Enrichments and Rescoring

Trent Balius
AMS 535 / CHE 535
Directory of Useful Decoys

Benchmarking Sets for Molecular Docking

Niu Huang, Brian K. Shoichet, and John J. Irwin

J. Med. Chem., 2006, 49 (23), 6789-6801
Outline

• Introduction
  – Docking Introduction
  – Docking Validations
  – Enrichment
• DUD Background
• DUD Enrichments
• DUD Cross-Enrichments
• Binding pose predictions
• Conclusions
Introduction
Docking Introduction

- Objectives of Docking programs
  - generate binding modes (or poses)
  - select the true pose out all poses generated with a scoring function

- Uses of Docking programs
  - pose reproduction
    (pdb of receptor but not of complex)
  - virtual screening
    find a new drug lead by screening virtual databank (e.g. ZINC)
Docking Validations Studies

• Pose reproduction:
  – regenerating the known binding mode of a ligand in the context of the protein with a docking program
  – protein-ligand complex needed

• Enrichments:
  – after docking a database of known actives and decoys the actives are top scoring
  – protein structure needed
Enrichment

- **Positives**: 100%
- **Negatives**: 50%

**Score**
- **Positives**: 80%
- **Negatives**: 20%

**Throw away**
Enrichment Studies

unknowns may have activity

Small molecule database seed with actives

Score Sort

↑ Keep

Threshold

↓ Throw away
Enrichment Studies

• Active and inactive is not known
  – Why not run an assay on all the small molecules?
    • expensive
    • takes time
    • multiple levels of experiments (needs to compare several assays. e.g. HTPS)
  – Seed the population with known active compounds
  – See how many bubble to the top.

• Enrichment curves

• Receiver operating characteristic (ROC) curves
DUD Background
Overview of DUD

• Directory of Useful Decoys (DUD)
  – used for enrichment studies
  – # of systems = 40 targets (proteins)
  – # of ligands = 2,950 molecules (actives)
  – # of decoys = 98,266 (presumed non-binders)
  – every active molecule has 36 decoys
    • $36 \times 2950 = 106,200 \neq 98,266$ because there are some decoys shared among ligands.
  • decoys are physically similar
  • topologically distinct

J. Med. Chem. 2006, 49, 6789-6801
Overview of DUD (cont'd)

• Directory of Useful Decoys (DUD)
  – Systems chosen for the following reasons:
    availability of annotated ligands
    crystal structures
    previous docking studies
  – Designed to remove sorting bias on "gross features"
  – Decoys are "chemically distinct" from active ligand "unlikely binders"

J. Med. Chem. 2006, 49, 6789-6801
Protocol for DUD prep

Annotated ligands + ZINC (~ 3.5 million compounds)

2D dissimilarity analysis

Annotated ligands + 1.5 million of ZINC compounds (Tc < 0.9* against any known ligands)

1D similarity analysis (MW, HBacc, HBdon, LogP and RB)

Annotated ligands + DUD decoys (36 decoys per ligand)

J. Med. Chem. 2006, 49, 6789-6801
Tanimoto Coefficient

\[ Tc = \frac{|A \cap B|}{|A \cup B|} \]

- intersection is \# of ON bits common in both A and B
- union is \# of ON bits present in either A or B
- Examples of Daylight Fingerprint descriptors:
  - ring systems
  - common functional groups
  - which elements are present
  - unusual electronic configurations.
## DUD systems

### Table 1. Enrichments of the Annotated Ligands Using the Decoys in DUD for Forty Targets by Docking$^a$

<table>
<thead>
<tr>
<th>protein</th>
<th>PDB code</th>
<th>resolution (Å)</th>
<th>no. of ligands$^b$</th>
<th>no. of decoys</th>
<th>EF$_{max}$</th>
<th>EF$_1$</th>
<th>EF$_{20}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nuclear Hormone Receptors</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. AR</td>
<td>1xq2</td>
<td>1.9</td>
<td>74 (a,b)</td>
<td>2630</td>
<td>60.2</td>
<td>33.5</td>
<td>3.8</td>
</tr>
<tr>
<td>2. ER$_{agonist}$</td>
<td>112i</td>
<td>1.9</td>
<td>67 (a-c)</td>
<td>2361</td>
<td>29.6</td>
<td>19.2</td>
<td>4.5</td>
</tr>
<tr>
<td>**3. ER$_{antagonist}$</td>
<td>3ert</td>
<td>1.9</td>
<td>39 (a–d)</td>
<td><strong>1399</strong></td>
<td><strong>101.6</strong></td>
<td><strong>12.7</strong></td>
<td><strong>1.3</strong></td>
</tr>
<tr>
<td>4. GR</td>
<td>1m2z</td>
<td>2.5</td>
<td>78 (a)</td>
<td>2804</td>
<td>31.7</td>
<td>8.9</td>
<td>1.4</td>
</tr>
<tr>
<td>5. MR</td>
<td>2aa2</td>
<td>1.9</td>
<td>15 (a)</td>
<td>535</td>
<td>330.0</td>
<td>46.2</td>
<td>3.7</td>
</tr>
<tr>
<td>6. PPARg</td>
<td>1fm9</td>
<td>2.1</td>
<td>81 (a)</td>
<td>2910</td>
<td>1.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>7. PR</td>
<td>1sr7</td>
<td>1.9</td>
<td>27 (a)</td>
<td>967</td>
<td>2.9</td>
<td>0.0</td>
<td>2.0</td>
</tr>
<tr>
<td>8. RXRa</td>
<td>1mvc</td>
<td>1.9</td>
<td>20 (a)</td>
<td>708</td>
<td>148.5</td>
<td>24.8</td>
<td>2.2</td>
</tr>
<tr>
<td>Kinases</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. CDK2</td>
<td>1ckp</td>
<td>2.1</td>
<td>50 (e,f)</td>
<td>1780</td>
<td>19.8</td>
<td>13.9</td>
<td>1.4</td>
</tr>
<tr>
<td>10. EGFr</td>
<td>1m17</td>
<td>2.6</td>
<td>416 (g)</td>
<td>14914</td>
<td>3.8</td>
<td>2.1</td>
<td>2.4</td>
</tr>
<tr>
<td>11. FGFr1</td>
<td>1agw</td>
<td>2.4</td>
<td>118 (g)</td>
<td>4216</td>
<td>1.0</td>
<td>0.0</td>
<td>0.2</td>
</tr>
<tr>
<td>12. HSP90</td>
<td>1uy6</td>
<td>1.9</td>
<td>24 (h)</td>
<td>861</td>
<td>10.8</td>
<td>8.6</td>
<td>2.0</td>
</tr>
<tr>
<td>**13. P38 MAP</td>
<td>1kv2</td>
<td><strong>2.8</strong></td>
<td><strong>234 (g)</strong></td>
<td><strong>8399</strong></td>
<td><strong>4.1</strong></td>
<td><strong>2.1</strong></td>
<td><strong>2.4</strong></td>
</tr>
<tr>
<td>14. PDGFrβ</td>
<td>model</td>
<td>n/a</td>
<td>157 (g)</td>
<td>5625</td>
<td>1.2</td>
<td>0.0</td>
<td>0.6</td>
</tr>
<tr>
<td>15. SRC</td>
<td>2src</td>
<td>1.5</td>
<td>162 (g)</td>
<td>5801</td>
<td>3.1</td>
<td>1.2</td>
<td>1.5</td>
</tr>
<tr>
<td>**16. TK</td>
<td>1kim</td>
<td><strong>2.1</strong></td>
<td><strong>22 (a,d,i)</strong></td>
<td><strong>785</strong></td>
<td><strong>63.0</strong></td>
<td><strong>54.0</strong></td>
<td><strong>5.0</strong></td>
</tr>
<tr>
<td>17. VEGFr2</td>
<td>1vr2</td>
<td>2.4</td>
<td>74 (j)</td>
<td>2647</td>
<td>2.2</td>
<td>1.3</td>
<td>1.4</td>
</tr>
<tr>
<td>Serine Proteases</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18. FXa</td>
<td>1f0r</td>
<td>2.7</td>
<td>142 (e,f,k)</td>
<td>5102</td>
<td>34.9</td>
<td>14.6</td>
<td>3.8</td>
</tr>
<tr>
<td>19. thrombin</td>
<td>1ba8</td>
<td>1.8</td>
<td>65 (e,l,m)</td>
<td>2294</td>
<td>18.3</td>
<td>13.7</td>
<td>2.9</td>
</tr>
<tr>
<td>20. trypsin</td>
<td>1bju</td>
<td>1.8</td>
<td>43 (e,l)</td>
<td>1545</td>
<td>22.5</td>
<td>22.5</td>
<td>2.6</td>
</tr>
</tbody>
</table>

J. Med. Chem. 2006, 49, 6789-6801
DUD systems (cont'd)

<table>
<thead>
<tr>
<th>protein</th>
<th>PDB code</th>
<th>resolution (Å)</th>
<th>no. of ligands$^b$</th>
<th>no. of decoys</th>
<th>EF$_{\text{max}}$</th>
<th>EF$_1$</th>
<th>EF$_{20}$</th>
</tr>
</thead>
</table>

**Metalloenzymes**

<p>| | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>21</td>
<td>ACE</td>
<td>1o86</td>
<td>2.0</td>
<td>49 (a,m)</td>
<td>1728</td>
<td>141.4</td>
<td>40.4</td>
</tr>
<tr>
<td>22</td>
<td>ADA</td>
<td>1stw</td>
<td>2.0</td>
<td>23 (a,e)</td>
<td>822</td>
<td>21.5</td>
<td>12.9</td>
</tr>
</tbody>
</table>

**Folate Enzymes**

<p>| | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>DHFR</td>
<td>3dfr</td>
<td>1.7</td>
<td>201 (m)</td>
<td>7150</td>
<td>28.7</td>
<td>21.7</td>
</tr>
<tr>
<td>26</td>
<td>GART</td>
<td>1c2t</td>
<td>2.1</td>
<td>21 (n)</td>
<td>753</td>
<td>70.7</td>
<td>42.4</td>
</tr>
</tbody>
</table>

**Other Enzymes**

<p>| | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>27</td>
<td>AChE</td>
<td>1eve</td>
<td>2.5</td>
<td>105 (a,e,m)</td>
<td>3732</td>
<td>3.1</td>
<td>1.9</td>
</tr>
<tr>
<td>28</td>
<td>ALR2</td>
<td>1ah3</td>
<td>2.3</td>
<td>26 (o)</td>
<td>920</td>
<td>76.2</td>
<td>38.1</td>
</tr>
</tbody>
</table>

**1NDW → 1Q4G**

<p>| | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>29</td>
<td>AmpC</td>
<td>1xgj</td>
<td>2.0</td>
<td>21 (p)</td>
<td>734</td>
<td>23.6</td>
<td>17.1</td>
</tr>
<tr>
<td>30</td>
<td>COX-1</td>
<td>1p4g</td>
<td>2.1</td>
<td>25 (i)</td>
<td>850</td>
<td>9.9</td>
<td>4.0</td>
</tr>
<tr>
<td>31</td>
<td>COX-2</td>
<td>1cx2</td>
<td>3.0</td>
<td>349 (c,f,m)</td>
<td>12491</td>
<td>29.1</td>
<td>20.1</td>
</tr>
<tr>
<td>32</td>
<td>GPB</td>
<td>1a8i</td>
<td>1.8</td>
<td>52 (e,m)</td>
<td>1851</td>
<td>28.6</td>
<td>22.8</td>
</tr>
<tr>
<td>33</td>
<td>HIVPR</td>
<td>1hpx</td>
<td>2.0</td>
<td>53 (a,e)</td>
<td>1888</td>
<td>9.3</td>
<td>3.7</td>
</tr>
<tr>
<td>34</td>
<td>IIIVRT</td>
<td>1rt1</td>
<td>2.6</td>
<td>40 (q)</td>
<td>1439</td>
<td>49.5</td>
<td>5.0</td>
</tr>
<tr>
<td>35</td>
<td>HMGR</td>
<td>1hw8</td>
<td>2.1</td>
<td>35 (a,i)</td>
<td>1242</td>
<td>198.0</td>
<td>33.9</td>
</tr>
<tr>
<td>36</td>
<td>InhA</td>
<td>1p44</td>
<td>2.7</td>
<td>85 (r)</td>
<td>3043</td>
<td>1.0</td>
<td>0.0</td>
</tr>
</tbody>
</table>

**36. InhA**

<p>| | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>37</td>
<td>NA</td>
<td>1a4g</td>
<td>2.2</td>
<td>49 (c,e,i)</td>
<td>1745</td>
<td>0.6</td>
<td>20.2</td>
</tr>
<tr>
<td>38</td>
<td>PARP</td>
<td>1efy</td>
<td>2.2</td>
<td>33 (s)</td>
<td>1178</td>
<td>6.3</td>
<td>6.0</td>
</tr>
<tr>
<td>39</td>
<td>PNP</td>
<td>1b8o</td>
<td>1.5</td>
<td>75 (e,t)</td>
<td>884</td>
<td>158.4</td>
<td>31.7</td>
</tr>
<tr>
<td>40</td>
<td>SAIII</td>
<td>1a7u</td>
<td>2.8</td>
<td>33 (i)</td>
<td>1159</td>
<td>120.0</td>
<td>78.0</td>
</tr>
</tbody>
</table>

*J. Med. Chem. 2006, 49, 6789-6801*
Six DUD systems

<table>
<thead>
<tr>
<th>protein</th>
<th>PDB code</th>
<th>resolution (Å)</th>
<th>no. of ligands</th>
<th>no. of decoys</th>
<th>EFmax</th>
<th>EF1</th>
<th>EF20</th>
</tr>
</thead>
<tbody>
<tr>
<td>ER antagonist</td>
<td>3ert</td>
<td>1.9</td>
<td>39</td>
<td>1399</td>
<td>101.6</td>
<td>12.7</td>
<td>1.3</td>
</tr>
<tr>
<td>P38 MAP</td>
<td>1kv2</td>
<td>2.8</td>
<td>234</td>
<td>8399</td>
<td>4.1</td>
<td>2.1</td>
<td>2.4</td>
</tr>
<tr>
<td>TK</td>
<td>1kim</td>
<td>2.1</td>
<td>22</td>
<td>785</td>
<td>63.0</td>
<td>54.0</td>
<td>5.0</td>
</tr>
<tr>
<td>ADA</td>
<td>1stw</td>
<td>2.0</td>
<td>23</td>
<td>822</td>
<td>21.5</td>
<td>12.9</td>
<td>2.4</td>
</tr>
<tr>
<td>ALR2</td>
<td>1ah3</td>
<td>2.3</td>
<td>26</td>
<td>920</td>
<td>76.2</td>
<td>38.1</td>
<td>2.3</td>
</tr>
<tr>
<td>InhA</td>
<td>1p44</td>
<td>2.7</td>
<td>85</td>
<td>3043</td>
<td>1.0</td>
<td>0.0</td>
<td>0.3</td>
</tr>
</tbody>
</table>

- Families chosen for the following reasons:
  - ER and TK -- strong ligand enrichment and substantial number of published docking studies
  - P38 MAP kinase -- poorly performing protein kinases
  - ADA -- failed with the fully automated docking engine and rescued by the semiautomated procedure
  - ALR2 -- intermediate enrichment.
  - InhA -- failure of the docking method

*J. Med. Chem. 2006, 49, 6789-6801*
Molecule Properties

The physical property distributions

- brown -- annotated ligands (2950 compounds)
- blue -- the DUD decoys (95,316 compounds)
- green -- properties of the MDDR database (98,000 compounds)
- orange -- Jain’s decoys (randomly selected 1000 ZINC druglike compounds)
- cyan -- Rognan’s decoys (randomly selected 990 ACD compounds).

• The physical property
  • # of HB acceptors
  • # of HB donors
  • xlogp
  • Molecular Weight
  • # of rotatable bonds

supporting material
J. Med. Chem. 2006, 49, 6789-6801
Automated Docking Pipeline

supporting material

J. Med. Chem. 2006, 49, 6789-6801
DUD Enrichments
$EF_{\text{subset}} = \frac{\frac{\text{ligands}_{\text{selected}}}{N_{\text{subset}}}} {\frac{\text{ligands}_{\text{total}}}{N_{\text{total}}}}$

$= \frac{\text{ligands}_{\text{selected}}}{\text{ligands}_{\text{total}}} \times \frac{N_{\text{total}}}{N_{\text{subset}}}$

$EF_1 = \frac{\text{ligands}_{\text{selected, top1\% database}}}{\text{ligands}_{\text{total}}} \times \frac{100}{1}$

$EF_{20} = \frac{\text{ligands}_{\text{selected, top20\% database}}}{\text{ligands}_{\text{total}}} \times \frac{100}{20}$
The docking ranked database

the percentage of known ligands found

six representative systems are highlighted in light yellow.

gray -- random

blue DUD database (98 266 compounds)

red target subset decoy

ROC curves

\[ TP_{Rate} = Se_{subset} = \frac{\text{ligands}_{selected}}{\text{ligands}_{total}} \]

\[ FP_{Rate} = (1 - Sp)_{subset} = \frac{\text{decoys}_{selected}}{\text{decoys}_{total}} \]

Se - Sensitivity, Sp - Specificity
Computational Prediction vs. Experimental Evidenced

<table>
<thead>
<tr>
<th></th>
<th>Activity</th>
<th>Inactivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predicted Activity</td>
<td>True Positive</td>
<td>False Positive</td>
</tr>
<tr>
<td>Predicted Inactivity</td>
<td>False Negative</td>
<td>True Negative</td>
</tr>
</tbody>
</table>
ROC curves

- ROC -- Receiver Operating Characteristic

http://www.anaesthetist.com/mnm/stats/roc/Findex.htm
ROC curves

ROC curves for 12 targets
DUD (blue)
MDDR (green)
Jain’s decoys (orange)
Rognan’s decoys (cyan)
random (gray)
DUD Cross-Enrichments
Cross-Docking

<p>| | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="green-c.png" alt="" /></td>
<td><img src="red-c.png" alt="" /></td>
<td>...</td>
<td><img src="blue-c.png" alt="" /></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><img src="green-c.png" alt="" /></td>
<td><img src="red-c.png" alt="" /></td>
<td>...</td>
<td><img src="blue-c.png" alt="" /></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><img src="green-c.png" alt="" /></td>
<td><img src="red-c.png" alt="" /></td>
<td>...</td>
<td><img src="blue-c.png" alt="" /></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><img src="green-c.png" alt="" /></td>
<td><img src="red-c.png" alt="" /></td>
<td>...</td>
<td><img src="blue-c.png" alt="" /></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

![Image of cross-docking](https://via.placeholder.com/150)
"Cognate" Enrichment Study
Cross-Enrichments
Cross-Enrichments

Matrix of Cross-Enrichments

Very good (black): 
ET\text{max} \geq 30 \text{ and } ET\text{20} \geq 3

good (red): 
30 > ET\text{max} \geq 20 \text{ and } 3 > ET\text{20} \geq 2.5

medium (green): 
20 > ET\text{max} \geq 10 \text{ and } 2.5 > ET\text{20} \geq 2

poor (white): 
ET\text{max} < 10 \text{ and } ET\text{20} < 2

boxes are drawn around related targets.

J. Med. Chem. 2006, 49, 6789-6801
Matrix of Cross-Enrichments

Very good (black):
ETmax ≥ 30 and
ET20 ≥ 3

good (red):
30 > ETmax ≥ 20 and
3 > ET20 ≥ 2.5

medium (green):
20 > ETmax ≥ 10 and
2.5 > ET20 ≥ 2

poor (white):
ETmax < 10 and ET20 < 2

boxes are drawn around related targets.

J. Med. Chem. 2006, 49, 6789-6801
# Statistics and Timings

**Table 3.** Docking Statistics on Six Representative Targets

<table>
<thead>
<tr>
<th>receptor</th>
<th>unique molecules scored(^a)</th>
<th>total molecules scored(^b)</th>
<th>orientations sampled per molecule</th>
<th>conformations sampled per molecule</th>
<th>total configurations scored(^b)</th>
<th>total time (h)(^c)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ER</td>
<td>97 427</td>
<td>416 990</td>
<td>1 895</td>
<td>6 543</td>
<td>2.69 × 10(^{10})</td>
<td>54.4</td>
</tr>
<tr>
<td>P38 MAP</td>
<td>93 887</td>
<td>294 917</td>
<td>592</td>
<td>7 875</td>
<td>8.97 × 10(^{9})</td>
<td>20.1</td>
</tr>
<tr>
<td>TK</td>
<td>37 240</td>
<td>180 451</td>
<td>3 437</td>
<td>4 302</td>
<td>2.67 × 10(^{9})</td>
<td>21.9</td>
</tr>
<tr>
<td>ADE</td>
<td>85 053</td>
<td>297 400</td>
<td>14 632</td>
<td>5 308</td>
<td>2.19 × 10(^{10})</td>
<td>65.5</td>
</tr>
<tr>
<td>ALR2</td>
<td>98 724</td>
<td>430 313</td>
<td>4 272</td>
<td>10 109</td>
<td>1.44 × 10(^{11})</td>
<td>296.4</td>
</tr>
<tr>
<td>InhA</td>
<td>97 668</td>
<td>429 579</td>
<td>2 325</td>
<td>6 809</td>
<td>5.87 × 10(^{10})</td>
<td>123.5</td>
</tr>
</tbody>
</table>

\(^a\) Only orientations and configurations passing the steric filter were scored. \(^b\) Some molecules were represented in the database in multiple rigid fragment, protonation, and tautomeric forms. \(^c\) Scaled to reflect time on a 2800-MHz Pentium IV.
Binding pose predictions

- six representative targets
- docked ligands (green)
- crystallographic structures (colored by atom type)
- Key hydrogen bonds (yellow)

Conclusions

• DUD is designed to match physical properties of active ligands

• Other databases used in enrichment studies are more physically dissimilar from the actives

• DUD gives poorer enrichment over other databases
  – better to gauge a docking program's abilities

• Most systems have no cross-enrichment with notable exceptions including TK
Rescores

http://dud.docking.org/
DUD Release 2: http://dud.docking.org/r2/

Notes accompanying release 2 as found on http://dud.docking.org/r2/

"Why is the ratio of decoys to annotated ligands described as 36 to 1 in the paper, yet there are on average only 33 to 1 in DUD? This is due to overlap, as the same decoy could be used for multiple targets, particularly in the kinase class where there was so much overlap.

Two DUD decoy compounds (ZINC154632 for RXR decoys and ZINC608655 for ER decoys) were structurally identical/similar to the crystal ligands of RXR and ER, individually. This problem was caused by failing to include the crystallographic ligands in our annotated ligands set, and will be fixed in the next version of DUD. Thanks to Paul Hawkins of OpenEye for bringing this to our attention.

Also: PDB code for COX-1 structure in given as 1P4G but should be 1Q4G. We regret this error, and thank alert reader Paul Hawkins of OpenEye for this information Also, Hao Li of UCSF Pharm Chem points out that the PDB id of ADA in the paper is wrong. It should be 1ndw."
Structural Interaction Fingerprint (SIFt): A Novel Method for Analyzing Three-Dimensional Protein-Ligand Binding Interactions

Zhan Deng, Claudio Chuaqui, and Juswinder Singh

SIFt Introduction

• Structural Interaction Fingerprints (SIFt)
• Identification of Ligand Binding Site Residues
  – non-hydrogen protein atoms solvent accessibility loss upon ligand binding
  – protein atoms h-bonding with the ligands
• Extraction and Classification of Binding Interactions

*J. Med. Chem. 2004, 47, 337-344*
SIFt Introduction

• Seven different types of interactions
  (1) residue is in contact with the ligand
  (2) backbone is in contact
  (3) sidechain is in contact
  (4) polar interaction
  (5) non-polar interaction
  (6) h-bond acceptor
  (7) h-bond donor

• Concatenating all figure prints together

*J. Med. Chem. 2004, 47, 337-344*
SIFt Introduction

• Three applications of SIFt in Drug Discovery:
  – sorting, clustering, and organizing docking poses (identifying like binding poses)
  – organizing and clustering 90 crystal complexes
  – filtering virtual screening results to find ligands with certain binding mode and interaction patterns

*J. Med. Chem.* 2004, 47, 337-344
Tanimoto Coefficient

\[ T_c = \frac{|A \cap B|}{|A \cup B|} \]

- intersection is # of ON bits common in both A and B
- union is # of ON bits present in either A or B

Docking studies

• Study #1 (single ligand)
  – ligand SB203580 docked to p38 (pdb code 1a9u)
  – poses generated with FlexX in Sybyl
  – 100 poses generated

• Study #2 (enrichment study)
  – 16 known p38 inhibitors
  – 1000 with diverse chemical structures
  – docked database to p38 (pdb code 1a9u)
  – 30 480 (30 1016) poses generated

*J. Med. Chem.* 2004, 47, 337-344
SB203580 Clusters in P38

Figure shows the 100 poses generated in Docking study #1, SB203580 docked to p38

SIFT Clusters

- Scores are not able to identify the binding mode (SIFT)

Enrichment

- comparison of SIFt with 2 alternative scoring functions
- SIFt gives good enrichment

Table 1. Comparison of the Database Enrichment Performances of SIFt with ChemScore and PMF Score

<table>
<thead>
<tr>
<th>filtering method</th>
<th>EF&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>PMF Score</td>
<td>2.0</td>
</tr>
<tr>
<td>ChemScore</td>
<td>5.4</td>
</tr>
<tr>
<td>SIFt</td>
<td>37.0</td>
</tr>
<tr>
<td>SIFt + ChemScore</td>
<td>42.3</td>
</tr>
</tbody>
</table>
Crystal Structure study

• Study #3 (Kinase family analysis)
  – 89 kinase-ligand complexes
    • inhibitor or substrate in ATP binding cleft
    • all active site residues are present in structure
  – 25 different kinases
  – 14 different protein kinase subfamilies
  – 54 unique compounds

Conclusions

• SIFt is a powerful tool
  – pose clustering
  – family clustering
  – filtering screening results

• possible improvements
  – incorporate more types of interactions in the fingerprint
  – uses only subset of residues
  – uses scaled numeric data representing interactions