The reuse of structural data for fragment binding site prediction

Richard Hall
Motivation

- many examples of fragments binding in a ‘phenyl shaped pocket’ or a ‘kinase slot’
- good shape complementarity between the ligand and receptor
- can we ascertain the shapes of pockets filled by fragments
  - use the wealth of structural data from the Astex database
- can we generate a descriptor for this kind of pocket
The Phenyl Shaped Pocket
Astex Diverse Set Fragments

1N1M: Dipeptidyl Peptidase IV
1W1P: Chitinase B
1N2V: TGT
1Q41: GSK-3beta
1LRH: Auxin binding protein 1
1SG0: Quinone Reductase 2

www.ccdc.cam.ac.uk/products/life_sciences/gold/validation/astex_diverse
Other Pocket Finding Algorithms Are Available


  • lists 30 algorithms to search for binding pockets
  • we are particularly interested in areas where *fragments* are likely to bind
Ligsite

- Ligsite is a simple algorithm for finding pockets in receptor structures.
- Create an excluded volume grid.
- At each point on the grid:
  - Count the number of vectors along which an atom hits the surface.
- Each point scores between 0 (very exposed) and 7 (very enclosed).
ADS Ligsite
Digsite: Distant Dependant Ligsite

- digsite evolved from the ligsite algorithm
- added to the number of vectors (additional 6), reducing rotational variance
- bin each distance (0-6 Å in 0.5Å bins, plus one ‘mop’ bucket) from ligand atoms to the excluded volume of the receptor
- generate a fingerprint spectrum
- smooth spectrum using a 25:50:25 function
we selected good binding fragments for 17 Astex targets (including 6 kinase, 5 protease)
  • HAC <= 15
  • MW <= 300
  • ClogP <= 3
  • NDon <= 3
  • NAcc <= 3
• ranked by LE
• diverse

• 127 complexes
  • average HAC = 11.7
• calculated the spectrum for each atom
• 1492 spectra
Clustering The Spectra

- spectra were clustered in R (Euclidean distance, Wards method)
- dendrogram was cut to produce 15 clusters
Cluster Pruning

- We averaged the spectra of each cluster
  - Average for each target and again for kinases
- We wanted to find descriptors for tight binding
  - Any average cluster spectrum with >2 distances in the mop bucket was ignored
  - Rank average cluster spectra by a consensus count of distances in the low distance bins
- Top three clusters used as prototype spectra
at each grid point outside the excluded volume:

• generate a digsite spectrum
• calculate the minimum average z score to each of the three prototype spectra
• transform scores to max(2-zscore, 0)
• expand/smooth resulting grids
A Poor Digest Score
A Confession

- when we started this project, the idea was to weight scoring functions so as to bias docking towards fragment binding regions...
- it didn’t improve results

http://en.wikipedia.org/wiki/Hype_cycle
Fragment Binding Site Prediction

- build a digsite or ligsite grid for the entire protein
- look at how many **good scoring points** (digsite > 1.2, ligsite > 6) overlap with **fragment volume(s)** (VdW+1Å)
- do not count any point in the **receptor excluded volume** (<VdW+1Å) or distant (>VdW+5Å) from the receptor
- how well do we perform for pdb and astex structures
the astex diverse set are 85 diverse protein ligand targets that are relevant to drug discovery
- ligands are drug like
- but only 7/85 are fragment like
- there are other pdb structures that have good sequence identity with the astex diverse set
  - CLASP database
    - clusters where sequence similarity > 0.7
    - ligands extracted and stored separately
- can we find examples of fragment ligands bound to these similar structures
  - not rigorously qc'ed as per the original 85
PDB Fragment Selection

- similar fragment binding site
  - within 5Å of the centroid of the original ads ligand
- only organic atoms, >2 carbon atoms
- rule of three compliant
- 6 <= HAC <= 15
- found 114 fragments
- 80 unique smiles
- 27/85 clusters have at least one fragment bound
- choose one pdb code to represent each cluster
  - conserve original ADS fragments
PDB Fragment Analysis

• use a confusion matrix to quantify success for each complex

<table>
<thead>
<tr>
<th></th>
<th>ligsite</th>
<th>digsite</th>
</tr>
</thead>
<tbody>
<tr>
<td>true +ve</td>
<td>115</td>
<td>104</td>
</tr>
<tr>
<td>false +ve</td>
<td>1273</td>
<td>187</td>
</tr>
<tr>
<td>false –ve</td>
<td>1074</td>
<td>1085</td>
</tr>
<tr>
<td>true –ve</td>
<td>388574</td>
<td>389660</td>
</tr>
</tbody>
</table>

• compare precision, \((\text{true +ve})/(\text{true +ve and false +ve})\)
  • above example ligsite 0.08, digsite 0.36 (best case, 3KGP)
  • can be misleading when ligsite true+ve >> digsite true+ve
• compare false +ve ratio for the two methods
Precision Explanation
PDB Fragment Results

- both methods find a ligand binding site in 100% of cases
- digsite is more precise than ligsite
  - average ratio 1.7, median 1.5
- ligsite produces far more false positives than digsite
  - average ratio 7.4, median 6.5
Astex Fragments

- 67 reference structures (23 targets, 724 fragments)
- all fragment ligands aligned to reference structure and used to mark true binding regions
  - HAC <= 15 atoms
- metrics computed for ligsite and digsite

- 91% of digsite grids found true positives, 96% ligsite
- two targets were not amenable to digsite
  - open nature of binding sites
Astex Fragment Results

- digsite is more precise than ligsite
- average ratio 1.7, median 1.2

- ligsite produces far more false positives than digsite
- average ratio 12.1, median 6.5
Conclusions

• we used our database of crystal structures to derive a descriptor that can be used to highlight fragment binding sites

• a retrospective analysis of PDB and Astex data has shown that digsite can be used to find fragment binding sites

• digsite grids are much more focused than ligsite grids
Future Work

• more extensive pdb evaluation
• compare digsite to other pocket finding methods
  • eg ASP grids, pocketome method of An et al
• perform a prospective analysis of in house structures, to ensure we have found all potential binding sites
• look at island counts, rather than precision/false positives
• consider a coloured digsite spectra (metals, donors, acceptors, etc)
Acknowledgements

- Fred Ludlow
- Ilenia Giangreco
- Marcel Verdonk
- Astex CCI
- CCDC
Thank you
www.astex-therapeutics.com