Cross-mining in 3D-2D-1D the PDB, chemical libraries and structure activities to extract shared modes of binding for PDB ligand substructures

François Delfaud, MEDIT CEO
EuroQSAR - September 2014
Starting point = MED-SuMo = Heuristic to mine PDB in 3D

- Similar binding sites are likely to bind same ligands
- It’s more likely to occur with smaller interactions: fragments in (sub)-pockets

**MEDIT 2003-2011: explore multiple applications with MED-SuMo**

Since 2012, C2P project (Chemo-Proteomic Platform):

- Deconvolute Protein-Lig in Pocket-Frag interactions to clarify Affinity + Selectivity
- Mining together in 3D/2D/1D biostructures, SAR and chemical libraries to discover new applications: ex. shared modes of binding for PDB ligand substructures?
- Require powerful architecture to mine in 1D/2D/3D chemo-proteomic data
Introduction → [3Dpdb] mining with MED-SuMo

Functional Annotation:

3D superposition based on SCF (Structural Chemical Features) of the catalytic triad between Subtilisin and Thrombin (only 16% sequence identity)

Binding site characterization:

Comparing Thrombin (1DWC) to all serine proteases shows frequent or unique 3D interactions (Tyr60C Trp60D)

Classifying a SCOP superfamily:

**MEDP-SiteClassifier:**
- MED-SuMo pairwise comparison
- Matrix normalisation
- Markov Clustering

<table>
<thead>
<tr>
<th>Id</th>
<th>Cluster description</th>
<th>Cluster %</th>
<th>PDB</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>58%(22) DNA gyrase MutL</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>58%(15) Histidine Kinase</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>16%(6) DNA gyrase MutL</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>100%(78) HSP90</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>26%(10) DNA gyrase MutL</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>25%(1) BCK</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

= All bind to Radicicol drug

Full PDB classif.:
- Clusters
- Interfamily(Pfam/EC) pairs

http://80.15.144.179:49999/POPSWebPages/PF00225 PF00625

Drug repurposing:

1. PDB 3D binding site search:
   Query = Sorafenib / Braf kinase binding site

2. BumpScore QueryLigand in Hits
   ➞ among hits = Sorafenib co-crystallized with P38 kinase (24% seq Id)

Scaffolf Hopping between CDK2s:

1. PDB 3D binding site search
   Query = 2C6I (CDK2 kinase)

2. OverlapScore HitLigands in QueryProt.
   ➞ Triazolo Pyrimidin & Pyrazole superposition
Off target identification:

- Proteins with similar binding sites are likely to bind to similar ligands [Xie et al. Plos Comp. Biology SD supercomp.]
- Pentachlorophenol (PCP): herbicide, fungicide, disinfectant … but declined due to high toxicity. Classify in group B2 (probable human carcinogen) by US EPA agency
- MED-SuMo job from PCP pocket of succinate quinone oxidoreductase (2WDR)

Query: 2WDR (with PCP)

Hit list having similar 3D interaction surfaces with 2WDR

3HWC pocket: sharing arom., hydroph. and charges

The detected (no ligand co-cristallized in 3HWC) pocket of Chlorophenol-4-monooxygenase component 2 is likely to fit ligands like PCP
**Introduction → [2Dpdb-pubchem][3Dpdb] mining**

**Deconvolute Protein-ligand in protein-fragments interactions:**

- Pubchem small molecules
- PDB Protein-Ligands
- Substructure search in ligands
- Each hit = Protein-Fragment interaction =
  - fragment + dummy atoms (connections to the rest of the ligand)
  - corresponding SCF chemical features on protein surface

Comparing VEGF-r2 (2OH4) binding site toward this PDB Protein-Fragments database by MED-SuMo = Target-based library design:

- 50 best ranked intrafamily fragments (protein kinase but not vegfr)
- 50 best ranked interfamily fragments (not protein kinase)

### Combining in 3D superposed fragments

- detecting overlapping bond and rings (FC-Ligand software)
- FBDD application

<table>
<thead>
<tr>
<th>Frag</th>
<th>Hybrid .steps</th>
<th>3D filter</th>
<th>hybrid count</th>
<th>Murcko scaffold</th>
<th>PDB</th>
<th>Pub-Chem</th>
<th>PubChem bioassays</th>
</tr>
</thead>
<tbody>
<tr>
<td>2OH4 Tyr. kinase</td>
<td>1 474</td>
<td>4</td>
<td>218 085</td>
<td>9 936</td>
<td>175</td>
<td>549</td>
<td>27</td>
</tr>
<tr>
<td>2RH1 GPCR</td>
<td>400</td>
<td>4</td>
<td>3 818</td>
<td>729</td>
<td>67</td>
<td>110</td>
<td>43</td>
</tr>
<tr>
<td>1X88 Kinesin</td>
<td>48x2812</td>
<td>5</td>
<td>8676</td>
<td>6 696</td>
<td>1</td>
<td>229</td>
<td>39 BindingDB</td>
</tr>
</tbody>
</table>

Oguievetskaia K & Al, JCAMD, 23(8): 571-82 (2009)
Oriented hybridisation:
FBDD protocol on JAK2 kinase:

Adding 2Dfp similarity constraint to the SAR kinase Aureus database (< 50nM on JAK2 ; >100nM on JAK1 JAK3 and TYK2) upon each hybridisation step

Results = JAK2 selective Hybrids

Example:

Chemical Diversity of hybrids:
(FBDD on 2OH4 Tyr. Kinase):

PCA plots of scaffolds (ECFP scitegic fingerprint)
Attractive idea: extract shared modes of binding from PDB to propose 3D poses for non-cocrystalized ligand

Idea: use the current implemented Ligand deconvolution method (split ligand in multiple Pubchem small molecules)

NB: « PDB Fragment » includes many 2D fragment duplicates ... then, are we having conserved pocket interaction for those 2D duplicate fragments ?

1) PubchemCID=19 (2,3-dihydroxybenzoic) in 17 PDB files corresponding to 9 ligands

MED-SuMo detects superpositions of 3BY0(DBH) with 4 other highly similar DBH PDB files in 3 different pockets (Pi-stacking or double Lysine or hydroxyls)

But No binding site superposition detected between 3BY0(DBH) and 1HKL(113):
Results → MEDIT

[2D][2D][3Dpdb] mining

2) VERY Frequent PDB fragment:
Phenyl sulfonamide (Pubchem CID7370)
in 70 copies in the PDB

MED-SuMo superposition of Phenyl sulfonamide environment between 6 human CDK2

MED-SuMo superposition of Phenyl sulfonamide environment between 11 anhydrase (Zn)

(Proteins are hidden)
3) Frequent PDB LARGER scaffold: 
PubchemCID=1853 (phenanthridone) in 8 PDB files corresponding to 2 ligands P34+LDR

Despite low sequence identity, MED-SuMo detects a similar pocket (Hyd+2HBs) between 3CE0(P34) and 6 PDB.
Bioisostere pairs from the PDB:

PDB Protein-Ligands

Pubchem small molecules

MED-SuMo superposes 3D binding sites

Output ligands are converted in Frags (2D matches)

Seal score cutoff to detects overlapping Frag Query, Frag Hit

Can be mined in 2D

DB of pairs of 3D superposed fragments

Example:

Alkenyldiaryl methane (ADAM) = HIV non-nucleoside reverse transcriptase inhibitors

but metabolically unstable methyl-ester


$Seal(Frag_i, Frag_j) = \frac{2}{m+n} \sum_{i=1}^{m} \sum_{j=1}^{n} e^{-d_{ij}^2}$


Moriaud F & al, ACS Symposium 2011, series 1076, Chapter book 5:71-88
Results →

PDB-based bioisosteres compare to SAR-based bioisosteres:

1) Comparing to best SwissBiostere candidates with SBscore > 0.9

<table>
<thead>
<tr>
<th>B.Sites Pockets</th>
<th>Lig. Seal</th>
</tr>
</thead>
<tbody>
<tr>
<td>3PSO + 4F3I (100% seq. id.)</td>
<td><img src="image" alt="" /></td>
</tr>
<tr>
<td>3SV2 + 3P17 (99% seq. id.)</td>
<td><img src="image" alt="" /></td>
</tr>
<tr>
<td>3E51 + 3CWJ (100% seq. id.)</td>
<td><img src="image" alt="" /></td>
</tr>
</tbody>
</table>
2) Comparing to other SwissBiostere candidates (0.8 < SBscore > 0.9)

- **3MWU** + **3MA6** (50% seq. id.)
- **1SVG** + **1RE8** (92% seq. id.)
- **2IW9** + **1H08** (99% seq. id.)
- **3MWU** + **1QCF** (21% seq. id.)
Results ➔

2D-duplicates in PDB-based bioisostere pairs (FC-Bioisostere):

- BioisostereDB = 5139 binding site queries toward the PDB (PDB protein-ligand structures on March 2014 with Xray resolution < 2.5Å ; 350<MW<550 , LigandFrequencyInThePDB < 11)
- 56% of the fragment pairs are related to protein partners sharing the same Uniprot ID
- 87% of the fragment pairs are related to protein partners sharing the same PFAM ID
- Distribution of 2D duplicates of pairs:
  - Example: Carboxylate / Tetrazole in 18 occurrences

- 53.78% 32.20% 7.81% 4.58% 1.65% 32.20%

- 0% 10% 20% 30% 40% 50% 60% 1 2-4 5-9 10-19 20-111

18 pairs in 3D

Best score and worst score
Conclusion

Matching SAR ligands onto 3D PDB via common 2D fragments:
• significant number of 2D (PDB-based) fragment matches to go further
• explore MMP Matched Molecular Pairs protocol for more homogeneous fragments
• consider optimal data architecture to save time (C2P new data model)

FC-Bioisostere:
• a new application by crossmining in 2D/3D PDB and chemical libraries
• complementary source of bioisosteres to SAR with 3D information

Working on the C2P Chemo-Proteomic Platform initiative

• Many roads to explore PDB, SAR, ChemLibs with 3D/2D/1D similarities and logics
• Toward a better understanding of affinity / selectivity mechanisms
• C2P is open to new partners (now MEDIT, Sup biotech and Felix Concordia)

THANKS TO:
Dr Stewart Adcock, Dr Fabrice Moriaud, Dr Jean-Yves Trosset (SupBiotech school)

CONTACT: francois@medit.fr +33 6 7513 0847 located in Paris area