PERFORMANCE
OF
SUB-PHARMACOPHORE MODELS
AS SEEDS IN
DRUG DISCOVERY

Workshop in
Cheminformatics
Sep 2014
EPFL Lausanne

Modest Korff, Joel Freyss, Thomas Sander
TOPICS

- Task
- Flexophore descriptor
- Sub-Flexophores
- Improving pharmacophore models
- Results and conclusions
TASK

Improve pharmacophore models for successful virtual screening
THE FLEXOPHORE DESCRIPTOR

Complete graph
Summarized

Nodes:
Pharmacophore points

Edges:
Distance histograms

Frequency
Distance bins
FLEXOPHORE COMPARISON

Similarity via interaction at protein
SIMILARITY METRIC FOR BIOISOSTERS

Protein-ligand interaction statistics

<table>
<thead>
<tr>
<th>PP Types</th>
<th>1.</th>
<th>2.</th>
<th>3.</th>
<th>4.</th>
<th>5.</th>
<th>6.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. C Alkane (prim.)</td>
<td>1.00</td>
<td>0.72</td>
<td>0.83</td>
<td>0.36</td>
<td>0.37</td>
<td>0.30</td>
</tr>
<tr>
<td>2. C Alkane (sec.)</td>
<td>0.72</td>
<td>1.00</td>
<td>0.51</td>
<td>0.38</td>
<td>0.40</td>
<td>0.38</td>
</tr>
<tr>
<td>3. C Carbonyl</td>
<td>0.83</td>
<td>0.51</td>
<td>1.00</td>
<td>0.41</td>
<td>0.45</td>
<td>0.45</td>
</tr>
<tr>
<td>4. O Carbonyl</td>
<td>0.36</td>
<td>0.38</td>
<td>0.41</td>
<td>1.00</td>
<td>0.98</td>
<td>0.88</td>
</tr>
<tr>
<td>5. O Ether</td>
<td>0.37</td>
<td>0.40</td>
<td>0.45</td>
<td>0.98</td>
<td>1.00</td>
<td>0.85</td>
</tr>
<tr>
<td>6. O Alcohol</td>
<td>0.30</td>
<td>0.38</td>
<td>0.45</td>
<td>0.88</td>
<td>0.85</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Sim$_{A,B}$ = f(Pharmacophore point pairs bioisosterism)
HOW TO IMPROVE A PHARMACOPHORE MODEL?

Two changeable components

- Pharmacophore points
  - Substructure based
  - Easy to manipulate on/off

- Distance histograms
  - Edges of complete graph \(((n \times n) - n) / 2\)
  - I.e. 7 PPNodes → 21 distance histograms
  - Histograms, not ranges
  - Meaningful manipulation difficult
SUB PHARMACOPHORES
COMBINATORIAL EXPLOSION

Possible combinations \( \binom{n}{k} \)

<table>
<thead>
<tr>
<th># PPPoints in Sub Pharmacophore</th>
<th>PPPPoints 4</th>
<th>PPPPoints 6</th>
<th>PPPPoints 8</th>
<th>PPPPoints 10</th>
<th>PPPPoints 12</th>
<th>PPPPoints 14</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>4</td>
<td>20</td>
<td>56</td>
<td>120</td>
<td>220</td>
<td>364</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>6</td>
<td>56</td>
<td>252</td>
<td>792</td>
<td>2002</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td>8</td>
<td>120</td>
<td>792</td>
<td>3432</td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td>10</td>
<td>220</td>
<td>2002</td>
</tr>
<tr>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>12</td>
<td>364</td>
</tr>
</tbody>
</table>

How many subpharmacophores?
HOW TO MEASURE IMPROVEMENT?

Need for validation

DUD dataset

- 39 targets
- 241 data sets
- 10-50 seeds (actives)
- Filled up to 1000 with decoys (assumed inactives)
  - Physico chemical properties like seeds

TEST SETUP

Select molecule

- Seeds (actives)
- Decoys (inactives)

Split

Train (500)  Test (500)
RUNNING A TEST

Probe SubFlexophores

Compare SubFlexophores with train Flexophores

Compare best performing SubFlexophores with test Flexophores

Calculate enrichment
RESULTS ALL TESTS

- Data sets 241
- Probe SubFlexophores: 20-200 per probe molecule
- Compare to train
- Top SubFlexophores 1-10 per probe molecule
- Compare to test

Summary

Enrichment SubFlexophore test 127

Enrichment full Flexophore test 86

Best SubFlexophore better than full Flexophore 91
SUCCESSFUL SCAFFOLD HOPPING

Query molecule

Found seeds
FLEXOPHORE EDITOR: PLAYGROUND FOR SCIENTISTS
IMPLEMENTATION DETAILS

- In-house development
- Java programming language
- Jmol for pharmacophore visualization
- Multi thread similarity calculations

Training data

SubFlexophores

Pharmacophore editor
CONCLUSIONS

- Impossible to outsmart combinatorial explosion of possible pharmacophore models
- Medicinal chemist may construct test and validation data
- Combination of brute force computer power plus medicinal chemist know how resulted in improved pharmacophore models
THANK YOU!