charges)

ASR: Update these with latest version with correct rho/rmse values

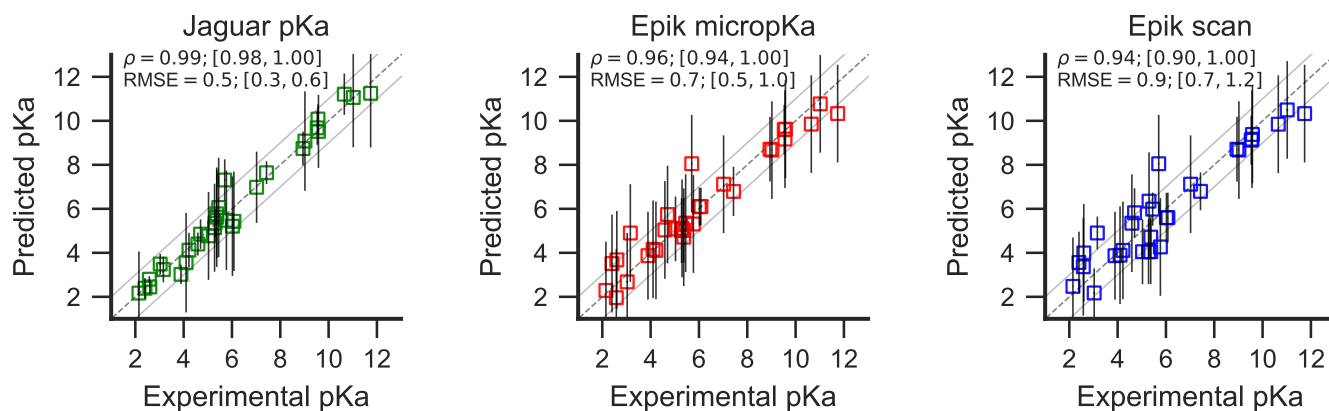


Figure 3. Comparison of computed pKa values to experimental macroscopic pKa values. *Left:* Microscopic pKa values for Jaguar were matched to the experimental dataset using algorithm 1. An alternative alignment of pKa values using the Hungarian algorithm (??) is available as a supplementary figure. *Middle:* Microscopic pKa values for Epik were matched to the experimental dataset using algorithm 1. An alternative alignment of pKa values using the Hungarian algorithm (??) is available as a supplementary figure. *Right:* Macroscopic pKa values for Jaguar were matched to the experimental dataset using algorithm 1. An alternative alignment of pKa values using the Hungarian algorithm (??) is available as a supplementary figure.

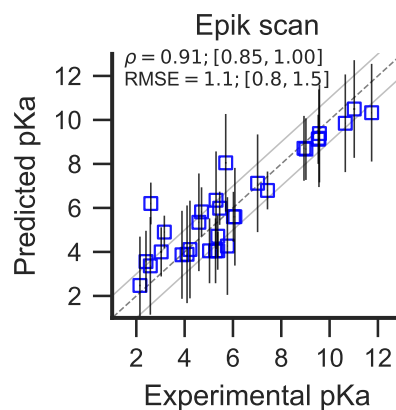


Figure 4. Epik macroscopic pKa values compared to macroscopic experimental pKa values, sequentially matching pKa values Comparison of the Epik sequential scan with entire experimental data set is shown, allowing only sequential pKa values to match experiment.

Table 3. Macroscopic pKa values computed via Epik's sequential scan procedure. Experimental and computed pKa values matched using sequential alignment of pKa values.

Molecule	Experimental	Epik-scan
SM01	9.53 ± 0.01	9 \pm 1
SM02	5.03 ± 0.01	4 \pm 1
SM03	7.02 ± 0.01	7 \pm 2
SM04	6.02 ± 0.01	6 \pm 1
SM05	4.59 ± 0.01	5 \pm 2
SM06	3.03 ± 0.04	4 \pm 1
SM06	11.74 ± 0.01	10 \pm 2
SM07	6.08 ± 0.01	6 \pm 2
SM08	4.22 ± 0.01	4 \pm 2
SM09	5.37 ± 0.01	4.0 \pm 0.9
SM10	9.02 ± 0.01	9 \pm 1
SM11	3.89 ± 0.01	4 \pm 2
SM12	5.28 ± 0.01	4 \pm 1
SM13	5.77 ± 0.01	4 \pm 2
SM14	2.58 ± 0.01	3 \pm 2
SM14	5.30 ± 0.01	6 \pm 2
SM15	4.70 ± 0.01	6 \pm 1
SM15	8.94 ± 0.01	9 \pm 1
SM16	5.37 ± 0.01	4.7 \pm 0.9
SM16	10.65 ± 0.01	10 \pm 2
SM17	3.16 ± 0.01	4.9 \pm 0.7
SM18	2.15 ± 0.02	2 \pm 2
SM18	9.58 ± 0.03	9 \pm 2
SM18	11.02 ± 0.04	10 \pm 2
SM19	9.56 ± 0.02	9 \pm 2
SM20	5.70 ± 0.03	8 \pm 2
SM21	4.10 ± 0.01	4 \pm 2
SM22	2.40 ± 0.02	4 \pm 1
SM22	7.43 ± 0.01	6.8 \pm 0.9
SM23	5.45 ± 0.01	6.0 \pm 0.8
SM24	2.60 ± 0.01	4 \pm 1

Table 4. Microscopic pKa per molecule for each method as matched by the "closest" algorithm (algorithm 1), compared to experiment. The uncertainty indicated is the standard error from the experiment, and the reported standard error by the prediction method.

Add deviation from experiment columns

Molecule	Experimental	Epik-micropka	Jaguar	Epik-scan
SM01	9.53 ± 0.01	10 ± 1	9.7 ± 0.8	9 ± 1
SM02	5.03 ± 0.01	5 ± 1	5 ± 2	4 ± 1
SM03	7.02 ± 0.01	7 ± 2	7 ± 2	7 ± 2
SM04	6.02 ± 0.01	6.1 ± 0.9	5 ± 2	6 ± 1
SM05	4.59 ± 0.01	5 ± 2	4.4 ± 0.5	5 ± 2
SM06	3.03 ± 0.04	3 ± 2	3.5 ± 0.5	2.2 ± 0.9
SM06	11.74 ± 0.01	10 ± 2	11 ± 2	10 ± 2
SM07	6.08 ± 0.01	6.1 ± 0.9	5 ± 2	6 ± 2
SM08	4.22 ± 0.01	4 ± 2	4.1 ± 0.8	4 ± 2
SM09	5.37 ± 0.01	5 ± 1	6 ± 2	4.0 ± 0.9
SM10	9.02 ± 0.01	9 ± 2	9 ± 2	9 ± 1
SM11	3.89 ± 0.01	4 ± 2	3.0 ± 0.4	4 ± 2
SM12	5.28 ± 0.01	5 ± 1	5 ± 2	4 ± 1
SM13	5.77 ± 0.01	5 ± 2	5 ± 2	4 ± 2
SM14	5.30 ± 0.01	5 ± 2	5.3 ± 0.7	6 ± 2
SM14	2.58 ± 0.01	3 ± 2	2.8 ± 0.1	3 ± 2
SM15	8.94 ± 0.01	9 ± 1	8.7 ± 0.8	9 ± 1
SM15	4.70 ± 0.01	6 ± 1	4.8 ± 0.7	6 ± 1
SM16	5.37 ± 0.01	5 ± 2	6 ± 2	4.7 ± 0.9
SM16	10.65 ± 0.01	10 ± 2	11.2 ± 0.9	10 ± 2
SM17	3.16 ± 0.01	4.9 ± 0.8	3.2 ± 0.6	4.9 ± 0.7
SM18	9.58 ± 0.03	10 ± 2	10 ± 2	9 ± 2
SM18	2.15 ± 0.02	2 ± 2	2 ± 2	2 ± 2
SM18	11.02 ± 0.04	11 ± 2	11 ± 2	10 ± 2
SM19	9.56 ± 0.02	9 ± 2	10.1 ± 0.7	9 ± 2
SM20	5.70 ± 0.03	8 ± 2	7.3 ± 0.9	8 ± 2
SM21	4.10 ± 0.01	4 ± 2	4 ± 2	4 ± 2
SM22	7.43 ± 0.01	6.8 ± 0.9	7.6 ± 0.5	6.8 ± 0.9
SM22	2.40 ± 0.02	4 ± 2	2.4 ± 0.5	4 ± 1
SM23	5.45 ± 0.01	5 ± 2	6 ± 2	6.0 ± 0.8
SM24	2.60 ± 0.01	4 ± 2	2.5 ± 0.3	4 ± 1

Discussion

Problematic Lewis structures

Epik generates some structures that seem to have invalid lewis structures. fig. 8.

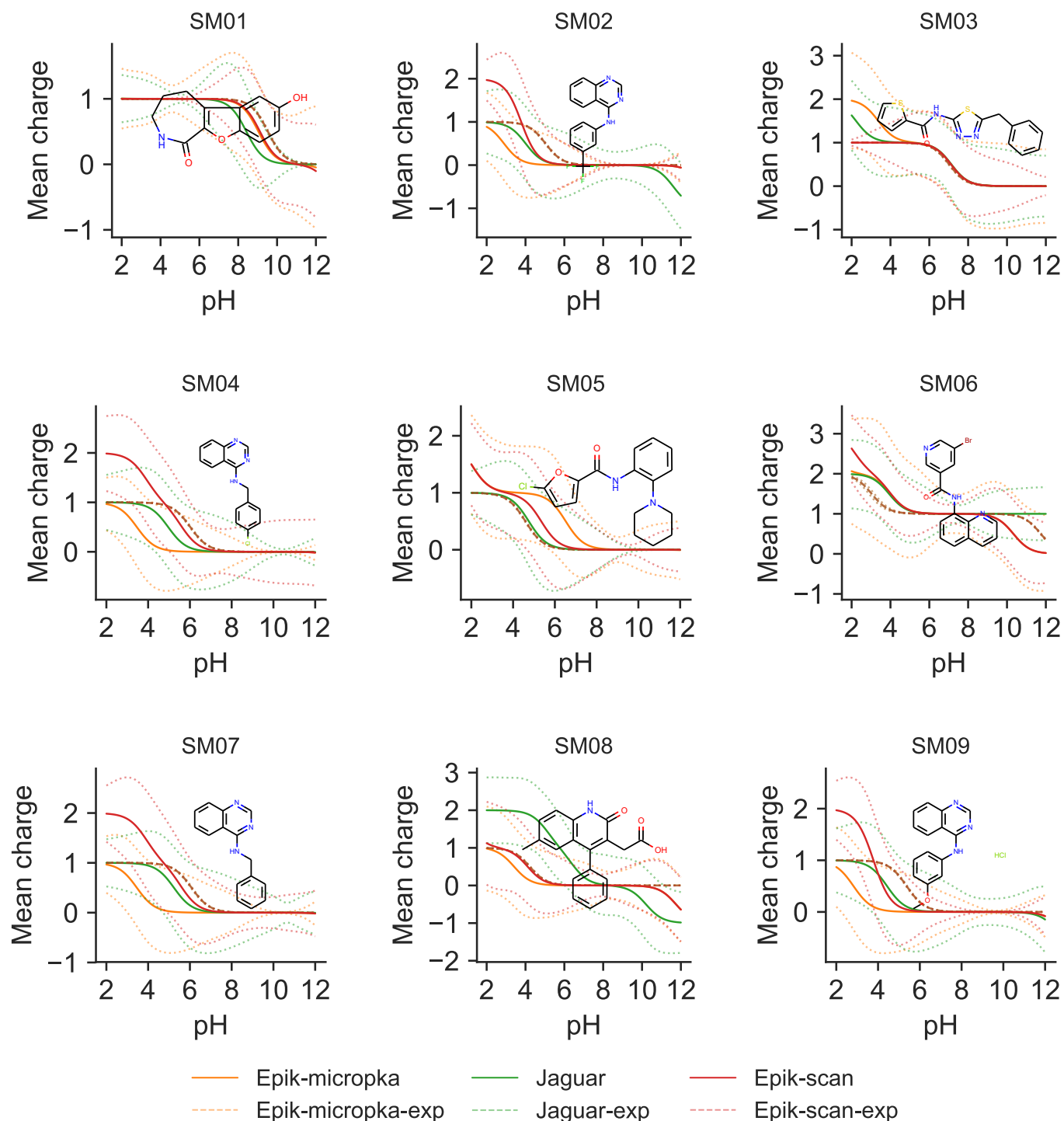


Figure 5. Bootstrap titration curves for each method compared to experiment for molecules SM01-SM09. The titration curve for each method is shown as a solid line, and 96 % confidence intervals from bootstrap have been shown as dotted lines. Since the absolute experimental charge is not available in general, experimental curves (dashed) were aligned to each prediction independently using an integer offset that minimizes the area between curves. A numerical comparison to experiment is presented in table 2.

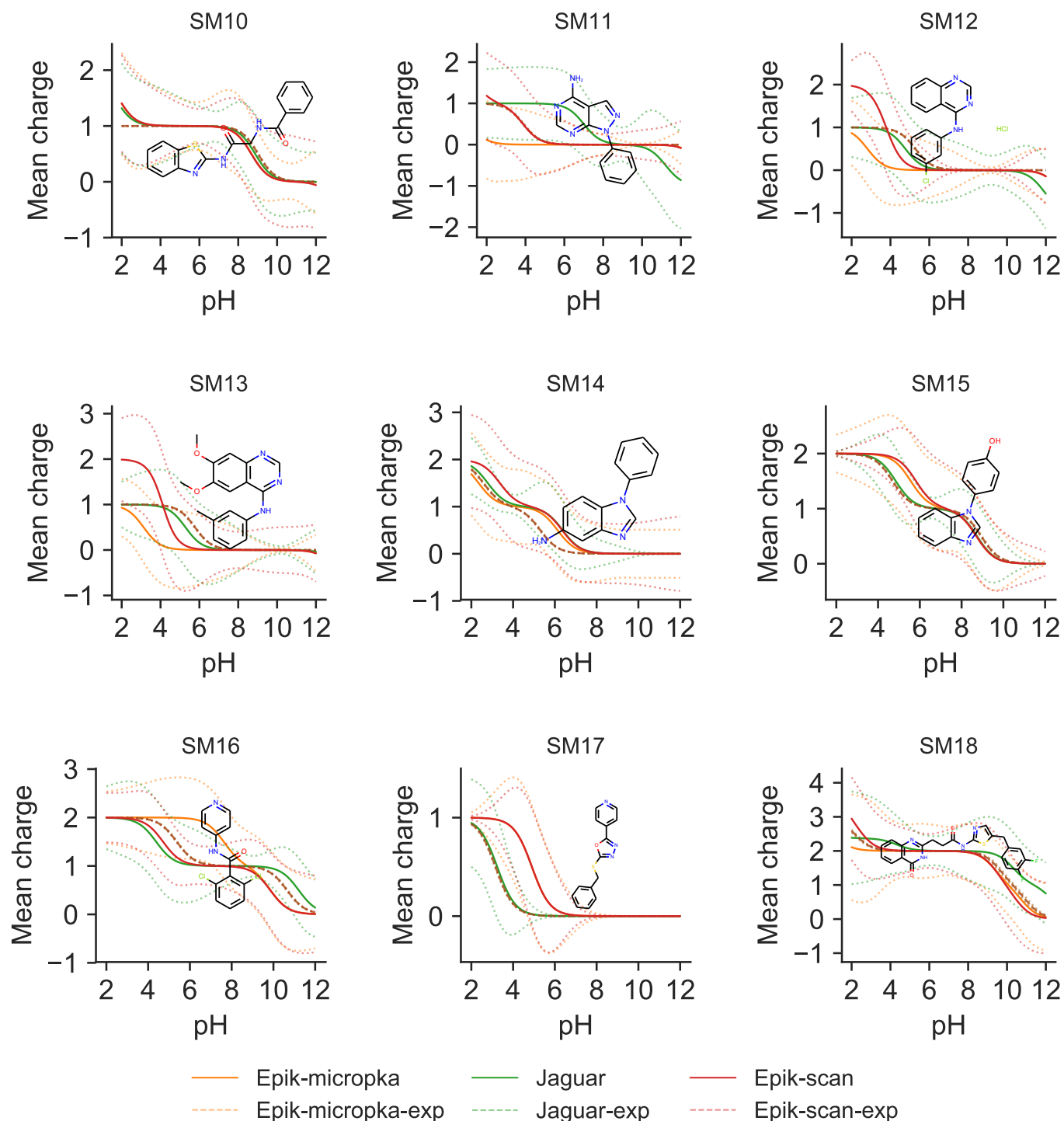


Figure 6. Bootstrap titration curves for each method compared to experiment for molecules SM10-SM18. The titration curve for each method is shown as a solid line, and 96 % confidence intervals from bootstrap have been shown as dotted lines. Since the absolute experimental charge is not available in general, experimental curves (dashed) were aligned to each prediction independently using an integer offset that minimizes the area between curves. A numerical comparison to experiment is presented in table 2.

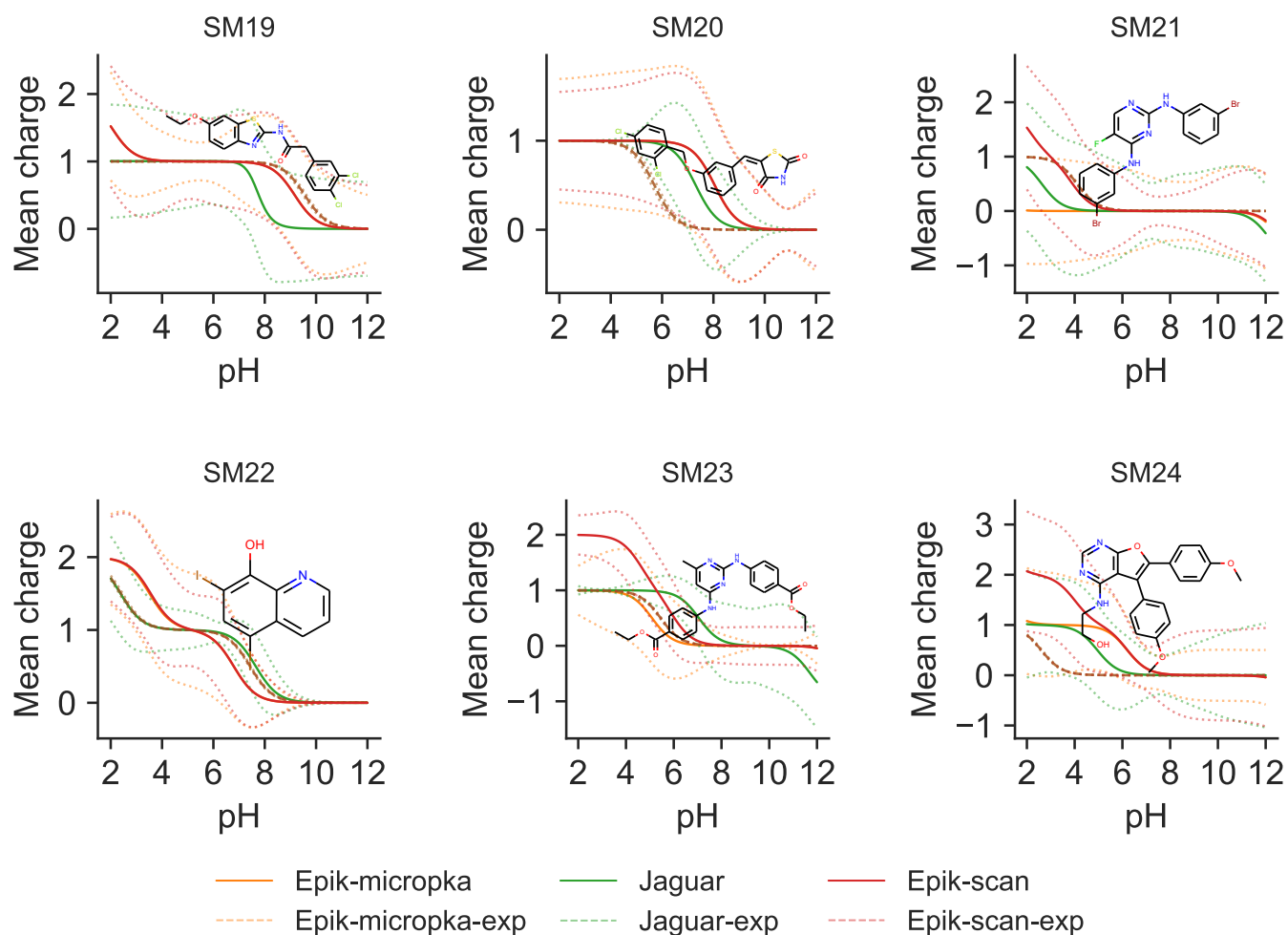


Figure 7. Bootstrap titration curves for each method compared to experiment for molecules SM19-SM24. The titration curve for each method is shown as a solid line, and 96 % confidence intervals from bootstrap have been shown as dotted lines. Since the absolute experimental charge is not available in general, experimental curves (dashed) were aligned to each prediction independently using an integer offset that minimizes the area between curves. A numerical comparison to experiment is presented in table 2.

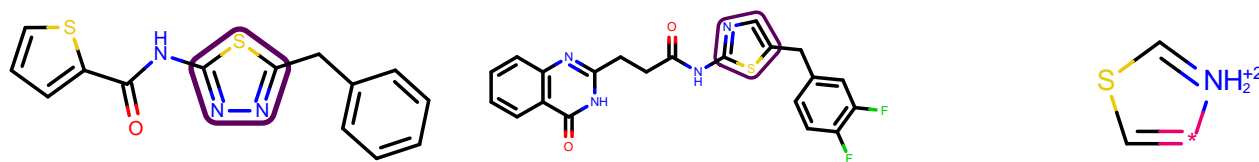


Figure 8. Problematic lewis structure generated by Epik. Epik generated protonation states with invalid Lewis structure for SM03 and SM18, with a pentavalent nitrogen present in the thiazole/thiadiazole (the protonation pattern displayed on the right). One of these was also present in the set provided as part of SAMPL6 microstates, namely SM03_micro021. A valid Lewis structure could not be figured out for this protonation state.

Descriptor analysis

- which moieties are harder to predict?
- Do certain descriptors correlate with variance/absolute errors?
- Do number of rotatable bonds affect epik/jaguar results (conformations missing from prediction)?

Suggestions for future challenges

- Future challenges could benefit from NMR experiments for microstates
- Potentiometric titrations could capture states that may have been left out
- Probabilistic models (i.e. bayesian hierarchical models) could be constructed for the analysis of experiments.

Conclusions

- Epik and Jaguar perform well when matching both pKa and curves
- Microscopic pKa comparison directly to macroscopic may be deceiving, so the titration curve pose an alternative interpretation that can help assess
- Epik has a few artifacts (weird states, state penalties)
- future challenges could benefit from more NMR experiments and potentiometric titration data
- How a future challenge could make this easier

Methods

Epik micropKa predictions

- were performed between pH 2-12
- Reported states with a minimum population of $e^{-10/RT}$
- States proposed by SAMPL6 but not predicted by Epik were not considered in analysis.

Epik scan: Sequential titration of dominant species

Started at pH 7.0, add protons sequentially. Produces the dominant microspecies as a way to mimic the likely macrostate.

Jaguar: Ab initio quantum chemical pKa predictions

- Ran using default settings (5 conformations) for each of the pre-enumerated specified microstate pairs.
- Input structures were first minimized using MMFF.
- Additional minimization was performed for structures for which scf/geopt did not converge

Code and data availability

- Input files and analysis scripts are available at <https://github.com/choderalab/SAMPL6-reference-pka-calculations>

Author Contributions

(Follow the [CRediT Taxonomy](#))

Acknowledgments

ASR, MI, AR, PBG, and JDC acknowledge support from the Sloan Kettering Institute. JDC acknowledges support from NIH grant P30 CA008748.

Disclosures

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

References

- [1] **Aguilar B**, Anandakrishnan R, Ruscio JZ, Onufriev AV. Statistics and Physical Origins of pK and Ionization State Changes upon Protein-Ligand Binding. *Biophys J*. 2010 Mar; 98(5):872–880. doi: [10.1016/j.bpj.2009.11.016](#).
- [2] **Jordan IK**, Kondrashov FA, Adzhubei IA, Wolf YI, Koonin EV, Kondrashov AS, Sunyaev S. A Universal Trend of Amino Acid Gain and Loss in Protein Evolution. *Nature*. 2005 Feb; 433(7026):633–638. doi: [10.1038/nature03306](#).
- [3] **Martin YC**. Let's not forget tautomers. *Journal of computer-aided molecular design*. 2009 Oct; 23:693–704. doi: [10.1007/s10822-009-9303-2](#).
- [4] **Greenwood JR**, Calkins D, Sullivan AP, Shelley JC. Towards the comprehensive, rapid, and accurate prediction of the favorable tautomeric states of drug-like molecules in aqueous solution. *Journal of computer-aided molecular design*. 2010 Jun; 24:591–604. doi: [10.1007/s10822-010-9349-1](#).
- [5] **Bax B**, Chung CW, Edge C. Getting the chemistry right: protonation, tautomers and the importance of H atoms in biological chemistry. *Acta crystallographica Section D, Structural biology*. 2017 Feb; 73:131–140. doi: [10.1107/S2059798316020283](#).
- [6] **Pickard FC**, König G, Tofoleanu F, Lee J, Simmonett AC, Shao Y, Ponder JW, Brooks BR. Blind Prediction of Distribution in the SAMPL5 Challenge with QM Based Protomer and pK_a Corrections. *J Comput Aided Mol Des*. 2016 Nov; 30(11):1087–1100. doi: [10.1007/s10822-016-9955-7](#).
- [7] **Shelley JC**, Cholleti A, Frye LL, Greenwood JR, Timlin MR, Uchimaya M. Epik: A Software Program for pK_a Prediction and Protonation State Generation for Drug-like Molecules. *J Comput Aided Mol Des*. 2007 Dec; 21(12):681–691. doi: [10.1007/s10822-007-9133-z](#).
- [8] **Bochevarov AD**, Harder E, Hughes TF, Greenwood JR, Braden DA, Philipp DM, Rinaldo D, Halls MD, Zhang J, Friesner RA. Jaguar: A High-Performance Quantum Chemistry Software Program with Strengths in Life and Materials Sciences. *Int J Quantum Chem*. 2013 Sep; 113(18):2110–2142. doi: [10.1002/qua.24481](#).
- [9] **Perrin DD**, Dempsey B, Serjeant EP. pK_a prediction for organic acids and bases, vol. 1. Springer; 1981.
- [10] SMARTS, SMiles ARbitrary Target Specification, is a registered trademark of Daylight Chemical Information Systems;.
- [11] **Lim C**, Bashford D, Karplus M. Absolute pK_a calculations with continuum dielectric methods. *The Journal of Physical Chemistry*. 1991; 95(14):5610–5620.
- [12] **Philipp DM**, Watson MA, Yu HS, Steinbrecher TB, Bochevarov AD. Quantum chemical prediction for complex organic molecules. *International Journal of Quantum Chemistry*. 2018; 118(12):e25561.

Supplementary Information

- Titration curves and free energy plots for each compound, by each method
- The mean charge/deviation curves for each compound and each method
- .mae and .sdf files with results
- scripts and or jupyter notebooks for analysis
- csv version of tables

Result: Mapping of each experimental pKa i to predicted pKa j

C is constructed, where every row i is an experiment, and every column j a prediction;

$C_{ij} = \text{cost}(\text{pKa}_{\text{exp},i}, \text{pKa}_{\text{pred},j});$

while $C.\text{size} > 0$ **do**

$k, l = \arg \min(C_{ij});$

 assign experimental value k to prediction l ;

 remove row k , column l from C ;

end

remaining values are unmatched;

Algorithm 1: This algorithm matches experiment with prediction based on how close each value is, one pKa value at a time. Unless the matrix C is square, some values will be unmatched. Those leftover pKa values are returned at the end. It uses a cost function, such as root mean square deviation, to assess how close two values are.

Result: Mapping of each experimental pKa i to predicted pKa j

I = sorted experimental pKa values ;

J = sorted predicted pKa values ;

length = $\max(I.\text{size}, J.\text{size})$;

Append placeholders such that $I.\text{size} = \text{length}$;

Append placeholders such that $J.\text{size} = \text{length}$;

min = ∞ ;

solution = J rolled 0 times;

for n in $0..\text{length}$ **do**

$S = J$ rolled n times ;

 total = 0.0;

for m in $0..\text{length}$ **do**

if I_m is placeholder **then** unmatched experiment

if $J_m \leq 7.0$ **then**

$I_m = 0.0$

else

$I_m = 14.0$

end

else if J_m is placeholder **then** unmatched prediction

if $I_m \leq 7.0$ **then**

$J_m = 0.0$

else

$J_m = 14.0$

end

end

 total = total + $\text{cost}(I_m, J_m)$;

end

if total < min **then** solution is better

 min = total;

 solution = S ;

end

end

Algorithm 2: Sequential pKa mapping. It uses a cost function to measure cost, and it rolls (shifts by one, and reintroduces last element as first). Any unmatched pKa values are represented by matching with a placeholder value. To calculate the cost, the placeholder is replaced by either 0, or 14, depending on whether the unmatched value is above or below 7.0.