SAMPL6 pKₐ 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
Simplified overview of pK$_a$ modeling

Ionizable atom in a microstate

2D Atomic Descriptors for ionizable atom in its molecular environment

<table>
<thead>
<tr>
<th>A1</th>
<th>A2</th>
<th>A3</th>
<th>A4</th>
<th>A5</th>
<th>A6</th>
<th>A7</th>
<th>A8</th>
<th>A9</th>
<th>A10</th>
<th>A11</th>
<th>A12</th>
<th>A13</th>
<th>A14</th>
<th>A15</th>
</tr>
</thead>
</table>

optimal subset from ~130 atomic descriptors

QSPR

Predicted micro pK$_a$
The predictive model, $S+pK_a$

- 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)

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

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

Details of data split:
Training pool: 25509 pKₐ values
  • 10594 from public sources
  • 14915 from Bayer Pharma
Internal test set: 8131 pKₐ values
  • 3582 from public sources
  • 4549 from Bayer Pharma

Internal test set prediction statistics:
MAE = 0.34
RMSE = 0.48
R² = 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

Total Partial Charge

\[ R^2 = 0.9863 \]
\[ \text{RMSE(train)} = 0.048 \]
\[ N_{\text{train}} = 16332 \]
\[ \text{RMSE(test)} = 0.049 \]
\[ N_{\text{test}} = 2867 \]

Ab initio vs. EEM-Hückel Model

NIH SBIR Grants:
1R43CA130388-1
2R44CA130388-02A1
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
Deviations above 1 log unit

\[
\begin{align*}
\text{SM01_pKa1} & \quad +1.95 \\
\text{SM02_pKa1} & \\
\text{SM03_pKa1} & \\
\text{SM04_pKa1} & \\
\text{SM05_pKa1} & \\
\text{SM06_pKa1} & \quad -1.02 \\
\text{SM06_pKa2} & \\
\text{SM07_pKa1} & \\
\text{SM08_pKa1} & \\
\text{SM09_pKa1} & \\
\text{SM10_pKa1} & \quad +1.22 \\
\text{SM11_pKa1} & \\
\text{SM12_pKa1} & \quad -1.03 \\
\text{SM13_pKa1} & \\
\text{SM14_pKa1} & 
\end{align*}
\]
Deviations above 1 log unit

\[
\begin{align*}
\text{SM14\_pKa1} & \quad +1.07 \\
\text{SM14\_pKa2} & \\
\text{SM15\_pKa1} & \\
\text{SM15\_pKa2} & \\
\text{SM16\_pKa1} & \\
\text{SM16\_pKa2} & \\
\text{SM17\_pKa1} & \\
\text{SM18\_pKa1} & \\
\text{SM18\_pKa2} & \\
\text{SM18\_pKa3} & \\
\text{SM19\_pKa1} & \\
\text{SM20\_pKa1} & -1.18 \\
\text{SM21\_pKa1} & \\
\text{SM22\_pKa1} & \\
\text{SM22\_pKa2} & \\
\text{SM23\_pKa1} & \\
\text{SM24\_pKa1} & \\
\end{align*}
\]

\(\Delta pK_a\) error (calc - exp)
## SM03 prediction

<table>
<thead>
<tr>
<th>pKₐ</th>
<th>Macrostates</th>
<th>Microstates</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.97</td>
<td>M⁻</td>
<td><img src="image1" alt="Chemical Structure" /> 100.0%</td>
</tr>
<tr>
<td></td>
<td>HM</td>
<td><img src="image2" alt="Chemical Structure" /> 100.0%</td>
</tr>
<tr>
<td>0.20</td>
<td>H₂M⁺</td>
<td><img src="image3" alt="Chemical Structure" /> 58.6% 41.4%</td>
</tr>
<tr>
<td>-2.65</td>
<td>H₂M⁺²</td>
<td><img src="image4" alt="Chemical Structure" /> 100.0%</td>
</tr>
</tbody>
</table>

Experimental pKₐ = 7.02 UV-metric psKₐ, with Yasuda-Shedlovsky extrapolation
SM03 prediction- Training set analogs

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

Example:
First acidic dissociation is a mixture of two microstates.
SM10 prediction

<table>
<thead>
<tr>
<th>pKₐ</th>
<th>Macrostates</th>
<th>Microstates</th>
</tr>
</thead>
<tbody>
<tr>
<td>11.31</td>
<td>M²</td>
<td><img src="image1.png" alt="Structure" /> 100.0%</td>
</tr>
<tr>
<td></td>
<td>HM⁻</td>
<td><img src="image2.png" alt="Structure" /> 89.9%</td>
</tr>
<tr>
<td>10.24</td>
<td>H₂M</td>
<td><img src="image3.png" alt="Structure" /> 100.0%</td>
</tr>
<tr>
<td>1.79</td>
<td>H₃M⁺</td>
<td><img src="image4.png" alt="Structure" /> 100.0%</td>
</tr>
</tbody>
</table>

Experimental pKa = 9.02
UV-metric psKa, with Yasuda-Shedlovsky extrapolation
SM21 prediction

<table>
<thead>
<tr>
<th>pKₐ</th>
<th>Macrostates</th>
<th>Microstates</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.92</td>
<td>M</td>
<td>![Image1]</td>
</tr>
<tr>
<td></td>
<td>HM⁺</td>
<td>![Image2]</td>
</tr>
<tr>
<td>-0.43</td>
<td>H₂M⁺²</td>
<td>![Image3]</td>
</tr>
<tr>
<td></td>
<td>H₃M⁺³</td>
<td>![Image4]</td>
</tr>
<tr>
<td>-1.69</td>
<td>H₄M⁺⁴</td>
<td>![Image5]</td>
</tr>
</tbody>
</table>

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