SAMPL6 pK_a Challenge: Predictions of ionization constants performed by the S+pKa method implemented in ADMET Predictor™ software

Robert Fraczkiewicz and Marvin Waldman*

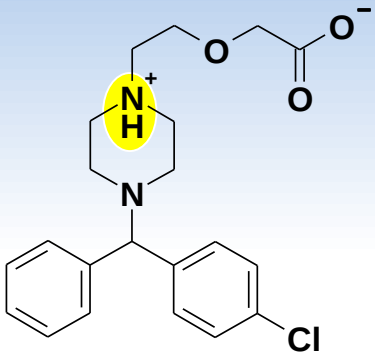
Simulations Plus, Inc.

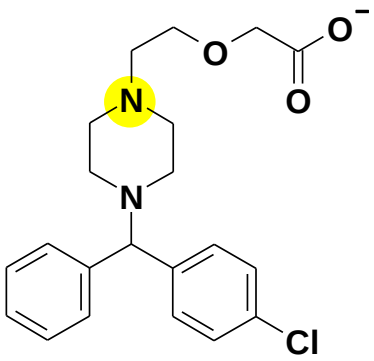
February 2018

SimulationsPlus
SCIENCE + SOFTWARE = SUCCESS

Simplified overview of pK_a modeling

Ionizable atom in a microstate





2D Atomic Descriptors for ionizable atom in its molecular environment

A1
A2
A3
A4
A5
A6
A7
A8
A9
A10
A11
A12
A13
A14
A15

optimal subset from ~130 atomic descriptors

QSPR

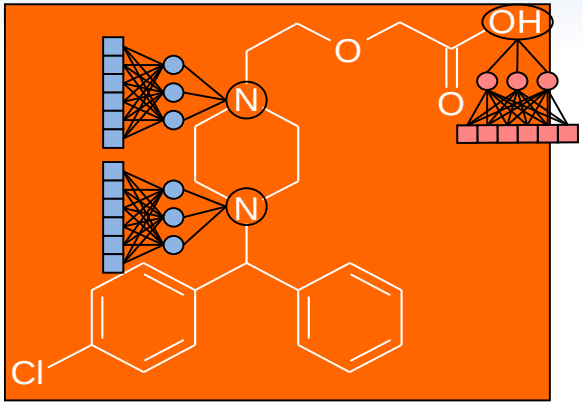
Predicted micro pk_a

SimulationsPlus
SCIENCE + SOFTWARE = SUCCESS

The predictive model, S+pKa

- 10 Artificial Neural Network Ensembles (ANNE);
  one ANNE for each of the following 10 classes of ionizable atoms:
  - (1) Hydroxyacids
  - (2) Acidic amides
  - (3) Acids of aromatic NH
  - (4) Thioacids
  - (5) Carboacids
  - (6) Amines
  - (7) Bases of aromatic N
  - (8) N-oxides
  - (9) Thiones
  - (10) Carbobases (protonatable C in some π–excessive rings)

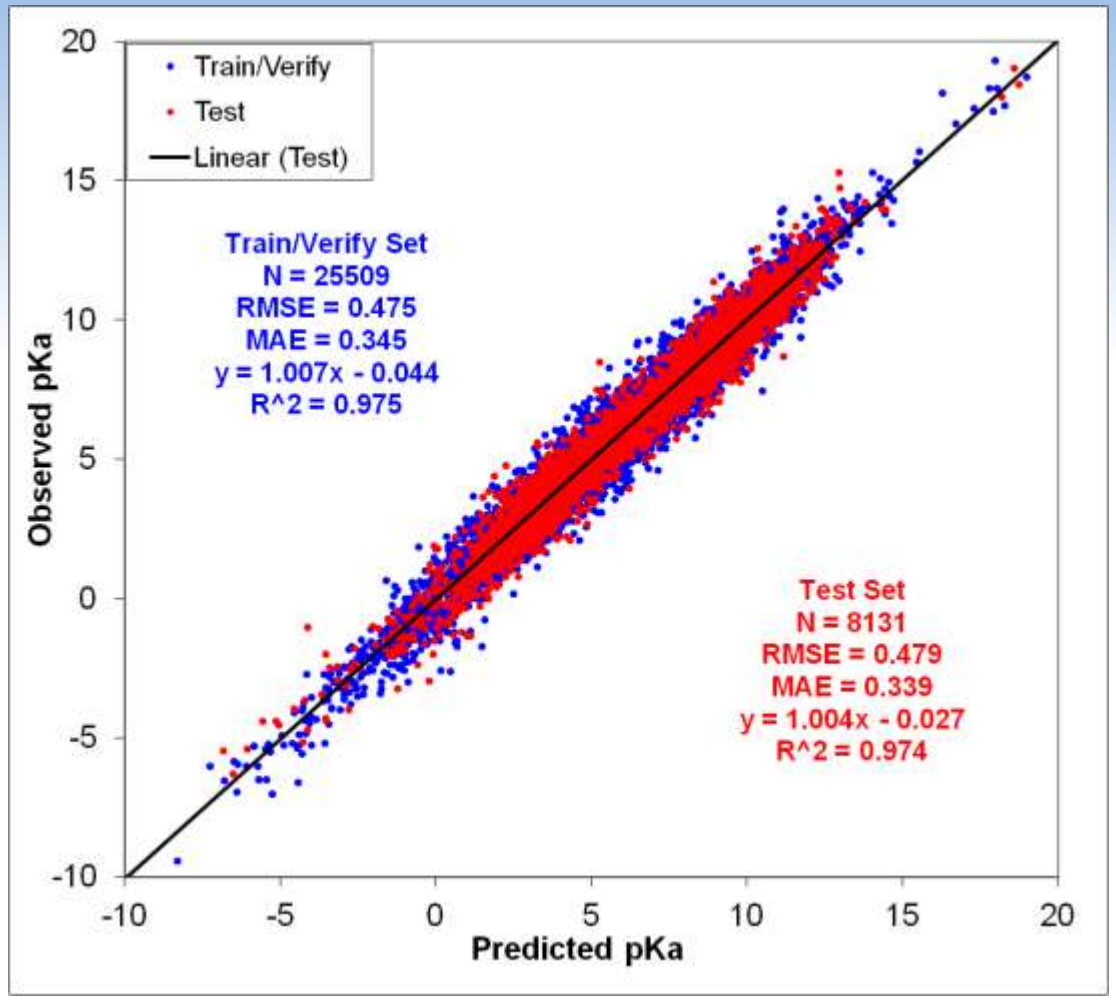
The main factor determining an atom's ionization is its type, followed by its local molecular environment



- ANNEs use localized atomic descriptors as inputs
- ANNEs predict ionization microconstants (micro pK_a)
- Macroconstants calculated with microequilibria theory

Fraczkiewicz, R.; Lobell, M.; Göller, A. H.; Krenz, U.; Schoenneis, R.; Clark, R. D.; Hillisch, A. "Best of Both Worlds: Combining Pharma Data and State of the Art Modeling Technology To Improve in Silico pKa Prediction." *Journal of Chemical Information and Modeling* **2014**, *55*, 389-397.

Model training and initial testing with internal test set*



Details of data split:

Training pool: 25509 pK_a values

- 10594 from public sources
- 14915 from Bayer Pharma

Internal test set: 8131 pK_a values

- 3582 from public sources
- 4549 from Bayer Pharma

Internal test set prediction statistics:

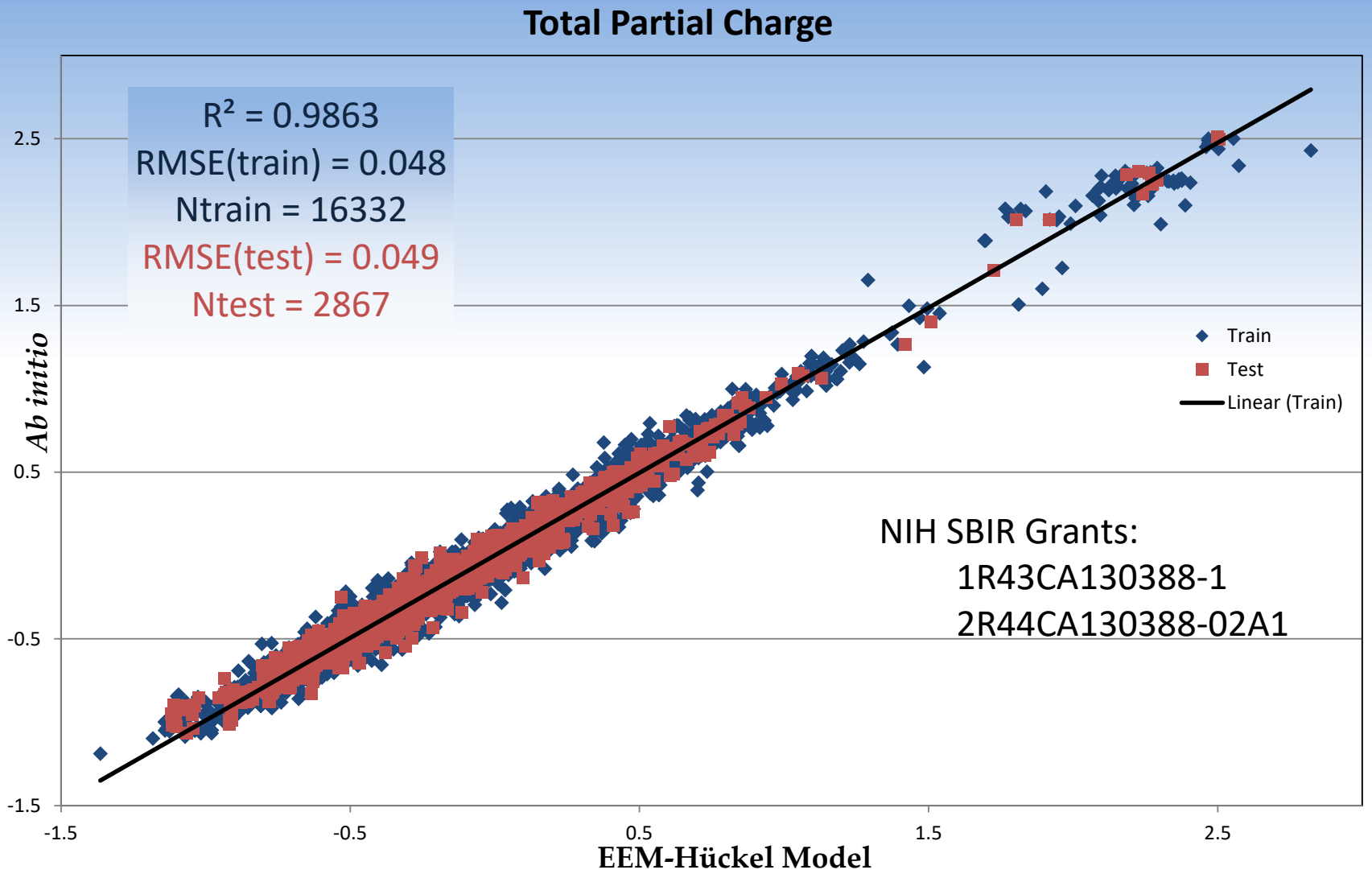
MAE = 0.34

RMSE = 0.48

R^2 = 0.97

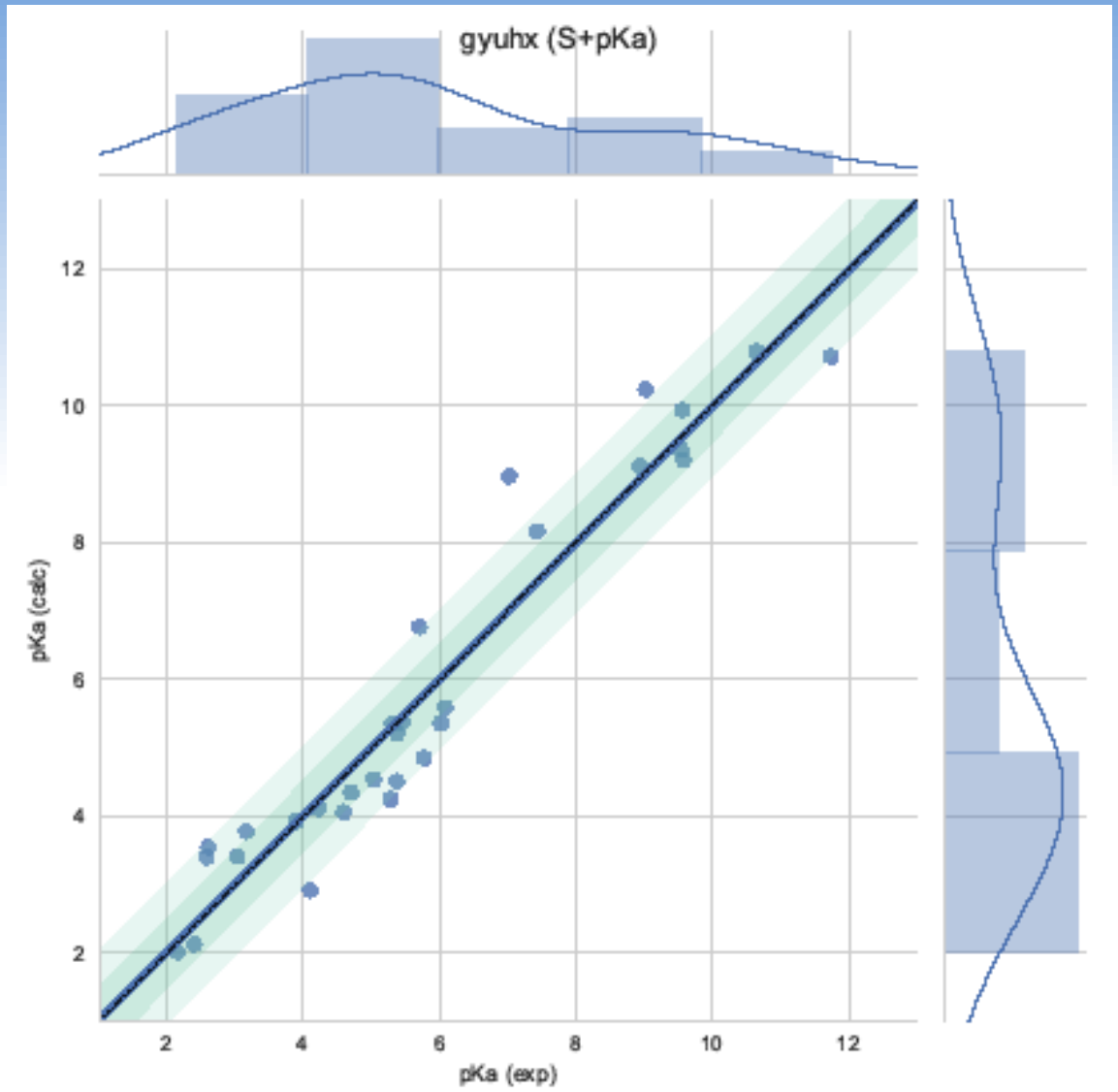
* Internal test set compounds have not been used for model training but have been used to select the 10 ANNEs to appear in the final model

Atomic Partial Charge Descriptors



SimulationsPlus
SCIENCE + SOFTWARE = SUCCESS

SAMPL6 pK_a challenge results



RMSE = 0.73

MAE = 0.579

ME = -0.009

R^2 = 0.925

Slope = 0.929

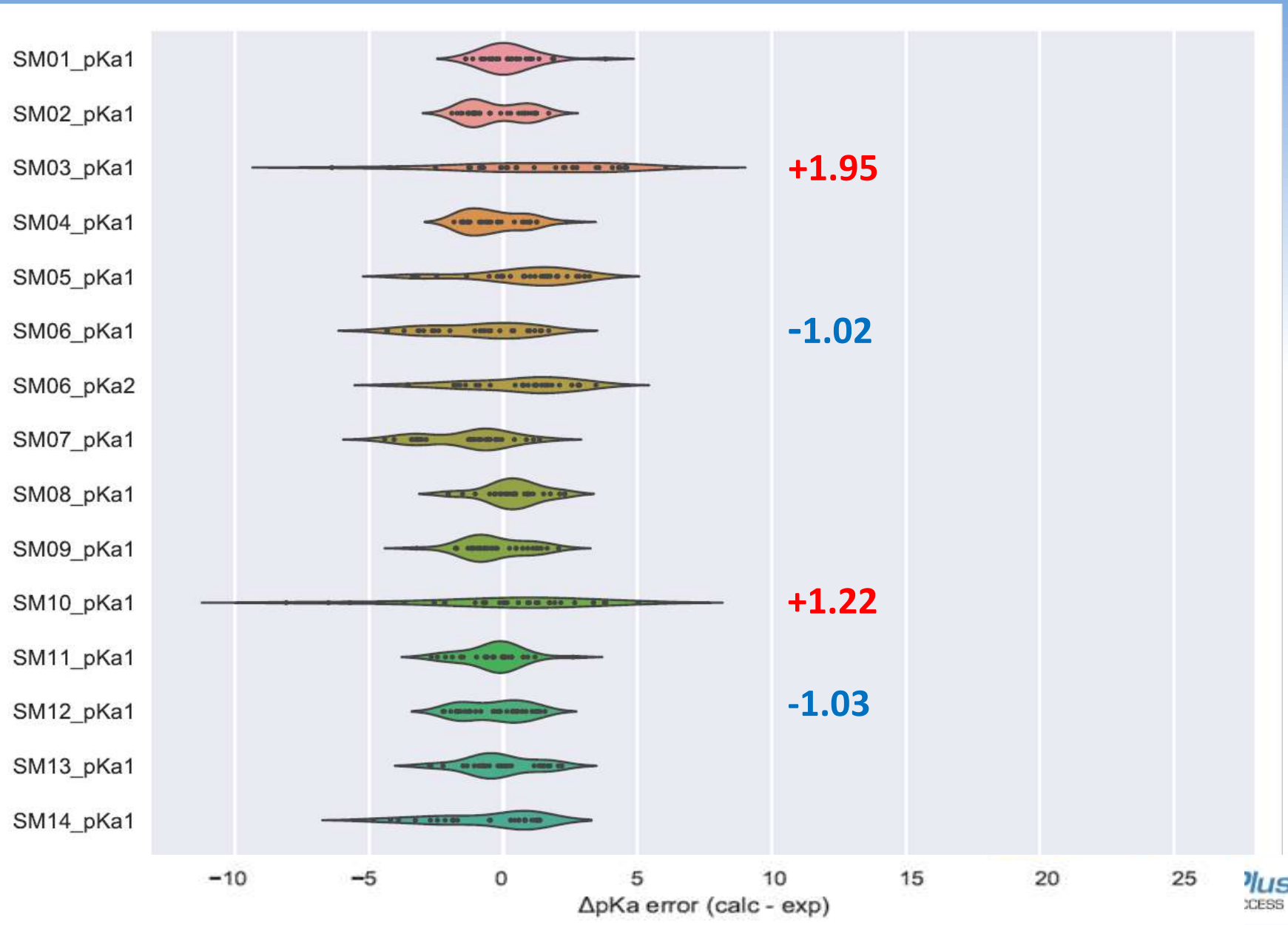
Computation time (wall clock) for all pK_a in all 24 compounds:

Under 2 seconds*

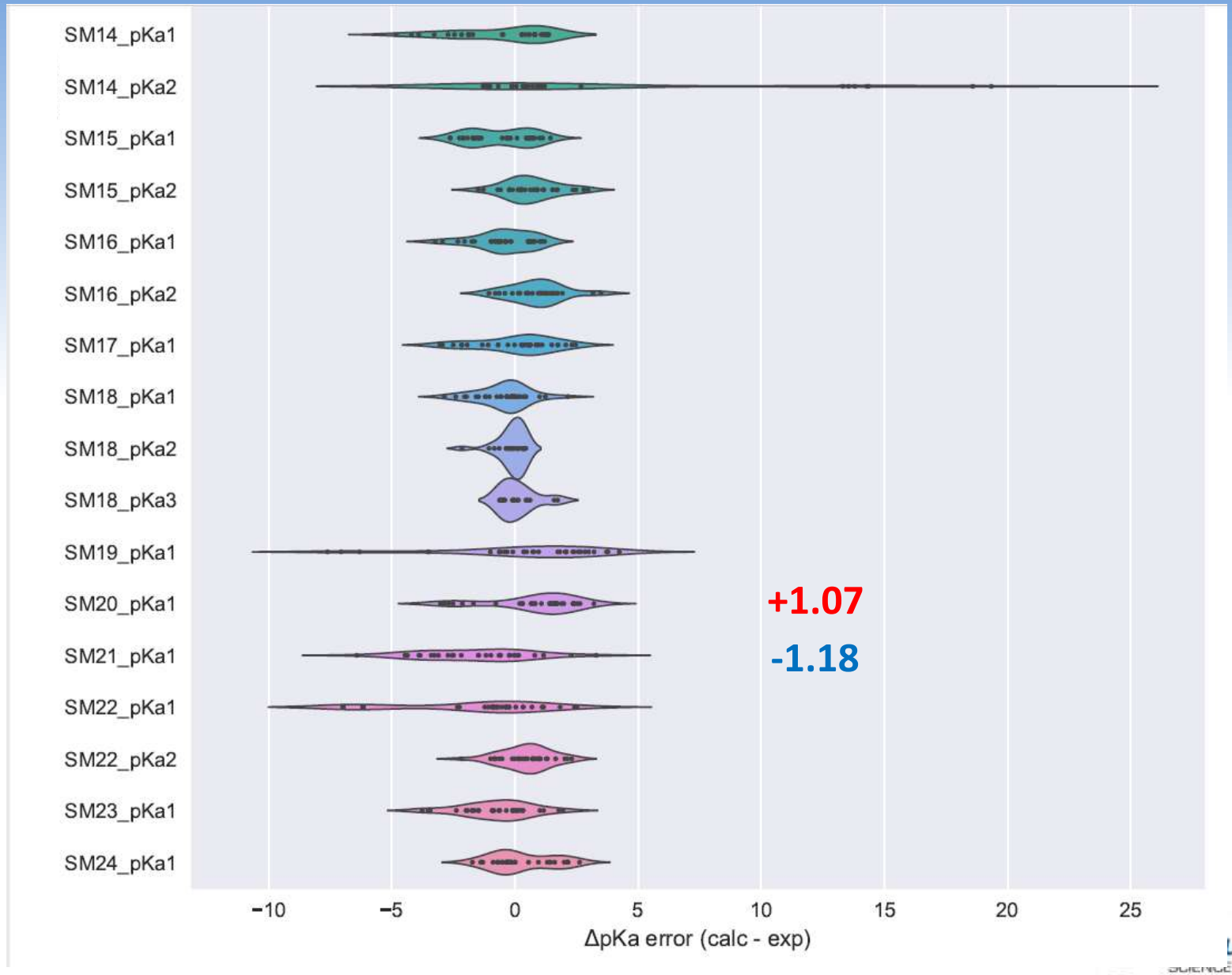
* This includes 145 other properties and ~400 molecular descriptors

SimulationsPlus
SCIENCE + SOFTWARE = SUCCESS

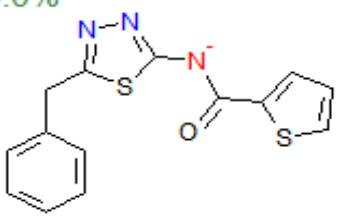
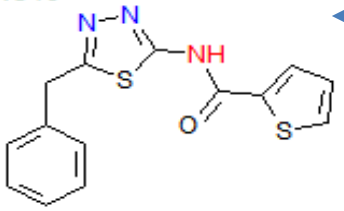
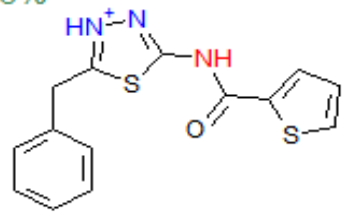
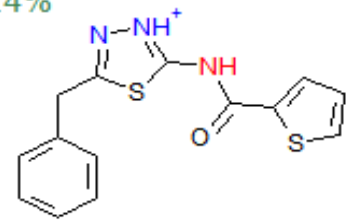
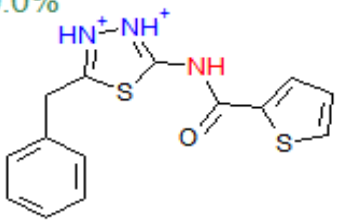
Deviations above 1 log unit



Deviations above 1 log unit



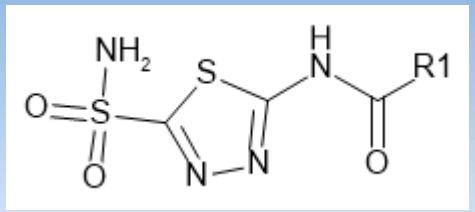
SM03 prediction

pKa	Macrostates	Microstates
	M^-	100.0% 
8.97		
	HM	100.0% 
0.20		
	H_2M^+	58.6%  41.4% 
-2.65		
	H_3M^{+2}	100.0% 

Experimental pKa = 7.02
UV-metric psKa,
with Yasuda-Shedlovsky
extrapolation

SimulationsPlus
SCIENCE + SOFTWARE = SUCCESS

SM03 prediction- Training set analogs



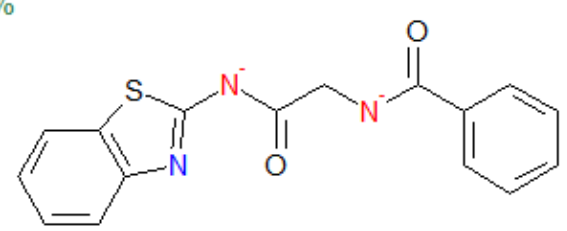
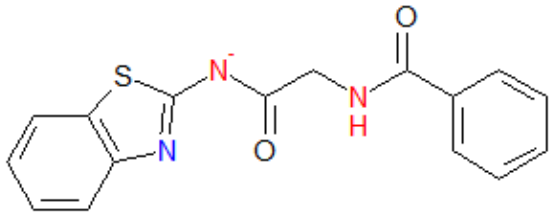
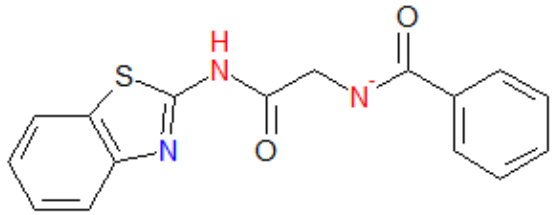
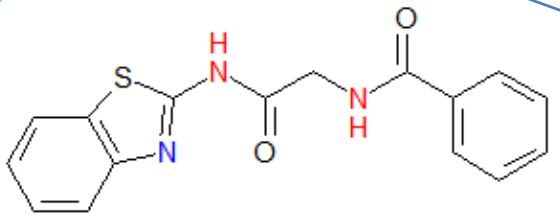
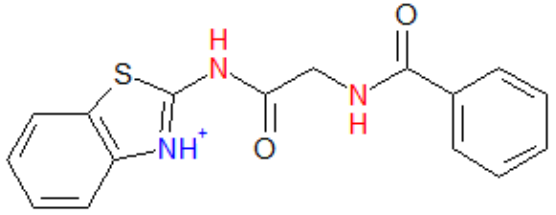
All analogs in training set contain sulfonamide group and have reported pKa's ranging from 4.8-9.3

pKa	Macrostates	Microstates
	M^{-2}	100.0%
9.27		
	HM^{-}	57.2% / 42.8%
8.19		
	H_2M	100.0%
-1.92		
	H_3M^{+}	99.3%
-6.19		
	H_4M^{+2}	100.0%

Example:
First acidic dissociation
Is a mixture of two
Microstates.

SimulationsPlus
SCIENCE + SOFTWARE = SUCCESS

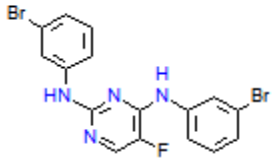
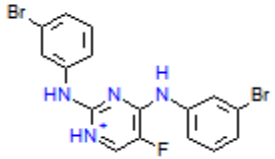
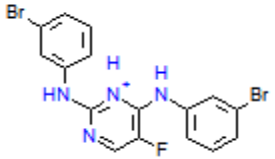
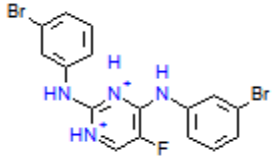
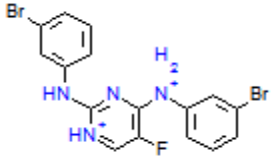
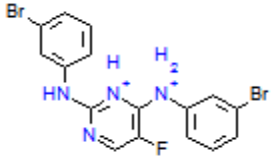
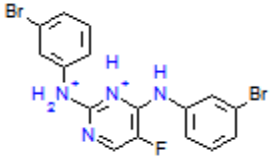
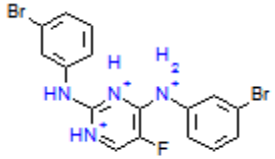
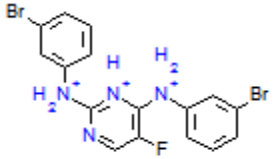
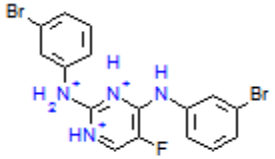
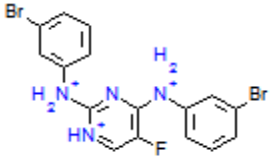
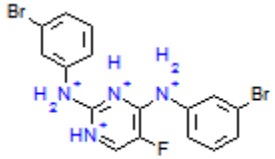
SM10 prediction

pKa	Macrostates	Microstates
	M^{-2}	100.0% 
11.31		
	HM^{-}	89.9%  10.1% 
10.24		
	H_2M	100.0% 
1.79		
	H_3M^{+}	100.0% 

Experimental pKa = 9.02
UV-metric psKa,
with Yasuda-Shedlovsky
extrapolation

SimulationsPlus
SCIENCE + SOFTWARE = SUCCESS

SM21 prediction

pKa	Macrostates	Microstates
	M	100.0% 
2.92		
	HM^{+}	86.2%  13.6% 
-0.43		
	H_2M^{+2}	47.8%  46.1%  3.5%  1.7% 
-1.69		
	H_3M^{+3}	72.4%  11.7%  8.6%  7.3% 
-2.83		
	H_4M^{+4}	100.0% 

Experimental pKa = 4.1
UV-metric psKa,
with Yasuda-Shedlovsky
extrapolation

SimulationsPlus
SCIENCE + SOFTWARE = SUCCESS