Relative Drug Likelihood: Going beyond ‘Drug-Likeness’

ACS Fall National Meeting, August 23rd 2012
Matthew Segall, Iskander Yusof
Overview

• ‘Drug-Like’ Properties

• Quantitative Estimate of Drug-Likeness (Bickerton et al.)
  – Multi-parameter Optimization
  – Desirability Functions

• Beyond ‘Drug-like’: Relative Drug Likelihood

• Results

• Conclusion
‘Drug-like’ Properties
Drug-like Properties

Background

• Rules for simple compound characteristics that drugs have in common

• Original and most influential: Lipinski’s Rule of Five

<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>

• Many others have been proposed, e.g.:
  – Rotatable bonds
  – Aromatic rings
  – Polar surface area
  – Fraction of sp3 carbons
Drug-like Properties
Strengths and Weaknesses

• Strengths
  – Easy to understand and apply
  – Compounds with ‘non drug-like’ properties lie in regions of property space with poor precedence
  – Good guide to avoid potential pitfalls

• Weaknesses
  – Simple characteristics are only weakly predictive of biological properties
  – Binary pass/fail rules
  – Tendency to apply over-rigorously (is MW of 501 worse than 499?)
  – Rules apply only to objective for which they were determined (most commonly oral bioavailability)
  – Many are derived only from analysis of drugs, i.e. what makes drugs similar
Quantitative Estimate of Drug-Likeness (QED)
Multi-Parameter Optimization
Desirability Functions

• Combine values of multiple characteristics into single measure of ‘quality’ of a compound*

• Desirability functions relate property values to how ‘desirable’ the outcome

![Desirability vs Property Graph]

Simple filter: >5

---

Multi-Parameter Optimization

Desirability Functions

• Combine values of multiple characteristics into single measure of ‘quality’ of a compound*

• Desirability functions relate property values to how ‘desirable’ the outcome

Multi-Parameter Optimization
Desirability Functions

- Combine values of multiple characteristics into single measure of ‘quality’ of a compound*

- Desirability functions relate property values to how ‘desirable’ the outcome

Multi-Parameter Optimization

Desirability Functions

• Combine values of multiple characteristics into single measure of ‘quality’ of a compound*

• Desirability functions relate property values to how ‘desirable’ the outcome

Ideal value: 5

Multi-Parameter Optimization
Desirability Functions

• Combine values of multiple characteristics into single measure of ‘quality’ of a compound*

• Desirability functions relate property values to how ‘desirable’ the outcome

*Trend: >8

Multi-Parameter Optimization
Desirability Functions

- Combine values of multiple characteristics into single measure of ‘quality’ of a compound*

- Desirability functions relate property values to how ‘desirable’ the outcome

- Combine multiple properties into ‘desirability index’
  - Additive: \[ D = \frac{d_1(Y_1) + d_2(Y_2) + \cdots + d_n(Y_n)}{n} \]
  - Multiplicative: \[ D = (d_1(Y_1) \times d_2(Y_2) \times \cdots \times d_n(Y_n))^{1/n} \]

QED*

• Combine values for 8 characteristics
  - Molecular weight (M_r)
  - Lipophilicity (alogP)
  - Number of hydrogen bond donors (HBD)
  - Number of hydrogen bond acceptors (HBA)
  - Polar surface area (PSA)
  - Number of rotatable bonds (ROTB)
  - Number of aromatic rings (AROM)
  - Count of alerts for undesirable substructures (ALERT)
• For each characteristic a desirability function was fitted to distribution for a set of 771 oral drugs

The desirabilities for the 8 characteristics are combined using a multiplicative approach:

\[
QED_w = \exp\left(\frac{w_{Mr} \ln d_{Mr} + w_{ALOGP} \ln d_{ALOGP} + w_{HBA} \ln d_{HBA} + w_{HBD} \ln d_{HBD} + w_{PSA} \ln d_{PSA} + w_{ROTB} \ln d_{ROTB} + w_{AROM} \ln d_{AROM} + w_{ALERT} \ln d_{ALERT}}{w_{Mr} + w_{ALOGP} + w_{HBA} + w_{HBD} + w_{PSA} + w_{ROTB} + w_{AROM} + w_{ALERT}}\right)
\]
• QED avoids the pitfalls of hard cut-offs
  – Provides a single metric for the ‘similarity’ of a compound to known oral drugs

• Bickerton et al. showed that QED correlates with chemists’ opinion on ‘beauty’ of compounds

• Benchmarked QED for selection of 771 oral drugs vs. 10,250 compounds from the PDB ligand dictionary
  – N.B. Not a fully independent test set of drugs

QED Benchmarking Results

The graph shows the sensitivity vs. FPR (1-specificity) for different methods:
- QEDw,u
- QEDw,mo
- Random

The green line represents QEDw,u, the orange line represents QEDw,mo, and the black line represents Random.
Beyond ‘Drug-like’: Relative Drug Likelihood
Similarity is Not Enough

- A compound with a characteristic that is ‘similar’ to known drugs does not necessarily have an increased chance of success.

- Some properties distinguish drugs from non-drugs better than others.
Relative Drug Likelihood
Bayesian probability theory

• Analysis of characteristics of known drugs gives us $P(X|\text{Drug})$

• We would like to know $P(\text{Drug}|X)$

• Bayes’ theorem allows us (in principle) to calculate this:

\[
P(\text{Drug} | X) = \frac{P(X | \text{Drug})P(\text{Drug})}{P(X)}
\]
Relative Drug Likelihood
Bayesian probability theory

• Compare with probability compound is not a drug:

\[ P(\text{not Drug} \mid X) = \frac{P(X \mid \text{not Drug})P(\text{not Drug})}{P(X)} \]

• We want to find compounds with high relative probability of being drug, so take ratio

\[ \frac{P(\text{Drug} \mid X)}{P(\text{not Drug} \mid X)} = \frac{P(X \mid \text{Drug})P(\text{Drug})}{P(X \mid \text{not Drug})P(\text{not Drug})} \]

Constant (v. small)
Relative Drug Likelihood  
Bayesian probability theory

Therefore, we define the desirability of a value $x$ of property $X$ as:

$$d(x) = \frac{P(X = x | \text{Drug})}{P(X = x | \text{not Drug})}$$

Need to choose appropriate negative set of non-drugs from which we would like to distinguish drugs

- Choose ChEMBL database* as representative of ‘med chem’ compounds
- Trained on random selection of 1000 compounds from ChEMBL and 771 compound oral drug set from Bickerton et al.

* https://www.ebi.ac.uk/chembldb/
Relative Drug Likelihood
Example – Molecular Weight
Relative Drug Likelihood
Analysis of 8 properties from QED
Relative Drug Likelihood

PSA
Relative Drug Likelihood

HBA
Relative Drug Likelihood

- Combine desirabilities of individual characteristics to give overall Relative Drug Likelihood (RDL)

- Multiplicative – analogous to QED

\[
RDL = \exp\left(\frac{1}{n} \sum_{i=1}^{n} \ln(d_i(x_i)))\right)
\]
Results
Identifying Drugs
Selecting from ‘med chem’ compounds

- 771 drug ‘test’ set from Bickerton et al. vs. >650k compounds from ChEMBL (independent of training set)

![ROC curve diagram](image-url)
Identifying Drugs
Selecting from PDB ligand dictionary

- 771 drug ‘test’ set from Bickerton et al. vs. 10,250 compounds from the PDB ligand dictionary

![Graph showing sensitivity vs. FPR (1-specificity) for different methods: RDL, QEDw,u, QEDw,mo, and Random.](image-url)
Comparing PDB Ligands with ChEMBL
Molecular weight distribution
Identifying Drugs
Selecting from PDB ligand dictionary

- PDB ligand dictionary is not representative of med chem compounds
- Retrain RDL using 500 compound ‘negative’ set from PDB ligand dictionary
- 771 drug ‘test’ set from Bickerton et al. vs. 9.750 compounds from the PDB ligand dictionary

![ROC curve diagram](image)
Conclusions

• Binary rules for selection of compounds are risky
  – Filters may throw away valuable opportunities

• The criteria to accurately identify good compounds depend on the population from which we are selecting
  – We have used ChEMBL as representative of ‘med chem’ compounds
  – ChEMBL is already biased by med chemists experience, so RDL shows added value over medicinal chemistry ‘instincts’

• Could be applied to different therapeutic classes

• Having a good RDL (or QED etc.) is not a guarantee of success
  – Relative drug likelihood
  – Remember the very small constant we ignored (P(Drug)/P(not Drug))
  – A compound with good ‘drug-like’ characteristics may fail for a large number of reasons

• Preprint and scripts to calculate RDL yourself can be downloaded from:
  – www.optibrium.com/community
Acknowledgements

• Optibrium team
  – Ed Champness
  – Chris Leeding
  – James Chisholm
  – Nick Foster
  – Hector Garcia

• Helpful comments from many others
  – Andrew Hopkins
  – Richard Bickerton
  – Stephen Smith