Estimation of Ionization Constants (pKₐ) for Drug-Like Compounds

Stephen Jelfs, Peter Ertl and Paul Selzer
Novartis Institutes for BioMedical Research, Basel, Switzerland

Abstract

A knowledge of the ionization constants (pKₐ) of compounds is important for much of the work carried out in the drug discovery process. These constants can have a profound affect on the physicochemical properties of a compound and, in rational drug discovery, are essential for the optimization of ADME characteristics. Notably, compounds in their unionized form tend to be less soluble but can more easily penetrate lipophilic barriers existing between them and a biological target of interest. Furthermore, the correct ionization state of a compound involved in ligand-receptor binding is required prior to docking studies and for the development of reliable SAR. We have therefore developed a pragmatic approach for the estimation of pKₐ. Importantly, in-house computational resources and experimental data were exploited to provide reliable predictions for drug-like compounds via the Novartis intranet.

Methodology

The approach utilizes a stepwise algorithm to assign pKₐ values to the acidic and basic groups in a compound. At each step, the algorithm finds the next most basic (least acidic) group based on a consensus of predictive models. These models were derived using descriptors including 2D molecular tree structured fingerprints and semi-empirical quantum chemical (QC) properties. The molecular tree derived for phenol, for example, comprises of an integer-vector encoding the frequency of atom-types and bond-distance (bonds), d = 0 1 2 3 4.

atom types, I = O,3, C=ar, C=ar, C=ar, C=ar

fingerprints (I,d) = 1 1 2 2 1

The QC properties included the partial charge and electrophilic superdelocalizability (SE) of the atoms undergoing (de)protonation. For multiprotic compounds, the absolute SE values were replaced by the SE of the neutral structure followed by a series of relative values calculated with combinations of the more basic/acidic groups appropriately ionized. This procedure overcame limitations highlighted in the literature to allow multiprotic, drug-like compounds to be modeled effectively.

Web-Application

The resultant performance of the approach has enabled the development of a web-application for pKₐ prediction (available on the Novartis intranet). The tool can be used to identify important acidic and basic groups within structures, such as the carboxylic acid and guanidine groups in Zanamivir:

The species distribution plot is also provided, indicating the major chemical species present in aqueous solution at varying pH-levels. For instance, Zanamivir is shown to be present in its zwitterion form under typical physiological conditions:

Further information about the web-application is provided via a Wiki discussion forum which summarizes the current status of the predictive models and highlights problematic structures, such as tautomers, that may be encountered. We have also recently extended the tool to include pH-dependent distribution constants (logD).

Results

Predictive models were successfully derived using partial least-squares (PLS), and validated using 7-fold cross-validation, for a variety of ionizable groups:

<table>
<thead>
<tr>
<th>Model</th>
<th>N</th>
<th>R²</th>
<th>RMSE</th>
<th>R² incr.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alcohol+</td>
<td>246</td>
<td>3</td>
<td>0.91</td>
<td>0.58</td>
</tr>
<tr>
<td>Amines</td>
<td>1453</td>
<td>4</td>
<td>0.88</td>
<td>0.52</td>
</tr>
<tr>
<td>Carboxylic Acids</td>
<td>732</td>
<td>4</td>
<td>0.90</td>
<td>0.33</td>
</tr>
<tr>
<td>Amines</td>
<td>134</td>
<td>4</td>
<td>0.97</td>
<td>0.67</td>
</tr>
<tr>
<td>Pyridines</td>
<td>447</td>
<td>4</td>
<td>0.95</td>
<td>0.58</td>
</tr>
<tr>
<td>Pyrimidines</td>
<td>141</td>
<td>4</td>
<td>0.94</td>
<td>0.46</td>
</tr>
<tr>
<td>Combined</td>
<td>3514</td>
<td>6.51</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

As mentioned, the predictive models are also applicable to multiprotic compounds:

![Species involved in ligand-receptor binding](image)

Where, the charged carboxylic acid and guanidine groups are essential for binding-site recognition, forming important interactions with the complementary arginine (ca. pKₐ 13.0) and glutamic (ca. pKₐ 4.3) amino acids, respectively.

Prior to docking studies, it is common to enumerate all potential chemical species that could possibly be involved in binding (ca. pH 7). We are therefore looking to improve the performance of our approach for the pre-treatment of compounds prior to high-throughput docking. The adverse affects of using the incorrect protonation state or tautomer form of compounds in docking studies may also form the bases of future work.

Acknowledgments

The authors would like to thank Joerg Muehlaber and Wolfgang Zipfel for their technical support, Bernard Faller and Frederique Louidet for providing experimental data, and Ansgar Schuffenhauser, John Priestle, and Bernhard Rohde for their insightful discussions. We are also grateful to all the medicinal chemists involved in testing the beta-version of the web-application.

References