Fragment-Based Screening, What can we learn from published hits?

Chris Swain
Cambridge MedChem Consulting
Fragment-Based Screening

- Fragment-based screening has become increasingly popular and has proven to be a viable alternative to high-throughput screening.
- Fragment space is smaller
  - A million compounds cover only a small fraction of the suggested $10^{60}$ Chemical Space, whilst 2000 compounds can probe much of the $10^6$ Fragment Space
- Hit rates for Fragment-based screening appear to be higher, typically 3-10%.
- Binding Efficiency for small molecules is likely to be higher.
Design of the Fragment Library

• Several approaches have been described in the design of fragment libraries. Most comply with the commonly accepted Astex "Rule-of-Three"
• Solubility is key requirement since screening carried out at higher concentrations
  – Often overlooked
• Rather than simply cull available molecules there have been recent attempts to design libraries based on known drugs, PDB ligands, natural products, or enhanced 3D structure.
What can we learn from known fragment hits?

• Compile database of published hits from fragment screens.

• Include:-
  – Screening technology
  – Target and Uniprot ID
  – Target type, using ChEMBL ontology

• Calculate
  – Physicochemical properties
  – LogP, LogD, PSA, HBA, HBD, RotB, pKa, shape descriptors, MR, HAC, fraction aromatic. (ChemAxon, MOE)
  – Functional groups (Checkmol)
Current Status (4 November 2013)

- 165 Publications
- 620 Published hits
- 116 Different targets
- 19 Detection technologies

- Finding the data is getting more of a challenge, it seems as fragment screening becomes more mainstream it is often not mentioned in the title or abstract.
Diversity

- Clustered using MACCS fingerprints in MOE. Tanimoto 0.85
- Majority are singletons
- Diverse fragments for same target
Suppliers of hits

<table>
<thead>
<tr>
<th>Supplier Name</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>MaybridgeAll.mdb</td>
<td>158</td>
</tr>
<tr>
<td>Maybridge_2500_Feb2013.mdb</td>
<td>96</td>
</tr>
<tr>
<td>LifeChemicals_frsags.mdb</td>
<td>47</td>
</tr>
<tr>
<td>Otava.mdb</td>
<td>44</td>
</tr>
<tr>
<td>Specs.mdb</td>
<td>41</td>
</tr>
<tr>
<td>KeyOrganicsAll.mdb</td>
<td>37</td>
</tr>
<tr>
<td>Enamine_frsags.mdb</td>
<td>37</td>
</tr>
<tr>
<td>Selcia_Frag_Library.mdb</td>
<td>30</td>
</tr>
<tr>
<td>Prestwick.mdb</td>
<td>29</td>
</tr>
<tr>
<td>Vitas.mdb</td>
<td>24</td>
</tr>
<tr>
<td>ChemDiv_Feb2013.mdb</td>
<td>19</td>
</tr>
<tr>
<td>ChemX.mdb</td>
<td>15</td>
</tr>
<tr>
<td>TimTec.mdb</td>
<td>13</td>
</tr>
<tr>
<td>Chembridge_frsags_feb2013.mdb</td>
<td>9</td>
</tr>
<tr>
<td>KeyOrganicsBionetPrem.mdb</td>
<td>8</td>
</tr>
<tr>
<td>Enamine_Golden.mdb</td>
<td>4</td>
</tr>
<tr>
<td>LCZenobia.mdb</td>
<td>3</td>
</tr>
<tr>
<td>Asinex.mdb</td>
<td>3</td>
</tr>
<tr>
<td>3DFragConsortium.mdb</td>
<td>1</td>
</tr>
<tr>
<td>WuXi_frsags_Feb2013.mdb</td>
<td>0</td>
</tr>
<tr>
<td>Pyxis.mdb</td>
<td>0</td>
</tr>
<tr>
<td>Infomatik3D.mdb</td>
<td>0</td>
</tr>
<tr>
<td>Analyticon.mdb</td>
<td>0</td>
</tr>
</tbody>
</table>

Maybridge are by far the most popular supplier
First major supplier to check solubility of fragments
Functional Group Analysis

- 590/620 contain an aromatic ring, 488 of which are heterocyclic
- 131/620 contain an arylhalide
- 117 contain an acidic group, 103 a basic group
- 15 contain a nitro group
- 115 contain a hydroxy, 72 an ether
- 231 contain an amine, 120 “anilines”
- 76 amides, 29 esters, 15 ureas
Most common scaffolds
You can only test what is available

- Some papers describe the source of the screening compounds, many do not.
- Looking at the hits we can make a guess at the likely source of the screening collection used.
- Use same tools to calculate profile of putative screening compounds.
Comparison of Molecular Weight

“Screening Collection”

Hits
Comparison of ionisation

“Screening Collection”

Hits
Comparison of Aromaticity

“Screening Collection”

Hits
Comparison of Shape
Conclusions

- Published fragments are lower molecular weight
- They contain a greater proportion of ionisable groups
- They contain a greater proportion of aromatics rings
- They contain a greater proportion of “disc-like” shaped molecules
- The role of increased 3D shape is unproven.
Detection technology and target type

![Chart showing frequency of various detection technologies](chart-image)

![Chart showing number of hits for different target types](chart-image)
Choice of technology

![Graph showing the relative throughput and protein yield for different technologies: Enzyme activity, SPR, Thermal shift, Capillary electroph., Ligand NMR, Mass spec, Protein crystals, Protein NMR. The graph is plotted on a log-log scale, with protein yield on the x-axis and relative throughput on the y-axis.](image-url)
Detection Technology
Detection Technology

- Evidence from literature that different technologies can identify hits for a single target.
- No evidence that detection technology influences the physiochemical properties of the hits identified.
  - Some technologies (e.g. SPR) are thought to have a higher false positive rate.
Multiple targets

- Over 80 fragment hits have been shown to be active against multiple targets.
- Whilst a few are active against similar targets (e.g. kinases), many are active against seemingly unrelated proteins.
Fragments active against multiple targets

- CDK2
- DNA Gyrase
- Factor Xa
- Urokinase
- Tryptase
- Thrombin
- Phenylethanolamine N-methyltransferase
- Urokinase
- Tryptase
- MMP-2
- Anthrax lethal factor
- Tyrosinase
- Stromelysin
- Inositol-3-phosphate synthase
- Thymidylate synthase
- Inositol-3-phosphate synthase
- ASIC3
- HIV Integrase
- Trypanosoma brucei Choline Kinase
- HIV-1 Integrase-Lens Epithelium-Derived Growth Factor/p75 (IN-LEDGF/p75) Interaction
- MMP-2
Effect of pKa and Target Type

Ion Channel and GPCR no acids but number of basic

PPI mainly acids

Enzymes mainly neutral
Target type physicochemical properties

- Range of molecular weight
- Number of aromatic atoms
- Range of LogP
- Range of LogD
Conclusions

- Fragment screening hits tend to be lower molecular weight, contain aromatic rings and ionizable groups.
- Some targets (GPCR, Ion channels, PPI) select for specific physicochemical properties.
- Detection technology does not appear to influence properties of hits identified.
Ongoing work..

- This work is part of an ongoing collaboration with Cheminformatics groups at University of Cambridge and The Institute of Cancer Research, London