Using “Consensus shape-based clustering” to help understand how multiple ligands might bind to multiple target sites

Violeta Pérez-Nueno, David Ritchie
Orpailleur Team, INRIA Nancy - Grand Est

LORIA (Laboratoire Lorrain de Recherche en Informatique et ses Applications),
INRIA Nancy – Grand Est, 615 rue du Jardin Botanique, 54506 Vandoeuvre-lès-Nancy, France
Presentation Overview

1. Introducing VS (Shape Matching)
2. Summary of spherical harmonics
3. Introducing SH “consensus shapes”
4. SH-based retrospective virtual screening of CXCR4 and CCR5 co-receptors
5. Analysing CCR5 ligands and binding sub-sites using SH consensus shape clustering
6. Examples of how Consensus Clustering can identify Multi-Site Pockets
1. Introducing VS (Shape Matching)
Virtual screening

• One of the main activities in the drug discovery process is to identify or design synthetic ligands which will bind to a given protein target and therefore serve as leads for a drug development.

• Three-dimensional virtual screening (3D VS) techniques are becoming an increasingly important approach for identifying and selecting biologically active molecules against specific targets from large chemical databases.

• In 3D VS, compounds are usually selected for testing from a database of small-molecules by applying different software screening tools and techniques. The approaches used can often be classified as either “structure-based” or “ligand-based.”
**VIRTUAL SCREENING**

**Ligand-based**
- Active Ligands
  - Conformational analysis
    - Similarity search. ADME models
    - Pharmacophore Shape matching
    - QSAR

**Structure-based**
- Target protein
  - Conformational analysis
    - Drug likeness filters and ADME properties
    - Receptor modelling
  - Compound library generation
  - Flexible docking and interaction fingerprints post-processing
  - Ligand Scoring
  - Post-processing hit ranking lists, synthesis and testing
  - Refinement
Virtual screening

VS techniques are usually used in two main ways. Often, known active compounds are first used in a retrospective VS study to determine the best docking or database search parameters and to validate the initial ligand-receptor model. Then, in prospective VS, the same parameters are used predictively to propose a small number of new candidate compounds to be synthesised and tested.

RETROSPECTIVE VIRTUAL SCREENING
- Using known active compounds.
- Validation of the virtual screening approaches. Selection of the best parameters.

PROSPECTIVE VIRTUAL SCREENING
- Application of the screening approaches to a real case.
- Identification of new compounds.
A VS study often involves applying a sequence of filters of increasing selectivity to reduce a large virtual library of known or potential molecular structures to a small collection of around 100 candidate molecules. This process may be considered in terms of a “filtering funnel” in which fast but crude filters are applied first, and then more sensitive but more computationally expensive filters are applied to distinguish the remaining candidates.

ADME: Absorption, Distribution, Metabolism and Excretion
Virtual screening procedures

- Compiling a database:
  Active compounds from the literature.
  Drug-like compounds to screen. Drug likeness filters.
  Presumed inactive compounds from Maybridge Screening Collection with 1D properties similar to those of the actives in order to avoid potential bias of the VS results. Drug likeness filters.
- Chemical similarity search and ADME properties using Qikprop/Qiksim.
- Pharmacophore modelling using MOE and Discovery Studio.
- Docking-based virtual screening using Autodock, Gold (with constraints for binding site residues), Fred, and Hex.
- Hit-ranking-lists post-processing according to the scores retrieved.
- Consensus “Rank-by-Vote” of all the first hit ranking lists compounds and compound selection.
Virtual screening evaluation

- **Enrichment plot:** ranking of compounds according to the scoring function. Representation of % Actives vs % Database screened.

- **ROC (Receiving Operating Characteristic) plot:** ranking of compounds according to the scoring function. Representation of Selectivity (%e) vs 1-Specificity (% Inactives chosen).
Ligand-Based Virtual screening

• Ligand-based VS approaches are based on the assumption that molecules which are structurally related should have similar biological activities.

• However, it is worth noting that sometimes small structural changes lead to large changes in activity, and that similar molecules sometimes show different binding modes. In spite of the above difficulties, fast and efficient ligand-based techniques have been very useful, especially when the structure of the receptor is not available.
Ligand-Based Virtual screening

- **Similarity-Based Approaches:** Similarity searches start from one or several target structures and their description by one or more structural descriptors, along with the description of the candidate compounds contained in the virtual library. The descriptors used to characterize these virtual libraries are classified as “1D” (which specify the atomic type solely), “2D” (which include topological connectivity of the molecule), and “3D” (which consider the 3D structure of molecule).

- **Pharmacophore Models:** A pharmacophore may be defined as a set of features common to a series of active molecules such as hydrogen-bond donors/acceptors, aromatic ring centroids, charged groups, and hydrophobic regions. The active compounds are superimposed to determine their common features and hence to provide a pharmacophore model that explains ligand-receptor binding.
Ligand-Based Virtual screening

- **Molecular Fields**: This method uses the molecular electrostatic potential (MEP) for similarity searches. In order for these methods to measure the similarity of a molecular pair, the molecules must first be superposed. Then the similarity is calculated by applying a similarity coefficient using the values of the fields within a 3D grid.

- **3D-QSAR**: Aims to correlate chemical structures with biological activities of a series of related compounds. They use a number of location-dependent measures that describe molecular properties. Generally, 3D-QSAR models are derived from the calculation of vectors of molecular electrostatic potentials (MEPs) mapped over the molecular surfaces of the ligands, as well as molecular field descriptors based on the description of receptor-ligand interactions by means of molecular interaction potentials (MIPs).

**Electrostatic Field**
- BLUE – positive charges
- RED – negative charges

**Steric Field**
- GREEN – area to be filled
- YELLOW – steric clash
Ligand-Based Virtual screening

- **Shape Matching:** Shape matching approaches are based on the comparison/superposition of the 3D shapes of a set of molecules against a well-known active molecule. The three-dimensional shape of an active molecule against a specific target is the complementary adapted one for the interaction with the active site of this target, therefore the compounds that have a three-dimensional shape similar to the one of the well-known active (shape-matching query) will have major probability of fitting in the biological receptor and consequently greater activity. The main problem with these techniques is the selection of the conformation of the query because the conformations and overall 3D shapes of two active ligands can sometimes be quite different.

2 main approaches to describe molecular surfaces:

- Gaussian Description of Molecular Shape
- Spherical Harmonic Molecular Surfaces
Ligand-Based Virtual screening

**Gaussian Description of Molecular Shape:** Grant *et al.* describe an algorithm that uses molecular volumes to represent and compare molecules. The ROCS program (Rapid Overlay of Chemical Structures) is the commercial implementation of this algorithm. ROCS is a fast shape comparison application, based on the idea that molecules have similar shape if their volumes overlay well and any volume mismatch is a measure of dissimilarity. It uses a smooth Gaussian function to represent the molecular volume, so it is possible to routinely minimize to the best global match.

- **Spherical Harmonic Molecular Surfaces:** SH-based shape matching is at the state of the art but it is still relatively new, and so far only our group was the first using it for VS. Bernard Maigret developed MSSH and SHEF to provide fast shape-based filters for protein-ligand docking, while Dave Ritchie developed ParaFit, to perform ligand-based screening using key quantum mechanical molecular surfaces properties calculated by Tim Clark's ParaSurf program.
2. Summary of spherical harmonics plus SH clustering example
Spherical Harmonic Surfaces

Surface shapes are represented as radial distance expansions of the molecular surface with respect to the center of the molecule.

- Real SHs: $y_{lm}(\theta, \phi)$
- Coefficients: $a_{lm}$
- Encode radial distances from origin as SH series…
- Solve coefficients by numerical integration…

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

ParaSurf calculates molecular shape and electronic properties from semi-empirical quantum mechanics theory

- From MOPAC or VAMP, calculate:
  - Density contours of $2 \times 10^{-4} \text{e/A}^3$ (i.e. approx = SAS)
  - MEP – electrostatic potential
  - $\text{IE}_L$ – ionization energy
  - $\text{EA}_L$ – electron affinity
  - $\alpha_L$ – polarizability

And encodes these properties as SH expansions to order $L=15$...

ParaFit calculates superpositions between pairs of molecules by exploiting the special rotational properties of the SH functions

- **Distance:** $D = \int (r_A(\theta, \phi) - r_B(\theta, \phi'))^2 d\Omega$
- **Orthogonality:** $D = |a|^2 + |b|^2 - 2 a.b'$
- **Rotation:** $b'_{lm} = \sum_{m'} R^{(l)}_{mm'}(\alpha, \beta, \gamma)b_{lm'}$
- **Carbo:** $S = a.b' / (|a|.|b|)$
- **Hodgkin:** $S = 2a.b' / (|a|^2 + |b|^2)$
- **Tanimoto:** $S = a.b' / (|a|^2 + |b|^2 - a.b')$
- **Multi-property:** $Q = pS + qS^{MEP} + rS^{IEl} + ...$

3. Introducing SH “consensus shapes”
Calculating Consensus Shapes

1. Do all-v-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. Compute new average seed
7. Rotate all onto new seed
8. Iterate until convergence...
9. Result = SH pseudo-molecule

4. SH-based retrospective virtual screening of CXCR4 and CCR5 co-receptors
HIV and HIV Entry Inhibitors

Acquired Immunodeficiency Syndrome (AIDS) is a group of symptoms and signs. It is not a hereditary disease.

- **Acquired Immune Deficiency**
  - Weakening and/or destruction

**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)

Adult prevalence (%)
- 15.0% – 28.0%
- 5.0% – 15.0%
- 1.0% – 5.0%
- 0.5% – 1.0%
- 0.1% – 0.5%
- < 0.1%
- No data available
HIV life cycle
HIV Cell Entry Mechanisms
HIV Cell Entry Mechanisms

<table>
<thead>
<tr>
<th>Target</th>
<th>Mechanism</th>
</tr>
</thead>
<tbody>
<tr>
<td>CD4 (cell)</td>
<td>Block CD4 binding by gp120 (A)</td>
</tr>
<tr>
<td>gp120 (virus)</td>
<td>Block gp120 conformational changes needed to interact with the chemokine receptor (B)</td>
</tr>
<tr>
<td>CCR5, CXCR4 (cell)</td>
<td>Block chemokine receptor binding by gp120 (C)</td>
</tr>
<tr>
<td>gp41 (virus)</td>
<td>Block gp41 structural changes needed for fusion (D, E)</td>
</tr>
<tr>
<td>Membrane (cell or virus)</td>
<td>Block lipid bilayer destabilization and mixing (F, G)</td>
</tr>
</tbody>
</table>

## HIV Cell Entry Mechanisms

The following table outlines the targets and mechanisms involved in HIV cell entry and inhibition:

<table>
<thead>
<tr>
<th>Target</th>
<th>Mechanism</th>
</tr>
</thead>
<tbody>
<tr>
<td>CD4 (cell)</td>
<td>Block CD4 binding by gp120</td>
</tr>
<tr>
<td>gp120 (virus)</td>
<td>Block gp120 conformational changes needed to interact with the chemokine receptor</td>
</tr>
<tr>
<td>CCR5, CXCR4 (cell)</td>
<td>Block chemokine receptor binding by gp120</td>
</tr>
<tr>
<td>gp41 (virus)</td>
<td>Block gp41 structural changes needed for fusion</td>
</tr>
<tr>
<td>Membrane (cell or virus)</td>
<td>Block lipid bi-layer destabilization and mixing</td>
</tr>
</tbody>
</table>

CXCR4 and CCR5 co-receptors
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 drivatives
11) 4-hydroxypiperidine derivatives
12) Phenylcyclohexitilamines
13) Anilide piperidine N-oxides
14) 1-phenyl-1,3-propanodiamines
15) AMD derivatives
16) Other

CXCR4 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
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 + disulphide constraints

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

CONGEN – open loop E2 (broken disulfide bond)
CXCR4/CCR5 modelling

CXCR4/CXCR5 modelling

CXCR4 co-receptor

CXCR4/CCL5 modelling

CXCR4-MODELLER co-receptor

BLIND DOCKING (Autodock)

Loop E2 closed conformation

Blocked binding pocket due to the closed loop E2 conformation

Starting conformation (black)

C-terminus

N-terminus

Asp₁₇₁

Asp₂₆₂

Glu₂₈₈

Loop E1

Loop E3
**CXCR4/CCR5 modelling**

CCR5-MODELLER co-receptor

*BLIND DOCKING (Autodock)*

Lowest docked energy conformation

Starting conformation
Starting conformation

Steric clash between ligand conformations and loop E2

CXCR4/CCR5 modelling

CCR5-MODELLER co-receptor

Binding mode DOCKING (Autodock)
CXCR4/CCR5 modelling

Loop E2 CXCR4

MODELLER - Loop E2

CONGEN - Loop E2 (preserved disulfide bond)

CONGEN - 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

Receptor-Based VS Enrichment Results

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

CXCR4 inhibitors

CCR5 inhibitors

SH Ligand-Based VS Set-Up

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

(a) (b) (c)

(d) (e) (f)

ParaFit ROCS Hex

• 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
SH Consensus Shapes of the Three Most Active Inhibitors

**CXCR4**

Consensus shape

KRH derivate superposition

Macrocyle derivate superposition

AMD derivate superposition

**CCR5**

Consensus shape

1,3,4-trisubstituted pyridine piperidine derivate superposition

SCH derivate superposition

Piperidine derivate superposition
Consensus Shape-Based VS

CXCR4

CCR5

Overall Results – CXCR4

Best scorers:

- ParaFit 3-Consensus
- ParaFit Tanimoto
- Fred Consensus
- ROCS Combo
Overall Results – CCR5

Best scorers:
- ParaFit 3-Consensus
- FRED Consensus
- ParaFit S-Consensus

5. Consensus Clustering: Exploring CCR5 Multiple Binding Sites
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 molecule.

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

Figure 2. CCR5 TM binding cavity. A. Amino acid sequence. Residues with side chain pointing toward the cavity are written in red and pointed using Ballesteros numbering.\textsuperscript{61} except in the extracellular domain 1 (EL1). Numbers indicated underneath the sequence summarize experimental mapping of receptor interaction site for nonpeptide antagonists 1–5; \textsuperscript{1} residues important for the efficiency of 1,\textsuperscript{22,23,24} \textsuperscript{2} residues important for the efficiency of 2 and Schering-Plough compound AD101,\textsuperscript{23–25} and \textsuperscript{3} residues important for the binding 4 and 5.\textsuperscript{25} B. The Connolly surface of the CCR5 receptor cavity (colored according to the lipophilic potential) is displayed together with the ribbon diagram of the seven TM helices. Side chains of key residues highlighted in the sequence are depicted using line representation. The bottom view is rotated about a vertical axis by 180° relative to top view.

Clustering the 424 CCR5 Ligands

- Because it is not clear *a priori* which ligands might belong to which group, we first performed Wards hierarchical clustering of chemical fingerprints...
- We then used Kelley’s method to find the optimal number of clusters (16)
- These were manually merged to 10 groups based on known CCR5 families

<table>
<thead>
<tr>
<th>CLUSTER</th>
<th>Compounds Found</th>
<th>Number of compounds</th>
<th>Consensus Shape</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(8) 1,3,4-trisubstituted pyrrolidinopiperidines</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1) 1,3,5-trisubstituted piperazines</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(5) N,N'-disubstituted carbamides</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2) TAK derivatives</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1) 4-piperidines</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1) others (MRK-1-CMPO 167)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>(1) 1,3,4-trisubstituted pyrrolidinopiperidines</td>
<td>41</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(6) 1,3,5-trisubstituted piperazines</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(13) 1-(phenyl)-1,3-propanoamines</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3) 4-piperidines</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1) AMD derivatives</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(9) Diketopiperazines</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1) SCH derivatives</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2) Phenylcyclohexylamines</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3) others (GSK, Merck1, Merck3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>(22) Amido piperidine N-oxides</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1) TAK derivatives</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1) others (1-benzazepine)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>(2) 1-phenyl-1,3-propanoamines</td>
<td>26</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(5) Phenylcyclohexylamines</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>(11) 1-phenyl-1,3-propanoamines</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>(12) 1-phenyl-1,3-propanoamines</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>(26) 4-amino piperidine or tropans</td>
<td>35</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(6) 4-piperidines</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2) Phenylcyclohexylamines</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1) others (Merck1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>(23) SCH derivatives</td>
<td>23</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>(20) SCH derivatives</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>(37) SCH derivatives</td>
<td>37</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>(22) SCH derivatives</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>(17) SCH derivatives</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>(19) TAK derivatives</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>(44) TAK derivatives</td>
<td>44</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>(33) Guanidine derivatives</td>
<td>33</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>(36) 4-hydroxy piperidine derivatives</td>
<td>36</td>
<td></td>
</tr>
</tbody>
</table>

- 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

- **Cluster 1**
- **Cluster 2**
- **Cluster 3**
- **Cluster 4**
- **Cluster 5,6**
- **Cluster 7**
- **Cluster 8,9,10,11,12**
- **Cluster 13,14**
- **Cluster 15**
- **Cluster 16**

**Super-consensus C**

**Super-consensus D**

**Super-consensus A**

**Super-consensus B**

**CCR5 big binding pocket**

**SC_A (87 compounds):** TAK derivatives
- Anilide piperidine N-oxides

**SC_B (69 compounds):** Guanylylhydrazone derivatives
- 4-hydroxypiperidine derivatives

**SC_C (184 compounds):** SCH derivatives
- 1,3,4-trisubstituted pyrrolidinepiperidines
- 1,3,5-trisubstituted pentacycles
- 5-oxopyrrolidine-3-carboxamides
- N,N’-diphenylureas
- Diketopiperazines
- AMD derivatives
- 1-phenyl-1,3-propanodiamines
- 4-piperidines

**SC_D (84 compounds):**
- 1-phenyl-1,3-propanodiamines
- Phenylcyclohexamides
- 4-aminopiperidine or tropanes
- 4-piperidines
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-As treated as actives for Site 1 (SCs B, C, D treated as inactives)
  - SC-Cs treated as actives for Site 2 (SCs A, B, D treated as inactives)
  - SC-B/Ds assumed active for Site 3 (SCs A and C treated as inactives)

A -> Site-1

B, D -> Site-3

C -> Site-2

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

- Berlex Science recently synthesised 69 guanyl-hydrozone and 4-piperidine-hydrazine derivatives which showed activity as CCR5 antagonists.

- We performed retrospective VS against 3388 decoys from Maybridge Screening Collection, with similar 1D properties to the actives using:
  - One high affinity query
  - Consensus of the 3 most active
  - Consensus of all actives...

CCR5 VS with Berlex Dataset

- Using Berlex actives as queries to previous 424/4696 dataset:

  Consensus of top 3 Berlex actives

Maraviroc = 1-phenyl-1,3-propanodiamine

Vicriviroc = SCH417690 (Schering Plough)

Aplaviroc = diketopiperazine
6. Examples of how Consensus Clustering can identify Multi-Site Pockets

A total of 2,950 active compounds against a total of 40 targets
Multi-Site Pockets: p38

- 353 p38 DUD ligands clustered using Wards hierarchical clustering of chemical fingerprints
- Kelley’s method to find the optimal number of clusters (15)
- SH consensus shapes were calculated for the 15 groups
- These were then compared in ParaFit (all-vs-all)
- Another round of Ward’s clustering proposed three super-consensus clusters
Multi-Site Pockets: p38

SC pseudomolecules were blind docked onto the p38 pocket using Hex software. Here are shown the blind docking results. The SC A pseudomolecule is docked onto one side of the pocket, SC B pseudomolecule is docked onto the opposite side, and SC C is docked in the same way as SCA.
Searching for literature reporting ligands binding to p38 pocket we found some examples of molecules binding to an allosteric sub-site on the right side of the pocket, making interactions with Glu71 and Asp168, or binding to the ATP binding sub-site, making mainly interactions with Met109. Compound 1 and BIRB bind to the allosteric sub-site, while compound 2 binds to the ATP sub-site.
Multi-Site Pockets: p38

- **SITE 1**: (ATP site)
  - Glu 71
  - Met 109
  - Asp 168

- **SITE 2**: (Allosteric site)
  - Glu 71
  - Met 109
  - Asp 168

- **SITE 1**: (ATP site)
  - ATP

- **SITE 2**: (Allosteric site)
  - ATP

- **Site 1** and **Site 2** are labeled appropriately.

Diaryl urea inhibitors and pyridinyl-imidazole inhibitors are discussed in the context of the pockets.

SC_A, SC_B, and SC_C denote different structural complexes or configurations.
Multi-Site Pockets: Alr2

• 26 alr2 DUD ligands clustered using Wards hierarchical clustering of chemical fingerprints

• Kelley’s method to find the optimal number of clusters (5)

• SH consensus shapes were calculated for the 5 groups

• These were then compared in ParaFit (all-vs-all)

• Another round of Ward’s clustering proposed three super-consensus clusters
Again, SC pseudomolecules were blind docked onto the alr2 pocket using Hex software. The SC A pseudomolecule was docked onto one side of the pocket, binding to residues of the catalytic cleft (Tyr48, His110, Trp111), whereas SC B and C pseudomolecules, which have larger shapes, occupy the whole pocket, binding to the catalytic cleft but also to residues on the other side of the pocket (Thr113, Leu300, Cys298).
There have been characterized mainly 5 binding pocket conformations. 1, 2, and 4 because of their small size and compact scaffold, bind to a side of the pocket, mainly the residues of the catalytic cleft (Tyr48, His110, Trp111), whereas, 3 and 5, which are larger, they occupy the whole pocket, also binding to residues in the opposite side of the catalytic pocket (Leu 300, Thr113, Cys298, Ala 299, Ser302).
Multi-Site Pockets: Alr2

Small ligand Sub-site 1 (catalytic cleft)

Super ligand Sub-site 2 (whole pocket)
Conclusions

- SH surfaces allow fast comparison and 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
- Good retrospective VS results on the Berlex actives
- Consensus clustering allows to detect multi-site targets (p38, alr2 examples).
Acknowledgements

- Dave Ritchie
- Vishwesh Venkatraman
- Lazaros Mavridis

- INRIA Nancy - Grand Est
- IQS, Universitat Ramon-Llull

- Agence Nationale de la Recherche (ANR-08-CEXC-017-01)
- Generalitat de Catalunya-DURSI (FI2008 and BE-DGR2009)
- FP7 Marie Curie IEF Fellowship (DOVSA 254128)

Papers: http://www.loria.fr/~pereznue/
         http://www.loria.fr/~ritched/
ParaSurf + ParaFit: http://www.ceposinsilico.de/
Thank you!