Capturing and Applying Knowledge to Guide Compound Optimisation

ACS Spring Meeting 2018

Matthew Segall, Edmund Champness, Peter Hunt and Tamsin Mansley
Overview

• Why capture and apply knowledge computationally?

• Capturing and applying different types of knowledge
  - Chemistry
  - Structure activity relationships
  - Project objectives

• Example application
  - Optimisation of high quality, selective Anagliptin analogues

• Conclusions
Why Capture Knowledge Computationally?

- **Different viewpoints**
  - Drug discovery is a multi-disciplinary field
  - Knowledge transfer

- **Limited experience**
  - Even within our own field, we each have limited experience

- **Limited memory**
  - We can only easily remember a fraction of what we have experienced

- **Computers have virtually unlimited capacity and perfect recall**
Why Apply Knowledge Computationally?

- Being able to search large databases of information/knowledge is useful, but limited
- Computers can apply stored knowledge much faster than a person
- Even scientists have hidden biases*
  - We choose which knowledge to ‘believe’
- Computers can explore ideas more quickly and rigorously than a person

Knowledge is like paint... It does no good until it is applied

- Doe Zantamata

Capturing and Applying Knowledge
Chemistry – What might we make?
Chemistry – What could we make?

Reactions and Reagents
Chemistry – What might we make?

Transformations

E.g. Bioisosteric replacement

SMIRKS: ([c;$(!n1@n@c@n@c@n1):1]n1[cH]n[cH]n1)\rightarrow([c;$(!C;@!N):1]C#N)
Chemistry – What might we make?

Transformations

E.g. Bioisosteric replacement

SMIRKS: ([c;$(@n1@c@n@c@n1):1]n1[nH]n[nH1])>>([c;$(!C#;@N):1]C#N)

Calculated with torch3D from Cresset
Chemistry – What might we make?

Transformations

E.g. Prodrug strategy

SMIRKS: ([C;$(C!@N):1][NH2])>>([C;$(C!@N)!@C!@C=;!@C1@C(@N2(@C(=;!@O)@C@C2@O1))!@C(!@O)=;!@O):1][NH][CH2][CH]=C1[C@@H](N2C(=O)[CH2][C@H]2O1)C([OH])=O)
Chemistry – What might we make?

Transformations

E.g. Ring opening/closing

SMIRKS: ([c;$\text{c1}\text{c}\text{c}(\text{n}\text{c}\text{n}1)!\text{N}!\text{C}!\text{C})\text{:2}]\text{c1}\text{c(n[\text{x2}:4]\text{n1})[\text{NH}][\text{C}:5][\text{C}:3]}>>\text{N}\text{21C}[\text{c}:1][\text{c}:2]\text{(N=}[\text{C};\text{x2};\text{c1}=;\text{N}\text{c}(\text{c}2\text{N1}\text{C}(\text{c}\text{N}=2)!\text{C}!\text{H})!\text{H})!\text{H})\text{:41}[\text{H}][\text{H}]\text{=N}[\text{C};\text{x2}:5][\text{CH}2\text{C}:3])
Chemistry – What might we make?

BIOSTER™

Database of ~30,000 precededented transformations from the chemistry literature

Chemistry – What could we make?

Applying Transformations
Capturing and Applying Knowledge
SAR – How are molecular properties likely to change?
SAR – How are Molecular Properties Likely to Change? Quantitative Structure-Activity Relationships

\[ y = f(x_1, x_2, x_3, \ldots) \pm \varepsilon \]

- **Data**
  - Quality data is essential
  - Public data need very careful curation* (and may not be good enough)

- **Descriptors, e.g.**
  - Whole molecule properties, e.g. logP, MW, PSA...
  - Structural descriptors, SMARTS, fingerprints...

- **Machine learning method, e.g.**
  - Artificial neural networks, support vector machines, random forest, Gaussian processes...

*Waldman et al. JCAMD (2015) DOI: 10.1007/s10822-015-9865-0*
SAR – How are Molecular Properties Likely to Change?
Quantitative Structure-Activity Relationships

Terfenadine: hERG pIC$_{50}$=6.7
Fexofenadine: hERG pIC$_{50}$=5.2
Capturing and Applying Knowledge

Project objectives
Project Objectives

Multi-parameter optimisation

Hit/Lead

- Potency
- Safety
- Absorption
- Solubility
- Metabolic stability

Drug
Project Objectives
Multi-parameter optimisation – Probabilistic Scoring

- Evaluates all available data against project criteria
- Accounts for the uncertainties in all compound-related data
- Objective assessment of compounds’ chances of success

User-defined scoring profile

<table>
<thead>
<tr>
<th>Property</th>
<th>Desired Value</th>
<th>Importance</th>
</tr>
</thead>
<tbody>
<tr>
<td>SHT1a affinity (pKi)</td>
<td>&gt; 7</td>
<td></td>
</tr>
<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>ZC9 pKi</td>
<td></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>

Compounds ranked by likelihood of success

Histograms for quick visual guide to compound properties

Example Application
Optimisation of high quality, selective Anagliptin analogues
Anagliptin

- Dipeptidyl Peptidase-4 (DPP-4) inhibitor
  - Treatment for type-2 diabetes mellitus

- Withdrawn in most markets due to animal toxicity*
  - Concerns raised about selectivity over DPP-2, DPP-8 and DPP-9
  - Since disproved...

- Explore strategies to improve DPP-4 activity and selectivity over other DPP isoforms

PDB 3WQH

Capturing SAR for DPP Activities

- QSAR models built with random forests method in StarDrop’s Auto-Modeller*
  - 2D SMARTS descriptors and whole molecule properties, e.g. logP, MW, TPSA...

- Data sets of pIC$_{50}$ data from ChEMBL†

- Validation on independent test sets:

  - **DPP-4**: $R^2=0.68$
  - **DPP-2**: $R^2=0.72$
  - **DPP-8**: $R^2=0.62$
  - **DPP-9**: $R^2=0.75$

---

Capturing the Optimisation Objectives

<table>
<thead>
<tr>
<th>Property</th>
<th>Desired Value</th>
<th>Importance</th>
</tr>
</thead>
<tbody>
<tr>
<td>DPP4_RF_model</td>
<td>7 -&gt; inf</td>
<td></td>
</tr>
<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>DPP2_RF_model</td>
<td>≤ 6</td>
<td></td>
</tr>
<tr>
<td>DPP8_RF_model</td>
<td>≤ 6</td>
<td></td>
</tr>
<tr>
<td>DPP9_RF_model</td>
<td>≤ 6</td>
<td></td>
</tr>
<tr>
<td>hERG pIC50</td>
<td>≤ 5</td>
<td></td>
</tr>
</tbody>
</table>

- Primary activity
- Good physchem properties
- Avoid off targets
Starting Point
CHEMBL1929395

CHEMBL1929395

DPP4_RF_model: 7.631  DPP Profile: 0.04626
DPP2_RF_model: 5.452
DPP8_RF_model: 7.074
DPP9_RF_model: 7.285
Guided Optimisation
Applying captured knowledge

- Cyanopyrroloidine conserved
- Selection applied at each generation based on multi-parameter profile
Results

~16,000 compounds explored
Results
~16,000 compounds explored
Results
Example compounds

CHEMBL1929395

DPP Profile: 0.4807

DPP Profile: 0.04634

DPP Profile: 0.5039

DPP Profile: 0.4918
Exploration of Chemical Space

DPP-4 Space

- DPP-4 compounds in ChEMBL
- CHEMBL192395
- Generated compounds
Conclusions

• Capturing and applying knowledge computationally enables broad exploration of new optimisation strategies
  – ‘Think outside the box’
  – Rigorous investigation of possibilities

• Chemistry knowledge
  – Transformations representing optimisation strategies used in the past

• SAR knowledge
  – Machine learning to build QSAR models trained on relevant data

• Project knowledge
  – Expertise of project team from different disciplines as a scoring profile

• Integrated and applied using an evolutionary algorithm

• For more information and references, please visit:
  – www.optibrium.com/stardrop/
  – www.optibrium.com/community/
Acknowledgements

• The Optibrium team, including
  - Chris Leeding
  - James Chisholm
  - Alex Elliott
  - Fayzan Ahmed
  - Coran Hoskin
  - Aishling Cooke

• Istvan Ujváry, Julian Hayward, John Barnard
  - BIOSTER