Predicting Sites of Metabolism with Artificial Neural Network Ensembles

Marvin Waldman, Robert Fraczkiewicz, Jinhua Zhang, Robert D. Clark, and Walter S. Woltosz

Simulations Plus, Inc., 42505 10th Street West, Lancaster, CA 93534, USA (www.simulations-plus.com)

Introduction

Hepatic first-pass metabolism of many drugs and prodrugs plays a key role in their oral bioavailability. The human cytochrome P450 enzymes are responsible for the metabolism of most drugs. Knowledge of likely sites of metabolic attack in a drug molecule can aid in designing out unwanted metabolic liabilities early on in the drug discovery process, as well as in the design of prodrugs where metabolic transformation is desired. Using datasets constructed from literature compilations and the industry’s largest commercially available metabolite database, we have constructed models based on artificial neural network ensembles that predict likely sites of metabolism for several CYP isoforms, including 2C9, 2D6, and 3A4.

Sources of Experimental Data

Extensively curated compilations:
• Accelrys Metabolite™ Database (formerly Symyx).
• Data published by Sheridan et al. [1].
• Review of literature – both old and new.

Model Building Process

Set of substrates for a given CYP
Generate molecular descriptors
Generate Kohonen Map
Training set of Molecules
Filter non-candidate atoms
Generate atomic descriptors
Build ANNC models
External Test set of Molecules
Predict
Select Best model

Descriptive Generation

Descriptors generated by ADMET Predictor™ v5.0:
• Partial Atomic Charges and Reactivities
• EEM-Hückel charge model (accounts for resonance effects)
• Parametrized from our ab initio database of partial atomic \( \sigma \) and \( \pi \) charges
• EEM (Electronegativity Equalization Method) for \( \sigma \) charges
• Reactivities
• EEM \( \sigma \) atomic Fukui indices
• Hückel \( \sigma \) frontier orbital atomic densities
• E-State indices
• Local shape descriptors
• Sheridan’s SPAN
• Atomic volumes
• Others
• Polarizability, electronegativity, autocorrelation vectors
• Specialty proprietary

Performance of Final Models

<table>
<thead>
<tr>
<th>Model</th>
<th>Top 1</th>
<th>Top 2</th>
<th>Top 3</th>
<th>Top 4</th>
<th>Top 5</th>
<th>Top 6</th>
<th>Top 7</th>
<th>Top 8</th>
<th>Top 9</th>
<th>Top 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>1A2</td>
<td>74.8%</td>
<td>91.5%</td>
<td>94.8%</td>
<td>97.4%</td>
<td>97.6%</td>
<td>97.9%</td>
<td>98.1%</td>
<td>98.6%</td>
<td>98.9%</td>
<td>99.1%</td>
</tr>
<tr>
<td>2C9</td>
<td>74.8%</td>
<td>91.5%</td>
<td>94.8%</td>
<td>97.4%</td>
<td>97.6%</td>
<td>97.9%</td>
<td>98.1%</td>
<td>98.6%</td>
<td>98.9%</td>
<td>99.1%</td>
</tr>
<tr>
<td>2D6</td>
<td>74.8%</td>
<td>91.5%</td>
<td>94.8%</td>
<td>97.4%</td>
<td>97.6%</td>
<td>97.9%</td>
<td>98.1%</td>
<td>98.6%</td>
<td>98.9%</td>
<td>99.1%</td>
</tr>
<tr>
<td>3A4</td>
<td>74.8%</td>
<td>91.5%</td>
<td>94.8%</td>
<td>97.4%</td>
<td>97.6%</td>
<td>97.9%</td>
<td>98.1%</td>
<td>98.6%</td>
<td>98.9%</td>
<td>99.1%</td>
</tr>
</tbody>
</table>

“False” positives in 2009 became true positives in 2010!

“False” positives in 2009 became true positives in 2010!

References


ACKNOWLEDGEMENTS

This work has been funded by the SBIR Grant 5R44CA130388-03 from the National Institute of Health.