Visualization and manipulation of Matched Molecular Series for decision support

Noel O’Boyle and Roger Sayle

NextMove Software
Matched (Molecular) Pairs

Coined by Kenny and Sadowski in 2005*
Easier to predict differences in the values of a property than it is to predict the value itself

MATCHED PAIR USAGE

• Successfully used for:
  – Rationalising and predicting physicochemical property changes
  – Finding bioisosteres

• Not very successful in improving activity
  – Activity changes dependent on binding environment

• Need to look beyond matched pairs
MATCHED SERIES OF LENGTH 2
= MATCHED PAIR

MATCHED SERIES OF LENGTH 3

$\{Cl, F, NH_2\}$
ORDERED MATCHED SERIES OF LENGTH 3

$$\text{pIC}_{50}$$

3.5

$$[\text{Cl} > \text{F} > \text{NH}_2]$$
THE MATCHED PAIR MENTALITY

• There can only be two
  – Like inhabitants of Flatland ignorant of a third dimension

• What is the equivalent of pair for three?
  – Triad, trio, triple?

• A matched pair represents a transformation from A->B
  – How would that work if there were three?
Matsy: Prediction using matched series
**Matched Series Have Preferred Orders**

<table>
<thead>
<tr>
<th>Series</th>
<th>Enrichment</th>
<th>Observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Br &gt; Cl &gt; F &gt; H</td>
<td>5.36*</td>
<td>256</td>
</tr>
<tr>
<td>Cl &gt; Br &gt; F &gt; H</td>
<td>3.14*</td>
<td>150</td>
</tr>
<tr>
<td><strong>H &gt; F &gt; Cl &gt; Br</strong></td>
<td><strong>1.53</strong>*</td>
<td><strong>73</strong></td>
</tr>
<tr>
<td>Br &gt; Cl &gt; H &gt; F</td>
<td>1.40</td>
<td>67</td>
</tr>
<tr>
<td>F &gt; Cl &gt; Br &gt; H</td>
<td>1.36</td>
<td>65</td>
</tr>
<tr>
<td>Cl &gt; F &gt; Br &gt; H</td>
<td>0.96</td>
<td>46</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td><strong>H &gt; F &gt; Br &gt; Cl</strong></td>
<td><strong>0.77</strong></td>
<td><strong>37</strong></td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td><strong>H &gt; Br &gt; F &gt; Cl</strong></td>
<td><strong>0.48</strong>*</td>
<td><strong>23</strong></td>
</tr>
<tr>
<td>Cl &gt; H &gt; F &gt; Br</td>
<td><strong>0.48</strong>*</td>
<td><strong>23</strong></td>
</tr>
<tr>
<td>Cl &gt; F &gt; H &gt; Br</td>
<td><strong>0.48</strong>*</td>
<td><strong>23</strong></td>
</tr>
<tr>
<td><strong>H &gt; Cl &gt; F &gt; Br</strong></td>
<td><strong>0.42</strong>*</td>
<td><strong>20</strong></td>
</tr>
<tr>
<td>Br &gt; F &gt; H &gt; Cl</td>
<td><strong>0.40</strong>*</td>
<td><strong>19</strong></td>
</tr>
<tr>
<td>F &gt; H &gt; Br &gt; Cl</td>
<td><strong>0.40</strong>*</td>
<td><strong>19</strong></td>
</tr>
<tr>
<td><strong>H &gt; Cl &gt; Br &gt; F</strong></td>
<td><strong>0.38</strong>*</td>
<td><strong>18</strong></td>
</tr>
<tr>
<td>F &gt; Br &gt; H &gt; Cl</td>
<td><strong>0.36</strong>*</td>
<td><strong>17</strong></td>
</tr>
<tr>
<td><strong>Br &gt; H &gt; F &gt; Cl</strong></td>
<td><strong>0.17</strong>*</td>
<td><strong>8</strong></td>
</tr>
</tbody>
</table>

The fact that certain orders are preferred may be used as the basis of a predictive method.
**Find R Groups that increase activity**

Query matched series

A > B

MATSY

<table>
<thead>
<tr>
<th>R Group</th>
<th>Observations</th>
<th>Obs that increase activity</th>
<th>% that increase activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>D</td>
<td>3</td>
<td>3</td>
<td>100</td>
</tr>
<tr>
<td>E</td>
<td>1</td>
<td>1</td>
<td>100</td>
</tr>
<tr>
<td>C</td>
<td>4</td>
<td>1</td>
<td>25</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

THE DATASET-CENTRIC APPROACH

• “Here is my dataset of molecules with activities – now tell me what to make next”

• Pro:
  – Easy for users to get up-and-running
  – Fits with their existing way of thinking
  • Don’t need to think too much about matched series

• Con:
  – User is one step removed from the matched series data on which the predictions are actually based
  – Dataset is fixed: cannot play with around with the prediction input
GOALS FOR THE INTERFACE

• Visual interface based around R-Groups as first-class objects arranged in ordered series
  – Promote new paradigm
  – Make it clear that the scaffold is not involved
• Should help break the “matched pair” mentality
  – Just a particular case of matched series
• Should be easy to play with
  – Easy to manipulate and quick to respond
• Drag-and-drop R Groups into slots to represent observed activity order
  – The query matched series
**PROS**

- Easy to **play** around with
  - Swap around order of R groups
  - See what happens if you follow the predictions
- May suggest **hypotheses**
- Useful for **searching** (not just for predictions)
- Tablet-friendly

**CONS**

- The user needs to be able to provide an ordered matched series as a query
  - You can’t just provide a dataset of molecules
NO CHEMISTRY REQUIRED

• Predictions are solely based on the order of R groups in a matched series
  – Not using any calculated properties

• Images of all R groups in ChEMBL can be generated in advance (~65K)

• ⇒ A cheminformatics toolkit is not required for the interface or even for making predictions

• In practice, we do use a toolkit to allow the user to enter R groups as SMILES
USE CASE #1
ARE MATCHED SERIES PREDICTIONS SYMMETRIC?
ARE MATCHED SERIES PREDICTIONS SYMMETRIC?

• If A>B>C>D is a highly preferred order
  – Then D>C>B>A also tends to be preferred

• Hypothesis:
  – if A reduces the activity given D>C>B
  \( \Rightarrow \) it will also improve the activity given B>C>D

• If true, then we have twice as much data to use for predictions
  – Let’s find out....
USE CASE #2
TOPLISS DECISION TREE
TOPLISS DECISION TREE

H > 4-Cl
- 4-OMe
  - 4-OMe > H
    - 4-N(Me)_2
  - 4-Cl ≥ 4-OMe
    - 3-Cl

H ~ 4-Cl
- 3,4-diCl > 4-Cl
  - 3-CF_3-4-Cl
    - 3-CF_3-4-NO_2
  - 3-CF_3
    - 4-Br
    - 4-I
    - 2,4-diCl
    - 4-NO_2

4-Cl > H
- 3,4-diCl
TOPLISS DECISION TREE

H > 4-Cl

4-OMe

4-OMe > H

4-N(Me)_2

4-Cl

4-Cl ≥ 3,4-diCl

3,4-diCl

4-CF_3

4-Br

4-I

2,4-diCl

4-NO_2

(17th)
CHEMBL-BASED DECISION TREE (ONE OF MANY)

H>4-Cl

4-OH

4-OH>H
H>4-OH>4-Cl
4-Cl>4-OH

3-pyridyl
2-OH
2-F
2-Cl

4-OMe
3-Me
4-F
4-Me

4-Cl>H

4-Cl>3,4-diCl>H

3,4-diCl

2-naphthyl
3,5-diCl
4-Br
4-NO₂
4-OMe

4-F
4-Br
3-Cl
4-OMe
Visualization and manipulation of Matched Molecular Series for decision support

http://nextmovesoftware.com
noel@nextmovesoftware.com
@nmsoftware