Advances in multi-parameter optimisation: Targeting the "best" profile for your project's objectives

ACS Spring National Meeting, March 16th 2013
Matthew Segall, Edmund Champness
Overview

• Multi-parameter optimisation in drug discovery

• Finding the ‘best’ profile for your project’s objective
  – Example: Selection to reduce toxicity risk

• ‘Hard’ vs. ‘soft’ boundaries
  – Example: Selection for CNS indications

• Testing the robustness of your decisions
  – Sensitivity analysis

• Conclusions
Multi-parameter Optimisation in Drug Discovery
The Objectives
Multi-parameter optimisation

- Identify chemistries with an optimal **balance** of properties

- Quickly identify situations when such a balance is not possible
  - Fail fast, fail cheap
  - Only when **confident**

*M.D. Segall Curr. Pharm. Des. 18(9) pp. 1292-1310 (2012)*
Multi-parameter Optimisation
Probabilistic Scoring*

<table>
<thead>
<tr>
<th>Profile</th>
<th>Desired Value</th>
<th>Importance</th>
</tr>
</thead>
<tbody>
<tr>
<td>logS</td>
<td>&gt; 1</td>
<td></td>
</tr>
<tr>
<td>HIA category</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>logP</td>
<td>0 -&gt; 3.5</td>
<td></td>
</tr>
<tr>
<td>BBB log([brain]:[blood])</td>
<td>-0.2 -&gt; 1</td>
<td></td>
</tr>
<tr>
<td>BBB category</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>P-gp category</td>
<td>no</td>
<td></td>
</tr>
<tr>
<td>hERG pIC50</td>
<td>≤ 5</td>
<td></td>
</tr>
<tr>
<td>2C9 pKi</td>
<td>≤ 6</td>
<td></td>
</tr>
<tr>
<td>2D6 affinity category</td>
<td>low medium</td>
<td></td>
</tr>
<tr>
<td>PPB90 category</td>
<td>low</td>
<td></td>
</tr>
</tbody>
</table>

Multi-parameter Optimisation
Probabilistic Scoring*

- Property data
  - Experimental or predicted
- Criteria for success
  - Relative importance
- Uncertainties in data
  - Experimental or statistical

Score (Likelihood of Success)
Confidence in score

Data do not separate these as error bars overlap

Bottom 50% may be rejected with confidence

Error bars show confidence in overall score

Finding the ‘Best’ Profile for your Project

Objectives

Patent pending
Finding Tailored Profiles

Objectives

• Use existing data to find scoring profiles that identify compounds with improved chance of success
  – Any drug discovery objective, e.g. clinical, PK, toxicity...
  – Once developed, a profile can be applied prospectively to find new compounds

• Identify most important data with which to distinguish between successful and unsuccessful compounds
  – Any data can be used as input, calculated or experimental

• Explore multi-parametric data
  – Consider properties simultaneously, not individually
  – Avoid ‘over counting’ of correlated factors

• Rules must be interpretable and modifiable
  – Avoid black boxes
  – Synergy between computer and experts

*Patent pending
What is a Rule?

• A **Rule** is a set of property criteria that in combination identify ‘good’ compounds, e.g.

<table>
<thead>
<tr>
<th>logP &lt; 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ligand efficiency &gt; 0.3</td>
</tr>
<tr>
<td>100 &lt; MW &lt; 450</td>
</tr>
<tr>
<td>PPB category = low</td>
</tr>
</tbody>
</table>

• For example, Lipinski RoF:

<table>
<thead>
<tr>
<th>logP&lt;5</th>
<th>MW&lt;500</th>
</tr>
</thead>
<tbody>
<tr>
<td>HBD&lt;5</td>
<td>HBA&lt;10</td>
</tr>
</tbody>
</table>
Finding Rules with PRIM

- **A Rule** is a box in multi-dimensional property space containing significantly more ‘good’ than ‘bad’ compounds
  - Use Patient Rule Induction Method (PRIM) by Friedman and Fisher* find rules in multi-dimensional data
  - Equivalent to a scoring profile

Example Application
Finding rules for selection of non-toxic compounds

- *In vitro* assay data from CEREP Bioprint®
  - Percentage inhibition of 185 targets including GPCF, kinase, NR, P450s...

- Drugs labelled as ‘cardiotoxic’, ‘hepatotoxic’ or ‘clean’
  - Based on FDA Adverse Event Reporting System
  - Reporting odds ratio (ROR) of 2.5 or above at System Organ Class level in MeDRA Ontology
  - Cardiotoxicity set: 408 ‘cardiotoxic’, 66 ‘non-cardiotoxic’

- Data sets divided into training, validation and test sets
  - Ratio 70:15:15
Example Application
Cardiotoxicity results

- Selected only 3 targets from 185
  - Rules ‘make sense’: Targets identified have known CV side effects
- 5/6 compounds meeting all criteria are non-cardiotoxic (83%)
- 19/20 compounds failing all criteria are cardiotoxic (95%)
Example Application
Hepatotoxicity results

Rules are (just) statistically significant, but don’t ‘make sense’
- Rules appear to be result of noise in small data set

Large majority of the targets in data set are not known to relate with hepatotoxicity
- In few examples, e.g. PPARγ there are a statistically insignificant number of inhibitors in the data set

Non ‘black-box’ method highlights limitations of data set
Desirability Functions*

- Relate property values to how ‘desirable’ the outcome

- Avoid hard cut-offs that draw artificially hard distinction between similar compounds

- Add ‘soft’ boundaries to ideal ranges

* Harrington EC. (1965) Ind. Qual. Control. 21 p. 494
Desirability Functions*

- Relate property values to how ‘desirable’ the outcome

- Avoid hard cut-offs that draw artificially hard distinction between similar compounds

- Add ‘soft’ boundaries to ideal ranges

---

* Harrington EC. (1965) Ind. Qual. Control. 21 p. 494
Desirability Functions*

- Relate property values to how ‘desirable’ the outcome

- Avoid hard cut-offs that draw artificially hard distinction between similar compounds

- Add ‘soft’ boundaries to ideal ranges

* Harrington EC. (1965) Ind. Qual. Control. 21 p. 494
Desirability Functions*

• Relate property values to how ‘desirable’ the outcome

```
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
|     | 0.0  | 0.2  | 0.4  | 0.6  | 0.8  | 1.0  |
+-----+-----+-----+-----+-----+-----+-----+
| Property | 0   | 2   | 4   | 6   | 8   | 10   |
+-----+-----+-----+-----+-----+-----+-----+
```

‘Soft’ range: 4-6

• Avoid hard cut-offs that draw artificially hard distinction between similar compounds

• Add ‘soft’ boundaries to ideal ranges

* Harrington EC. (1965) Ind. Qual. Control. 21 p. 494
Desirability Functions
Example: CNS MPO*

CNS MPO = sum of desirabilities for each parameter

- 74% of marketed CNS drugs achieved CNS MPO > 4 vs. 60% of Pfizer candidates
- Correlations observed between high CNS MPO score and good in vitro ADME properties, e.g. MDCK $P_{app}$, HLM stability, P-gp transport

Determining ‘Soft’ Box Boundaries

- Box bounds previously only output as hard cut-offs
- Sensitivity analysis of box bounds to data sampling
  - Particularly important for sparse data
  - Incorporate uncertainty into the generated box bounds
  - Cross validation between training/validation sets
Example Application
CNS Drugs

• Data set of 119 CNS Drugs and 108 Candidates published by Wager et al. in CNS MPO paper

• Divided into training, validation and test sets (55:25:20)

• Rule with hard cut-offs:

<table>
<thead>
<tr>
<th>Set</th>
<th>Mean Improvement (%)</th>
<th>Support (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Train</td>
<td>42</td>
<td>28</td>
</tr>
<tr>
<td>Val</td>
<td>56</td>
<td>32</td>
</tr>
<tr>
<td>Test</td>
<td>47</td>
<td>34</td>
</tr>
</tbody>
</table>
Example Application
CNS Drugs – Introducing ‘soft’ boundaries

MW

pKa

logP
Example Application
CNS Drugs – Comparison of ROC curves for test set

- Random CNS MPO (AUC=0.66)
- Rule Induction Hard Boundaries (AUC=0.76)
- Rule Induction Soft Boundaries (AUC=0.77)
Testing the Robustness of Your Decisions
Patent pending
Sensitivity Analysis

• What impact would changing a property criterion have on the decision we would make?
  – How large a change is necessary to have a significant impact?

• To which property criteria is compound priority most sensitive?
  – Which criteria/importance will, if modified, significantly change the order of compound priority?

• Highlight new avenues for exploration and avoid missed opportunities

• Considerations
  – Need to consider statistical significance of reordering (given uncertainties in scores)
  – Interested in changes to high-ranked compounds. Reordering of rejected compounds is not relevant
Sensitivity Analysis
Importance of uncertainty

Modified Spearman’s rank correlation coefficient accounts for uncertainty
Example Output

Sensitive parameter

What parameters are most sensitive?

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value Sensitivity</th>
<th>Importance Sensitivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>5HT1a affinity (pKi)</td>
<td>1.000</td>
<td>0.008</td>
</tr>
<tr>
<td>logP</td>
<td>0.310</td>
<td>0.096</td>
</tr>
<tr>
<td>BBB log([brain]:[blood])</td>
<td>0.249</td>
<td>0.015</td>
</tr>
<tr>
<td>hERG pIC50</td>
<td>0.096</td>
<td>0.207</td>
</tr>
<tr>
<td>2D6 affinity category</td>
<td>N/A</td>
<td>0.107</td>
</tr>
<tr>
<td>BBB category</td>
<td>N/A</td>
<td>0.055</td>
</tr>
<tr>
<td>logS</td>
<td>0.040</td>
<td>0.002</td>
</tr>
</tbody>
</table>

What magnitude of change has a significant impact?

What compounds are most affected?
Example Output

Insensitive parameter

What parameters are most sensitive?

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value Sensitivity</th>
<th>Importance Sensitivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>5HT1a affinity (pKi)</td>
<td>1.000</td>
<td>0.008</td>
</tr>
<tr>
<td>logP</td>
<td>0.310</td>
<td>0.096</td>
</tr>
<tr>
<td>BBB log([brain]:[blood])</td>
<td>0.249</td>
<td>0.015</td>
</tr>
<tr>
<td>hERG pIC50</td>
<td>0.096</td>
<td>0.207</td>
</tr>
<tr>
<td>2D6 affinity category</td>
<td>N/A</td>
<td>0.107</td>
</tr>
<tr>
<td>BBB category</td>
<td>N/A</td>
<td>0.055</td>
</tr>
<tr>
<td>logS</td>
<td>0.040</td>
<td>0.002</td>
</tr>
</tbody>
</table>

What magnitude of change has a significant impact?

What compounds are most affected?
Conclusion

• Rule induction can generate interpretable parameter scoring profiles tailored to specific project objectives

• ‘Soft’ boundaries provide more subtle distinctions between compounds

• Sensitivity analysis of scoring criteria is important to avoid missed opportunities due to the criteria we have chosen

  – 10.1016/j.drudis.2014.01.005
  – www.optibrium.com/community/publications

• See a live demo at Optibrium booth #1516