EXPLORING THE SCORING FUNCTION SPACE FOR STRUCTURE-BASED DRUG DESIGN

XXVIII Symposium on Bioinformatics and computer-aided drug discovery

Gabriela Bitencourt Ferreira
Pontifical Catholic University of Rio Grande do Sul, Brazil
MOLECULAR DOCKING

+ Scoring Function =

[Diagram showing molecular docking and scoring function]
MOLECULAR DOCKING

Classical Scoring Function
Table 1. Protein and Ligand Data Set Details

<table>
<thead>
<tr>
<th>protein</th>
<th>target type</th>
<th>no. of ligands</th>
<th>no. of ligand classes</th>
<th>no. of cocrystals</th>
<th>max affinity (nM)</th>
<th>min affinity (nM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chk1</td>
<td>kinase</td>
<td>193</td>
<td>2</td>
<td>15</td>
<td>7</td>
<td>&gt;10000</td>
</tr>
<tr>
<td>factorXa</td>
<td>serine protease</td>
<td>218</td>
<td>4</td>
<td>10</td>
<td>&lt;1</td>
<td>5000</td>
</tr>
<tr>
<td>gyrase B</td>
<td>isomerase</td>
<td>138</td>
<td>3</td>
<td>7</td>
<td>4</td>
<td>&gt;10000</td>
</tr>
<tr>
<td>HCV polymerase</td>
<td>polymerase</td>
<td>205</td>
<td>2</td>
<td>13</td>
<td>5.6</td>
<td>&gt;10000</td>
</tr>
<tr>
<td>Met tRNA synthetase</td>
<td>synthetase</td>
<td>144</td>
<td>2</td>
<td>31</td>
<td>1</td>
<td>&gt;10000</td>
</tr>
<tr>
<td>E. coli PDF</td>
<td>metalloprotease</td>
<td>199</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>&gt;10000</td>
</tr>
<tr>
<td>Strep PDF</td>
<td>metalloprotease</td>
<td>186</td>
<td>3</td>
<td>4</td>
<td>&lt;2</td>
<td>&gt;10000</td>
</tr>
<tr>
<td>PPARδ</td>
<td>nuclear hormone receptor</td>
<td>206</td>
<td>5</td>
<td>54</td>
<td>0.3</td>
<td>&gt;10000</td>
</tr>
</tbody>
</table>

Table 7. Best Correlation Coefficient $r$ between the $-\log$ Affinity (pAffinity) and Docking Score for All Programs across All Targets

<table>
<thead>
<tr>
<th>program</th>
<th>Chk1</th>
<th>FXa</th>
<th>gyrase B</th>
<th>HCVP</th>
<th>MRS</th>
<th>E. coli PDF</th>
<th>Strep PDF</th>
<th>PPARδ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dock4</td>
<td>-0.33</td>
<td>-0.31</td>
<td>-0.39</td>
<td>0.00</td>
<td>-0.13</td>
<td>-0.38</td>
<td>-0.34</td>
<td>0.07</td>
</tr>
<tr>
<td>DockIt</td>
<td>-0.49</td>
<td>-0.19</td>
<td>-0.37</td>
<td>0.04</td>
<td>-0.28</td>
<td>-0.13</td>
<td>-0.30</td>
<td>-0.34</td>
</tr>
<tr>
<td>FlexX</td>
<td>-0.57</td>
<td>-0.31</td>
<td>-0.39</td>
<td>-0.12</td>
<td>-0.01</td>
<td>-0.42</td>
<td>-0.25</td>
<td>-0.36</td>
</tr>
<tr>
<td>Flo+</td>
<td>-0.44</td>
<td>-0.38</td>
<td>-0.36</td>
<td>-0.09</td>
<td>0.05</td>
<td>-0.27</td>
<td>-0.39</td>
<td>-0.42</td>
</tr>
<tr>
<td>Fred</td>
<td>-0.14</td>
<td>0.01</td>
<td>-0.13</td>
<td>-0.07</td>
<td>0.13</td>
<td>0.07</td>
<td>-0.24</td>
<td>0.06</td>
</tr>
<tr>
<td>Glide</td>
<td>-0.47</td>
<td>-0.08</td>
<td>-0.21</td>
<td>-0.04</td>
<td>0.08</td>
<td>-0.13</td>
<td>-0.12</td>
<td>-0.35</td>
</tr>
<tr>
<td>Gold</td>
<td>-0.42</td>
<td>-0.05</td>
<td>-0.14</td>
<td>-0.09</td>
<td>0.04</td>
<td>-0.12</td>
<td>-0.11</td>
<td>-0.43</td>
</tr>
<tr>
<td>LigandFit</td>
<td>-0.45</td>
<td>-0.13</td>
<td>-0.39</td>
<td>-0.06</td>
<td>-0.15</td>
<td>-0.21</td>
<td>-0.49</td>
<td>-0.10</td>
</tr>
<tr>
<td>MOEDock</td>
<td>-0.29</td>
<td>0.00</td>
<td>0.07</td>
<td>-0.01</td>
<td>-0.13</td>
<td>0.08</td>
<td>0.20</td>
<td>0.17</td>
</tr>
<tr>
<td>MVP</td>
<td>-0.26</td>
<td>0.10</td>
<td>-0.33</td>
<td>-0.01</td>
<td>-0.18</td>
<td>-0.17</td>
<td>-0.16</td>
<td>-0.18</td>
</tr>
</tbody>
</table>

We suggest the use of targeted scoring functions, specific to the protein we are studying.
Log \( IC_{50} \):

\[ \log(IC_{50}) = \sum_{i=0}^{N} \omega_i x_i + \sum_{j=0}^{N} \alpha_j x_j^i \]

\[ \Delta G = \sum_{i=0}^{N} \omega_i x_i \]

\[ \Delta S = \alpha_j x_i - \sum_{i=1}^{N} x_i y_j \]

\[ \log(K_i) = \sum_{i=0}^{N} \omega_i x_i + \sum_{j=1}^{M} \sum_{i=1}^{N} \lambda \]
Machine Learning methods

Scoring Function Space

\[
\log(EC_{50}) = \sum_{i=0}^{N} \omega_i x_i + \sum_{j=0}^{N} \alpha_j x_j
\]

\[
A = \pi r^2
\]

\[
f(x) = a_0 + \sum_{n=1}^{\infty} (a_n \cos \frac{n\pi x}{L} + b_n \sin \frac{n\pi x}{L})
\]

\[
\log(EC_{50}) = \sqrt{x_i} + \sum_{j} x \quad \Delta G = \sum_{i=0}^{N} \omega_i x_i
\]

\[
\Delta S = \alpha_j - x_i \sum_{i=1}^{N} x_i y_j
\]

\[
\log(K_i) = \sum_{i=0}^{N} \omega_i x_i + \sum_{j=1}^{M} \sum_{i=1}^{N} \lambda
\]

\[
a^2 + b^2 = c^2
\]

\[
f = \alpha_j \beta_i + x
\]
Sixty-four different regression methods

**Scoring Function Space**

\[ \log(K) = \sum_{i=0}^{N} \omega_i x_i + \sum_{j=1}^{M} \sum_{i=1}^{N} \lambda \]

\[ A = \pi r^2 \]

\[ f(x) = a_0 + \sum_{n=1}^{\infty} \left( a_n \cos \frac{n \pi x}{L} + b_n \sin \frac{n \pi x}{L} \right) \]

\[ \log(EC_{50}) = \sqrt{x_i} + \sum_{j}^i x \]

\[ \Delta G = \sum_{i=0}^{N} \omega_i x_i \]

\[ a^2 + b^2 = c^2 \]

\[ \Delta S = \alpha_j - x_i \sum_{i=1}^{N} x_i y_j \]

\[ f = \alpha_j \beta_i + x \]
DATA

2c5n + 77 ligands
Sixty-four different regression methods
MODEL TEST SET

\[ \rho = 0.656 \quad p = 0.0009 \]

\[ \rho = -0.051 \quad p = 0.820 \]

\[ \rho = -0.298 \quad p = 0.820 \]

\[ \rho = 0.656 \quad p = 0.0009 \]
We concluded that the use of targeted scoring functions can be a new approach to predict the binding affinity.
Thank you!

g.bitencourt@edu.pucrs.br

https://azevedolab.net/