Molecule properties and compound developability – what’s new?

Dr Tim Ritchie
TJR-Chem, Ranco (VA), Italy
http://www.linkedin.com/in/timritchiechemistry
Dr Tim Ritchie - Background

1989 – 2005
Novartis Institute, London UK
Medicinal chemist, neuroscience area:
$BKB_1$, $BKB_2$, $nAChR\alpha_7$, $NK_1$, $TRPV_1$

2005 – 2008
GlaxoSmithKline, Stevenage UK
Medicinal chemistry design expert, respiratory area:
Medchem-Compchem interface
$CCR_4$, $DP_1$, $PI3K$, $Syk$

2008 – date
TJR-Chem, Ranco (VA), Italy
Independent chemistry consultant:
Data mining, structure-property analysis (rings),
ligand-based drug design, phys-chem tutorials

Tim Ritchie Cambridge 2013
Outline of talk

• Molecule properties
• Developability
• Analysis of compounds in different regions of clogP/MW space
  – Developability score from solubility, permeability, protein binding and P450 inhibition data
  – Which properties can best differentiate high/low developability?
• Summary
• Acknowledgements

Tim Ritchie Cambridge 2013
A balanced view of molecule properties?

Rules are rules: a clogP of 4.99 is OK but not 5.01!

Properties? The ADMET assays tell me what’s OK!
Scepticism about molecule properties

• Rule-of-Five: help or hindrance?
  – Too lax? 
  – Too strict? – limits creativity, opportunities
  – Applied too widely? – Not a ‘drug-like’ score!

• Do property rules apply to all chemical space?

• Can general ‘global’ models compete with more specific ‘local’ models?

• Drug discovery is complicated: can simple properties really predict anything useful?
Molecule properties – if only...

Likelihood of becoming a successful drug vs. Molecule property X

$r^2 = 0.959$
clogP and drug fraction in human plasma

From data published in J Med Chem 2010 53 1098

Tim Ritchie Cambridge 2013
‘Developability’ and ring type
Sol/Perm/PB/3A4 profile for compounds with 3 rings

Tim Ritchie Cambridge 2013
Molecule property proliferation

2000

LogP/D MW HBA HBD PSA

2013

LogP/D MW HBA HBD PSA
RB Ar/HA (i_aro)
Arom Ar-sp3 Bl
Fsp3 (i_ali) Chiral
PFI fMF

LE GE LipE LELP
PEI SILE BEI SEI

• Do we need all of them? Which are most important?

Tim Ritchie Cambridge 2013
Compound developability

• ‘Developability’ is determined by how well a compound fares in ADME-related assays that are relevant to drug development

• High developability compounds are defined as having:
  – High solubility
  – High permeability
  – Low protein binding
  – Low CYP450 3A4 inhibition
Compound scoring based on assays

Scores based on performance in assay

High SOL
High PERM
Low PB
Low 3A4
TOTAL

High developability
Medium developability
Low developability

Tim Ritchie Cambridge 2013
Developability in clogP/MW quadrants

30,000 diverse GSK molecules

Which molecule properties best differentiate low-scoring and high-scoring compounds?
Phys-chem properties and developability

- 40 molecule properties calculated
  - Aromatic, aliphatic, lipophilic, polar, ionisable, atom counts, flexibility, size

- O-PLS regression models generated for each clogP/MW quadrant
  - Separates variation in properties that is correlated with developability from variability that is uncorrelated (orthogonal)
  - Importance of each property in differentiating high/low developability determined for each quadrant

- Results indicate that:
  - Some properties are important in all quadrants
  - Some properties vary in importance depending on the quadrant
  - Others are less important to developability
Examples of property importance*

**Ar-sp3** has high importance to developability across all quadrants

**Heteroaliphatic ring count** importance to developability varies with quadrant

**Fluorine atom count** has little impact on developability

*Values >1 are considered more important*
Importance of Ar-sp3 and fluorine count

Ar-sp3: high importance to developability:

There is a significant difference in Ar-sp3 values between the low and high developability compounds.

Fluorine count: low importance to developability:

No difference in fluorine atom count between high and low developability compounds.

Tim Ritchie Cambridge 2013
Importance to developability: Aromatic-composite

Tim Ritchie Cambridge 2013
Importance to developability: Aromatic

![Graph showing the importance to developability of different aromatic structures.](image)
Importance to developability: Aliphatic

- Fsp3
- sp3 atom
- Chiral atom
- Non-Ar ring
- Het Ali ring
- Car Ali ring

Quadrant:
- Large & Greasy
- Greasy Not Large
- Large Not Greasy
- Not Large Not Greasy

Tim Ritchie Cambridge 2013
Importance to developability: Lipophilicity

![Graph showing importance to developability of lipophilicity. The graph is divided into quadrants representing different combinations of size and greasiness. The x-axis represents the quadrants: Large & Greasy, Greasy Not Large, Large Not Greasy, Not Large Not Greasy. The y-axis represents the importance to developability ranging from 0 to 2. The graph shows a decrease in importance as the quadrants transition from Greasy Not Large to Large Not Greasy to Not Large Not Greasy.]

Tim Ritchie Cambridge 2013
Importance to developability: **Polarity**

![Graph showing the importance of various molecular properties in different quadrants: Alpha, Pi, HBD, tPSA, HBA, BetaH.](image)

Tim Ritchie Cambridge 2013
Importance to developability: Ionisability

The graph illustrates the importance of ionisability to developability across different quadrants: Large & Greasy, Greasy Not Large, Large Not Greasy, and Not Large Not Greasy. The graph shows the distribution of ionisability across these quadrants, indicating the relative importance of developing compounds with specific ionisability characteristics.
Importance to developability: Atom counts

Tim Ritchie Cambridge 2013
Importance to developability: **Flexibility**

![Graph showing the importance of flexibility in different quadrants.](image)

- **Quadrants:** Large & Greasy, Greasy Not Large, Large Not Greasy, Not Large Not Greasy
- **Axes:** Importance to developability, Quadrant
- **Data Points:** Various flexibility measures (fMF, Flexibility, Chain atom, Rot bond)

Tim Ritchie Cambridge 2013
Importance to developability: **Size**
## Properties & developability: Summary

<table>
<thead>
<tr>
<th>‘Always important’</th>
<th>‘Important some of the time’</th>
<th>‘Less important’</th>
</tr>
</thead>
<tbody>
<tr>
<td>clogD+Ar ring ↓</td>
<td>Ar rings ↓</td>
<td>Atom counts ↓</td>
</tr>
<tr>
<td>clogD+(Ar-sp3) ↓</td>
<td>clogD ↓</td>
<td>Flexibility ↑</td>
</tr>
<tr>
<td>Ar-sp3 ↓</td>
<td>Positive ionisable groups ↑</td>
<td>Negative ionisable groups</td>
</tr>
<tr>
<td>Ar/HA ↓</td>
<td>Heteroaliphatic ring count ↑</td>
<td>Size ↓</td>
</tr>
<tr>
<td>Fsp3 ↑</td>
<td>HBDs ↓</td>
<td>HBAs</td>
</tr>
</tbody>
</table>

Lower values increase developability

Higher values increase developability

Tim Ritchie Cambridge 2013
Final points

• High clogP has a stronger negative impact on developability than high MW

• Aromatic-composite descriptors appear most important with respect to high/low developability
  – Large & Greasy compounds with less aromatic character / lower clogD = higher developability
  – Not Large & Not Greasy compounds with more aromatic character / higher clogD = lower developability

• N.B. This is a holistic view of GSK compounds
  – Is it representative of ‘Pharma’ structures?
  – Property importance may vary between different chemical classes / structural frameworks
Acknowledgements

• GlaxoSmithKline UK
  – Simon Macdonald
  – Simon Peace
  – Stephen Pickett
  – Chris Luscombe
Molecule properties and compound developability – what’s new?

Dr Tim Ritchie
TJR-Chem, Ranco (VA), Italy
http://www.linkedin.com/in/timritchiechemistry
List of descriptors used in analysis

- \( \log D + (\text{Ar-sp3}) = \text{ChemAxon clog D7.4} + (\text{Ar-sp3}) \)
- \( \log D + \text{aring} = \text{ChemAxon clog D7.4} + \text{aromatic ring count} \)
- \( \text{Ar-sp3} = \text{aromatic atom count minus sp3 carbon atom count} \)
- \( \text{Ar/HA} = \text{aromatic atom count/heavy atom count} \)
- \( \text{sp2 atom} = \text{sp2-hybridised atom count} \)
- \( \text{Ar ring} = \text{aromatic ring count} \)
- \( \text{R2} = \text{Abraham's R2 (E) (excess molar refraction) descriptor} \)
- \( \text{Het Aro ring} = \text{heteroaromatic ring count} \)
- \( \text{Car Aro ring} = \text{carboaromatic ring count} \)
- \( \text{Fsp3} = \text{sp3 carbon atom count/total carbon atom count} \)
- \( \text{sp3 atom} = \text{sp3-hybridised atom count} \)
- \( \text{Chiral atom} = \text{chiral atom count} \)
- \( \text{Non-Ar ring} = \text{non-aromatic ring count} \)
- \( \text{Het Ali ring} = \text{heteroaliphatic ring count} \)
- \( \text{Car Ali ring} = \text{carboaliphatic ring count} \)
- \( \log D = \text{ChemAxon clog D7.4} \)
- \( \log P = \text{daylight clog P} \)
- \( \text{Alpha} = \text{Abraham's Alpha (A) (H-bond acidity) descriptor} \)
- \( \text{Pi} = \text{Abraham's Pi (S) (combined dipolarity/polarisability) descriptor} \)
- \( \text{HBD} = \text{H-bond donor count} \)
- \( \text{tPSA} = \text{topological polar surface area} \)
- \( \text{HBA} = \text{H-bond acceptor count} \)
- \( \text{BetaH} = \text{Abraham's BetaH (B) (H-bond basicity) descriptor} \)
- \( \text{Pos ionisable} = \text{positive ionisable group count} \)
- \( \text{Neg ionisable} = \text{negative ionisable group count} \)
- \( \text{Arom sulfur} = \text{aromatic sulfur atom count} \)
- \( \text{Ring atom count} = \text{total ring atom count} \)
- \( \text{Sulfur} = \text{sulfur atom count} \)
- \( \text{Chlorine} = \text{chlorine atom count} \)
- \( \text{Nitrogen} = \text{nitrogen atom count} \)
- \( \text{Oxygen} = \text{oxygen atom count} \)
- \( \text{Carbon} = \text{carbon atom count} \)
- \( \text{Fluorine} = \text{fluorine atom count} \)
- \( \text{fMF} = \text{fraction of atoms in the molecular framework} \)
- \( \text{Flexibility} = (100 \ \text{rotatable bond count})/\text{total bond count} \)
- \( \text{Chain atom} = \text{chain atom count} \)
- \( \text{Rot bond} = \text{rotatable bond count} \)
- \( \text{Mol Wt} = \text{molecular weight} \)
- \( \text{Mol Ref} = \text{molecular refractivity} \)
- \( \text{Heavy atom} = \text{heavy atom count} \)

Tim Ritchie Cambridge 2013
# O-PLS regression models

## Table 1 OPLS models summary

<table>
<thead>
<tr>
<th>Quadrant</th>
<th>n(&lt;0.33)(^a)</th>
<th>n(&gt;0.66)(^b)</th>
<th>R2X(^c)</th>
<th>R2Y(^d)</th>
<th>Q2Y(^e)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Large &amp; Greasy</td>
<td>3726</td>
<td>1880</td>
<td>0.711</td>
<td>0.440</td>
<td>0.435</td>
</tr>
<tr>
<td>Greasy Not Large</td>
<td>873</td>
<td>647</td>
<td>0.564</td>
<td>0.510</td>
<td>0.497</td>
</tr>
<tr>
<td>Large Not Greasy</td>
<td>752</td>
<td>2952</td>
<td>0.422</td>
<td>0.329</td>
<td>0.322</td>
</tr>
<tr>
<td>Not Large Not Greasy</td>
<td>378</td>
<td>2910</td>
<td>0.471</td>
<td>0.271</td>
<td>0.261</td>
</tr>
</tbody>
</table>

\(^a\) The number of compounds with a developability score of <0.33.
\(^b\) The number of compounds with a developability score of >0.66.
\(^c\) The cumulative fraction of X variation modelled by the predictive component.
\(^d\) The cumulative fraction of Y variation modelled by the predictive component.
\(^e\) The cumulative fraction of Y variation predicted by the X model, by cross-validation (‘leave-many-out’, sevenfold).