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

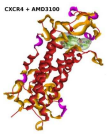
David Ritchie
INRIA, Nancy Grant Est

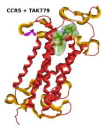
Violeta Pérez-Nueno
Institut Chimíque de Sarià

PERSONA CIÈNCIA EMPRESA
Universitat Ramon Llull

ESCUELA TÉCNICA SUPERIOR

Spherical Harmonic Virtual Screening – Talk Overview

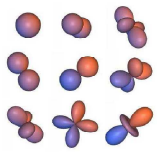




1. Summary of spherical harmonics
2. SH-based retrospective virtual screening of CXCR4 and CCR5 co-receptors
3. Introducing SH "consensus shapes"
4. Analysing CCR5 ligands and binding sub-sites using SH consensus shape clustering

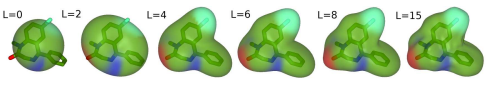
Spherical Harmonic Surfaces

- 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…

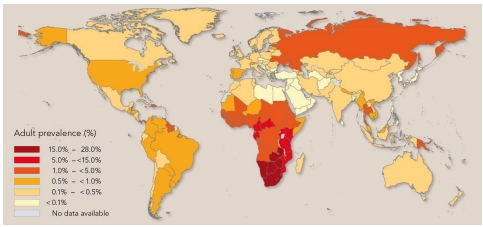
$$r(\theta,\phi) = \sum_{l=0}^{15} \sum_{m=-l}^{l} a_{lm} y_{lm}(\theta,\phi)$$



Ritchie, D.W. and Kemp, G.J.L. *J. Comp. Chem.* **1999**, *20*, 383–395.

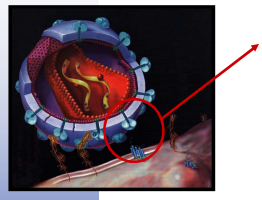
HIV and HIV Entry Inhibitors

A	Acquired	Group of symptoms and signs
I	Immune	Immunitary system
D	Deficiency	Weakening and/or destruction
S	Syndrome	It is not a hereditary disease

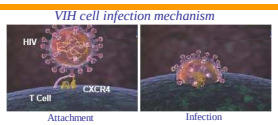


Number of people living with HIV in 2007	Total: 33,0 million (30–36)
People newly infected with HIV in 2007	Total: 2,7 million (2,2–3,2)
AIDS deaths in 2007	Total: 2,0 million (1,8–2,3)

HIV Cell Entry Mechanisms



VIH cell infection mechanism



VIH entry inhibition mechanism

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

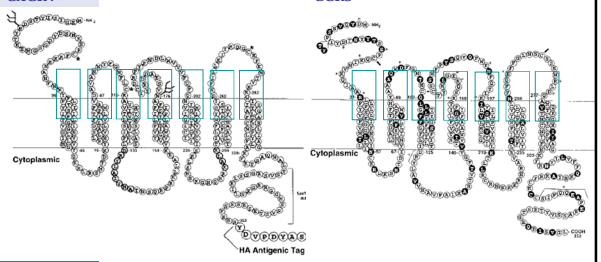
Shaheen, F.; Collman, R.G. *Curr. Opin. Infect. Dis.* **2004**, *17*, 7–16.

5/31

Targeting the CXCR4 and CCR5 Co-Receptors

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

CXCR4 *CCR5*



Cabrera, C. *et al. AIDS Res. Hum. Retrovir.* **1999**, *15*, 1535–1543.
Berson, J.F. *et al. J. Virol.* **2000**, *10*, 255–277.

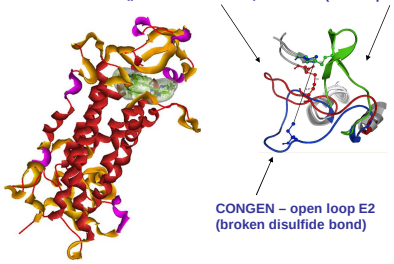
6/31

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)

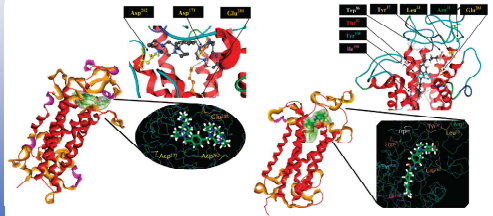


CONGEN – open loop E2 (broken disulfide bond)

7/31

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

Pérez-Nueno *et al. J. Chem. Inf. Model.* **2008**, *48*, 2146–2165.

8/31

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) Phenylcyclohexilamines
13) Anilide piperidine N-oxides
14) 1-phenyl-1,3-propanodiamines
15) AMD derivatives
16) Other

CXCR4 antagonists (248):

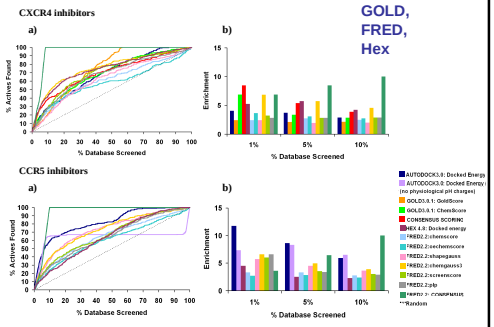
1) AMD derivatives
2) Macrocyles
3) Tetrahydroquinolinamines
4) KRH derivatives
5) Dipicolil amine zinc(II) complexes
6) Other

PLUS…

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

Receptor-Based VS Enrichment Results

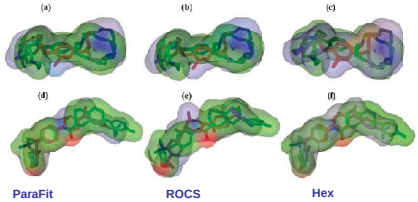
- Each ligand was docked and ranked using: Autodock, GOLD, FRED, Hex



Pérez-Nueno *et al. J. Chem. Inf. Model.* **2008**, *48*, 2146–2165.

SH Ligand-Based VS Set-Up

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

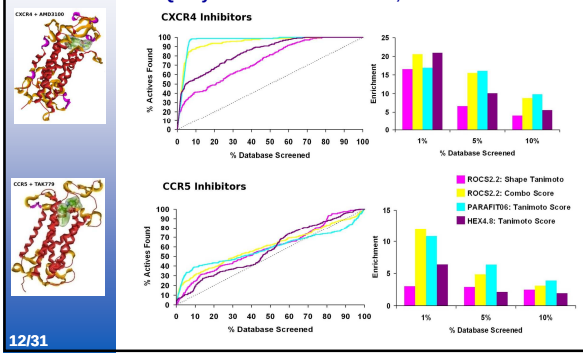


- This example shows the superpositions of (top) AMD3167 (blue), and (bottom) SCH417690) with the given queries
- NB. The database conformations were calculated by MOE FlexAlign… ROCS used Omega for 10 further conf.s

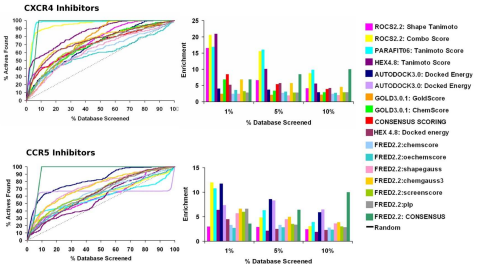
Pérez-Nueno *et al. J. Chem. Inf. Model.* **2008**, *48*, 2146–2165.

SH Ligand-Based VS Enrichment Results

- Query = AMD3100 for CXCR4; TAK779 for CCR5

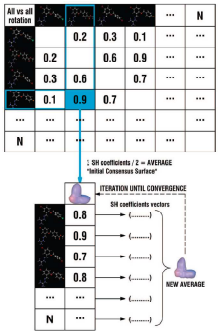


Comparing Ligand-Based and Receptor-Based VS



- Docking enrichments are better for CXCR4 than CCR5
- But shape-based scoring gives better overall enrichments

Calculating Consensus Shapes

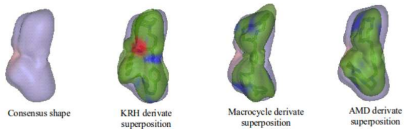


1. Do all-v-all SH comparison
2. Find best pair-wise match
3. Calculate SH average of pair
4. Treat average as new seed
5. Superpose all onto seed
6. Compute new average seed
7. Rotate all onto new seed
8. Iterate until convergence...
9. Result = SH pseudo-molecule

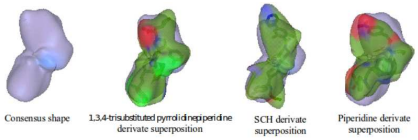
Pérez-Nueno *et al. J. Chem. Inf. Model.* **2008**, *48*, 2146–2165.

SH Consensus Shapes of the Three Most Active Inhibitors

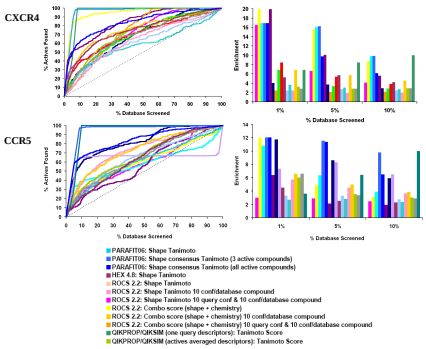
CXCR4



CCR5

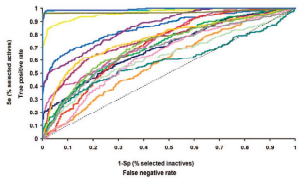


Consensus Shape-Based VS



Pérez-Nueno *et al. J. Chem. Inf. Model.* **2008**, *48*, 2146–2165.

Overall Results – CXCR4



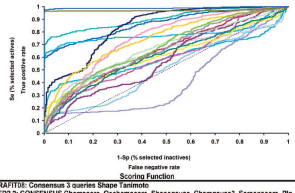
Best scorers:

- ParaFit 3-Consensus
- ParaFit Tanimoto
- Fred Consensus
- ROCS Combo

Scoring Function	AUC
PARAFIT08: Consensus 3 queries Shape Tanimoto	0.994
PARAFIT08: Shape Tanimoto	0.992
FRED2.2: CONSENSUS Chemscore, Oechemscore, Shapegauss, Chemgauss3, Screenscore, Plp	0.974
ROCS2.2: Combo Score (Shape Tanimoto + Scaled Color Score)	0.966
HEX4.8: Shape Tanimoto	0.962
PARAFIT08: Consensus all compounds queries Shape Tanimoto	0.937
FRED2.2: Chemgauss3	0.768
ROCS2.2: Shape Tanimoto	0.761
HEX4.8: Docked Energy	0.756
FRED2.2: Screenscore	0.717
CONSENSUS SCORING: Autodock Docked Energy, Gold GoldScore, Gold ChemScore	0.717
GOLD3.0.1: ChemScore	0.712
AUTODOCK3.0: Docked Energy	0.699
FRED2.2: Plp	0.687
FRED2.2: Shapegauss	0.637
FRED2.2: Chemscore	0.623
GOLD3.0.1: GoldScore	0.595
FRED2.2: Oechemscore	0.398

Pérez-Nueno *et al. J. Chem. Inf. Model.* **2008**, *48*, 2146–2165.

Overall Results – CCR5



Best scorers:

- ParaFit 3-Consensus
- FRED Consensus
- ParaFit S-Consensus

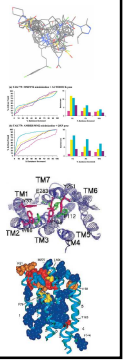
Scoring Function	AUC
PARAFIT08: Consensus 3 queries Shape Tanimoto	0.991
FRED2.2: CONSENSUS Chemscore, Oechemscore, Shapegauss, Chemgauss3, Screenscore, Plp	0.971
PARAFIT08: Superconsensus C Shape Tanimoto	0.905
PARAFIT08: Consensus all cmpds queries Shape Tanimoto	0.866
AUTODOCK3.0: Docked Energy	0.843
PARAFIT08: Superconsensus A Shape Tanimoto	0.789
FRED2.2: Shapegauss	0.780
FRED2.2: Chemgauss3	0.747
FRED2.2: Screenscore	0.603
FRED2.2: Chemscore	0.600
FRED2.2: Plp	0.499
HEX4.8: Docked Energy	0.483
PARAFIT08: Superconsensus D Shape Tanimoto	0.630
ROCS2.2: Combo Score (Shape Tanimoto + Scaled Color Score)	0.615
FRED2.2: Oehemscore	0.611
ROCS2.2: Shape Tanimoto	0.600
PARAFIT08: Shape Tanimoto	0.594
HEX4.8: Shape Tanimoto	0.581
PARAFIT08: Superconsensus B+D Shape Tanimoto	0.505
PARAFIT08: Superconsensus B Shape Tanimoto	0.413

Pérez-Nueno *et al. J. Chem. Inf. Model.* **2008**, *48*, 2146–2165.

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.



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...*

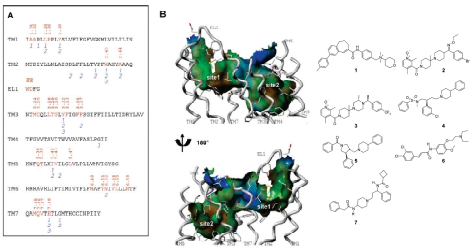
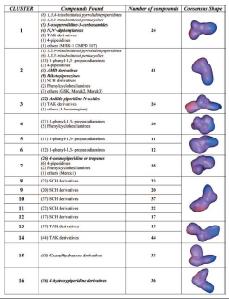


Figure 2. CCR5 TM binding cavity. A. Amino acid sequence. Residues with side chain pointing toward the cavity are written in red and pointed using Ballesteros numbering,[28] except in the extracellular domain 1 (EL1). Numbers indicated underneath the sequence summarize experimental mapping of receptor interaction site for nonpeptide antagonists **1**−**5**. 1: residues important for the efficiency of 1.[19,20,22] 2: residues important for the efficiency of 2 and Schering-Plough compound AD101.[19,20] and 3: residues important for the binding 4 and 5.[21] 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.

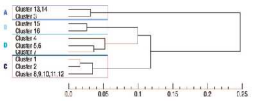
Kellenberg *et al. J. Med. Chem.* **2007**, *50*, 1294-1303.

Clustering the 424 CCR5 Ligands

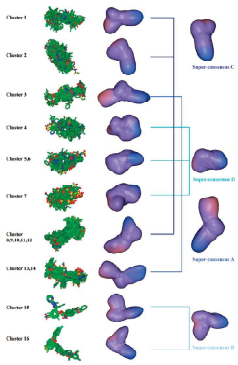
- Because it is not clear *a priori* which ligands might belong to which group, we first performed Wards hierarchical clustering of chemical fingerprints…
- We then used Kelley's method to find the optimal number of clusters (16)
- These were manually merged to 10 groups based on known CCR5 families

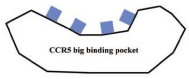


- SH consensus shapes were calculated for the 10 groups
- These were then compared in ParaFit (all-vs-all)
- Another round of Ward's clustering proposed four super-consensus clusters



From Consensus Shapes to Super-Consensus Clusters

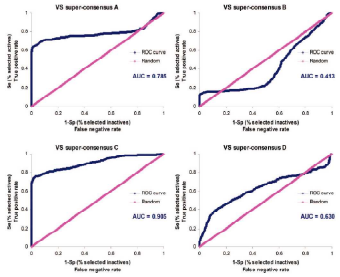




SC_A (87 compounds): **TAK derivatives**
Anilide piperidine *N*-oxides
SC_B (69 compounds): **Guanylhydrazone derivatives**
4-hydroxypiperidine derivatives
SC_C (184 compounds): **SCH derivatives**
1,3,4-trisubstituted pyrrolidinepiperidines
1,3,5-trisubstituted pentacyclics
5-oxopyrrolidine-3-carboxamides
***N,N'*-diphenylureas**
Diketopiperazines
AMD derivatives
SC_D (84 compounds): **1-phenyl-1,3- propanodiamines**
4-piperidines
1-phenyl-1,3- propanodiamines
Phenylcyclohexilamines
4-aminopiperidine or tropanes
4-piperidines

Using Super-Consensus Shapes as VS Queries

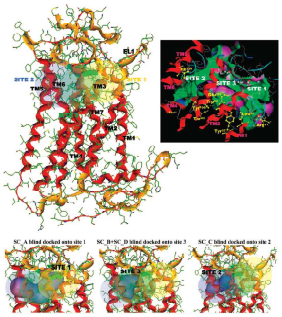
- **Each SC pseudo-molecule was used as a VS query:**



- **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**

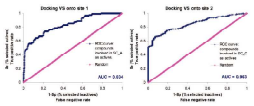


- **SC-A docks to Site-1 (TMs 1, 2, 3, 7)**
- **SC-C docks to Site-2 (TMs 3, 5, 6)**
- **B and D dock to Site-3 (TMs 3, 6, 7)**

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

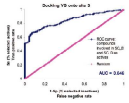
- To confirm the SC shapes were matched to their predicted target sites, docking based VS was repeated for each ligand using:
- SC-As treated as actives for Site 1 (SCs B, C, D treated as inactives)
- SC-Cs treated as actives for Site 2 (SCs A, B, D treated as inactives)
- SC-B/Ds assumed active for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

A -> Site-1



C -> Site-2

B,D -> Site-3



- 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
- IQS, Universitat Ramon-Llull

Papers: http://www.loria.fr/~dritchie/

ParaSurf + ParaFit: http://www.ceposinsilico.de/