Using Spherical Harmonic Virtual Screening Tools to Compare and Classify HIV Entry Inhibitors for the CXCR4 and CCR5 Co-Receptors

David Ritchie
INRIA, Nancy Grant Est

Violeta Pérez-Nueno
Institut Chimique de Sarre

1. Summary of spherical harmonics plus SH clustering example
2. SH-based retrospective virtual screening of CXCR4 and CCR5 co-receptors
3. Introducing SH “consensus shapes”
4. Analysing CCR5 ligands and binding sub-sites using SH consensus shape clustering

Spherical Harmonic Surfaces

- Use SHs as “building blocks,” i.e. components of shape, etc.
  - Real SHs: \( y_{lm}(\theta, \phi) \)
  - Coefficients: \( a_{lm} \)
  - Encode radial distances from origin as SH series...
  - Solve coefficients by numerical integration...

\[
    r(\theta, \phi) = \sum_{l=0}^{l=15} \sum_{m=-l}^{m=l} a_{lm} y_{lm}(\theta, \phi)
\]

PARA FIT Superposition Searches

- Uses icosahedral tessellation of sphere for Euler rotations
- Samples 22,000 orientations of about 8 degree steps
- Refine with a 16x16x16 grid of 1 degree steps
- Approx 20 pair-wise superpositions/sec on 1.8GHz Xeon PC
- Rotates everything from ParaSurf SDF file –
  - SH coefficients, dipole, quadrupole, moments, etc.,
  - density matrix elements, NAO-PCs, etc.
ParaFit Clustering Example

- Takane et al. collected 47 odour molecules: in 7 classes:
  - bitter, ambergris, jasmine camphor, rose, muguet, musk
- Takane et al. clustered into 10 groups using eigenvector analysis of QM vibrational frequencies...


ParaFit Clustering Example

- Clustering the Odour Dataset
  - Calculate SH shapes using ParaSurf, and cluster with ParaFit:
    - unix% PS_mopac_run
    - unix% PS_Parasurf_run
    - unix% ParaFit –matrix –dif
    - unix% dif2jpg –n10 o.dif _*.psf
    - unix% eog o.jpg

  - Clustering SH shapes gives better clusters than using vibrational frequencies...


HIV and HIV Entry Inhibitors

<table>
<thead>
<tr>
<th>Co-receptor</th>
<th>Targets of Inhibitors and Drugs</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCR5</td>
<td>Anti-viral drugs for Treatment</td>
</tr>
<tr>
<td>CXCR4</td>
<td>Anti-viral drugs for Treatment</td>
</tr>
<tr>
<td>CCR5</td>
<td>Anti-viral drugs for Treatment</td>
</tr>
<tr>
<td>CXCR4</td>
<td>Anti-viral drugs for Treatment</td>
</tr>
</tbody>
</table>

Number of people living with HIV in 2007
- Total: 33,0 million (30–36)
- People newly infected with HIV in 2007: Total: 2,7 million (2,2–3,2)
- AIDS deaths in 2007: Total: 2,0 million (1,8–2,3)

AIDS and HIV Entry Inhibitors

- Acquired Immune Deficiency Syndrome
- Inmunitary system Weakening and/or destruction
- It is not a hereditary disease
- Block infection
- Target Mechanism
  - CD4 (cell) Block CD4 binding by gp120
  - gp120 (virus) Block gp120 conformational changes needed to interact with the chemokine receptor
  - CCR5, CXCR4 (cell) Block chemokine receptor binding by gp120
  - gp41 (virus) Block gp41 structural changes needed for fusion
  - Membrane (cell or virus) Block lipid bi-layer destabilization and mixing


HIV Cell Entry Mechanisms

- The Co-receptor structures were built using Modeller
- But loop E2 was built with CONGEN + disulphide constraints


Targeting the CXCR4 and CCR5 Co-Receptors

- CXCR4 and CCR5 are members of the GPCR family
- We modelled them using bovine rhodopsin as template


Homology Modelling CXCR4/CCR5

- The Co-receptor structures were built using Modeller
- But loop E2 was built with CONGEN + disulfide constraints

CONGEN – open loop E2 (preserves disulfide) MODELLER – loop E2 (blocks pocket)

CONGEN – open loop E2 (broken disulfide bond)
Validating the Receptor Model Structures

- The receptor models were validated by docking selected high-affinity ligands: AMD3100 (CXCR4) and TAK779 (CCR5)

- The binding modes from Autodock were consistent with the available SDM evidence on key ligand-binding residues

Virtual Screening Datasets

- CCR5 Antagonists (424):
  1) SCH-C derivatives
  2) 1,3,5-trisubstituted pentacyclics
  3) Diketopiperazines
  4) 1,3,4-trisubstituted pyrrolidinepiperidines
  5) 5-oxopyrrolidine-3-carboxamides
  6) N,N'-Diphenylureas
  7) 4-aminopiperidine or tropanes
  8) 4-piperidines
  9) TAK derivatives
  10) Guanylhydrazone derivatives
  11) 4-hydroxypiperidine derivatives
  12) Phenylcyclohexilamines
  13) Anilide piperidine N-oxides
  14) 1-phenyl-1,3-propanodiamines
  15) AMD derivatives
  16) Other

- CXC4 antagonists (248):
  1) AMD derivatives
  2) Macrocycles
  3) Tetrahydroquinolinamines
  4) KRH derivatives
  5) Dipicolil amine zinc(II) complexes
  6) Other

PLUS...

4696 inactive compounds from the Maybridge Screening Collection with similar 1D properties to the actives

Receptor-Based VS Enrichment Results

- Each ligand was docked and ranked using: Autodock, GOLD, FRED, Hex

SH Ligand-Based VS Set-Up

- Each database compound was scored against the docked conformation of AMD3100 (CXCR4) and TAK779 (CCR5)

- This example shows the superpositions of (top) AMD3167 (blue), and (bottom) SCH417690) with the given queries

- NB. The database conformations were calculated by MOE FlexAlign... ROCS used Omega for 10 further conf.s

SH Ligand-Based VS Enrichment Results

- Query = AMD3100 for CXCR4; TAK779 for CCR5

- Docking enrichments are better for CXCR4 than CCR5

- But shape-based scoring gives better overall enrichments
Calculating Consensus Shapes

1. Do all vs all SH comparison
2. Find best pair-wise match
3. Calculate SH average of pair
4. Treat average as new seed
5. Superpose all onto seed
6. Rotate all onto new seed
7. Iterate until convergence...
8. Result = SH pseudo-molecule

Consensus Shape-Based VS

Overall Results – CCR5

Best scorers:
- ParaFit 3-Consensus
- FRD2 Consensus
- ParaFit 5-Consensus

Overall Results – CXCR4

Best scorers:
- ParaFit 3-Consensus
- ParaFit Testuto
- Fred Consensus
- ROCS Combo

Experimental Evidence for Multiple CCR5 Binding Sites

There is strong evidence that there are multiple sub-sites within the CCR5 extracellular pocket:

- It is very difficult to superpose all the different families of CCR5 active compounds.

- VS enrichment results are strongly dependent on the conformation of the query molecules.

- Site directed mutagenesis evidence suggests a large pocket (the SDM residues are spatially well distributed around the pocket).

- Not all SDM locations affect the binding of all ligands.

Consensus Shapes of the Three Most Active Inhibitors

SH Consensus Shapes of the
CXCR4
CCR5

Fred Consensus
ParaFit Tanimoto

ParaFit 3-Consensus
Best scorers:
ParaFit 5-Consensus

Overall Results – CXCR4

Overall Results – CCR5

Fred Consensus
ParaFit Tanimoto

Best scorers:
ParaFit 3-Consensus
ParaFit 5-Consensus

Experimental Evidence for Multiple CCR5 Binding Sites

There is strong evidence that there are multiple sub-sites within the CCR5 extracellular pocket:

- It is very difficult to superpose all the different families of CCR5 active compounds.

- VS enrichment results are strongly dependent on the conformation of the query molecules.

- Site directed mutagenesis evidence suggests a large pocket (the SDM residues are spatially well distributed around the pocket).

- Not all SDM locations affect the binding of all ligands.

Consensus Shapes of the Three Most Active Inhibitors

SH Consensus Shapes of the
CXCR4
CCR5

Fred Consensus
ParaFit Tanimoto

Best scorers:
ParaFit 3-Consensus
ParaFit 5-Consensus

Overall Results – CXCR4

Overall Results – CCR5

Fred Consensus
ParaFit Tanimoto

Best scorers:
ParaFit 3-Consensus
ParaFit 5-Consensus

Experimental Evidence for Multiple CCR5 Binding Sites

There is strong evidence that there are multiple sub-sites within the CCR5 extracellular pocket:

- It is very difficult to superpose all the different families of CCR5 active compounds.

- VS enrichment results are strongly dependent on the conformation of the query molecules.

- Site directed mutagenesis evidence suggests a large pocket (the SDM residues are spatially well distributed around the pocket).

- Not all SDM locations affect the binding of all ligands.

Consensus Shapes of the Three Most Active Inhibitors

SH Consensus Shapes of the
CXCR4
CCR5

Fred Consensus
ParaFit Tanimoto

Best scorers:
ParaFit 3-Consensus
ParaFit 5-Consensus

Overall Results – CXCR4

Overall Results – CCR5

Fred Consensus
ParaFit Tanimoto

Best scorers:
ParaFit 3-Consensus
ParaFit 5-Consensus

Experimental Evidence for Multiple CCR5 Binding Sites

There is strong evidence that there are multiple sub-sites within the CCR5 extracellular pocket:

- It is very difficult to superpose all the different families of CCR5 active compounds.

- VS enrichment results are strongly dependent on the conformation of the query molecules.

- Site directed mutagenesis evidence suggests a large pocket (the SDM residues are spatially well distributed around the pocket).

- Not all SDM locations affect the binding of all ligands.

Consensus Shapes of the Three Most Active Inhibitors

SH Consensus Shapes of the
CXCR4
CCR5

Fred Consensus
ParaFit Tanimoto

Best scorers:
ParaFit 3-Consensus
ParaFit 5-Consensus

Overall Results – CXCR4

Overall Results – CCR5

Fred Consensus
ParaFit Tanimoto

Best scorers:
ParaFit 3-Consensus
ParaFit 5-Consensus

Experimental Evidence for Multiple CCR5 Binding Sites

There is strong evidence that there are multiple sub-sites within the CCR5 extracellular pocket:

- It is very difficult to superpose all the different families of CCR5 active compounds.

- VS enrichment results are strongly dependent on the conformation of the query molecules.

- Site directed mutagenesis evidence suggests a large pocket (the SDM residues are spatially well distributed around the pocket).

- Not all SDM locations affect the binding of all ligands.
Exploring the CCR5 Multiple Binding Site Hypothesis

- There is a hypothesis that the CCR5 ligands form two or more groups, i.e., they have two or more binding modes.

Clustering the 424 CCR5 Ligands

- Because it is not clear a priori which ligands might belong to which group, we first performed Ward's hierarchical clustering of chemical fingerprints.
- We then used Chess's method to find the optimal number of clusters (16).
- These were manually merged to 10 groups based on known CCR5 families.
- SH consensus shapes were calculated for the 10 groups.
- These were then compared in ParaFit (all-vs-all).
- Another round of Ward's clustering proposed four super-consensus clusters.

From Consensus Shapes to Super-Consensus Clusters

- SC queries => CCR5 ligands form no less than FOUR groups

Using Super-Consensus Shapes as VS Queries

- Each SC pseudo-molecule was used as a VS query.
- NB, merging SC shapes significantly worsens the AUCs...
- SC queries => CCR5 ligands form no less than FOUR groups

Hex Blind Docking of SC Pseudo-Molecules to CCR5

- 3D pseudo-molecules were created as the union of all superposed ligands in each SC family for docking in Hex.
- SC-A docks to Site-1 (TM 1, 2, 3, 7)
- SC-C docks to Site-2 (TM 3, 5, 6)
- B and D dock to Site-3 (TM 3, 6, 7)

Autodock Docking VS w.r.t. Three CCR5 Sub-Sites

- To confirm the SC shapes were matched to their predicted target sites, docking based VS was repeated for each ligand using:
  - SC-A treated as active for Site 1 (SCs B, C, D treated as inactive)
  - SC-Cs treated as active for Site 2 (SCs A, B, D treated as inactive)
  - SC-B/Ds assumed active for Site 3 (SCs A and C treated as inactive)

- As before, merging SCs worsens the AUCs...
- SC docking => no less than THREE CCR5 pocket sub-sites
Conclusions

- SH surfaces allow fast comparison and clustering
  - SH-based clustering of Odour dataset superior to EVA clustering
- Our models of CXCR4 and CCR5 are consistent with SDM
- We built a VS library of 248 CXCR4 and 424 CCR5 inhibitors
- Ligand-based VS gives better enrichments than docking
- ParaFit and ROCS give the best overall VS enrichments
- Docking & SH-based VS results for CXCR4 better than CCR5
  - CXCR4 has smaller pocket and fewer ligands than CCR5
- Consensus clustering of CCR5 ligands -> FOUR super-families
- Docking CCR5 SC pseudo-molecules -> THREE sub-sites

Acknowledgments

- Violeta Pérez-Nueno
- Lazaros Mavridis
- Brian Hudson
- Vishwesh Venkatraman
- EPSRC
- University of Aberdeen
- I QS, Universitat Ramon-Llull

Papers: http://www.loria.fr/~dritchie/  
ParaSurf + ParaFit: http://www.ceposinsilico.de/