Tutorial 3 - Compound profiling using similarity between protein binding sites and shape analysis of ligand

Jérémy Desaphy, Guillaume Bret, Didier Rognan and Esther Kellenberger

University of Strasbourg, Medalis Drug Discovery Center, UMR7200, 74 route du Rhin, 67400 Illkirch

1. Introduction: aim and context

Over the previous decades, drug discovery efforts have focused on the design of selective drugs, assuming that targeting a key protein in a single biological process causes the beneficial therapeutic effect. Growing experimental evidences have recently shifted the single-target to multi-target paradigm,[1] thereby boosting the development of computational approaches to identify all possible ligands for all possible targets.[2-5] The new research field, called chemogenomics or *in silico* polypharmacology, has firstly proposed efficient ligand-centric methods for structure-activity data mining. Protein-centric methods have complemented the toolbox. They have allowed prediction for protein without known ligands, yet their usage necessitates the three-dimensional molecular structure of the protein.

In this tutorial, we will work on the issue of ligand profiling and answer the question "Can we find secondary targets of a ligand whose primary target is known?". To that end, we will test two methods based on 3D-shape comparison:

- *A protein-centric method: 3D similarity between protein binding sites*

  The ligand-binding site in its specific protein constitutes the reference, which is compared to each entry of a dataset made of druggable binding sites in therapeutically relevant proteins. The sites are defined from the crystal structure of ligand/protein complexes.

- *A ligand-centric method: 3D similarity between ligands*

  A conformational ensemble representing the ligand constitutes the reference, which is compared to the low energy structures of all high affinity ligands selected for the therapeutically relevant proteins.

The basic idea behind the protein-centric method is that two similar binding sites can accommodate the same ligand.[6] The ligand-centric method assumes that ligands with similar

shape are prone to bind to the same protein.[7] The general screening strategy is summarized on Figure 1.

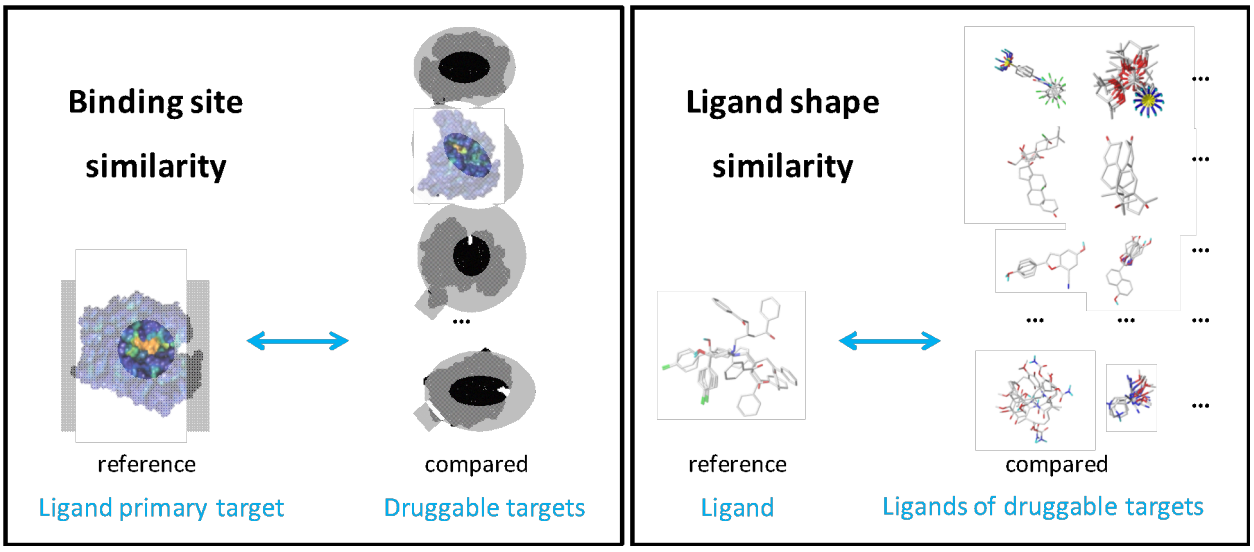

Figure 1: Protein-centric (left) and ligand-centric (right) strategies to ligand profiling.

The present tutorial aims at predicting known and putative secondary targets of haloperidol. Haloperidol (Figure 2) is a phenyl-piperidinyl-butyrophenone that is used primarily to treat schizophrenia and other psychoses. It is also used in schizoaffective disorder, delusional disorders, ballism, and tourette syndrome and occasionally as adjunctive therapy in mental retardation and the chorea of huntington disease. It is a potent antiemetic and is used in the treatment of intractable hiccups. Haloperidol have common adverse effects (>1% incidence) because it is not highly selective to its primary target, which is dopamine receptor. For example, the interaction of the drug with the receptors of acethylcholine causes constipation. Hypotension consequent to adrenergic receptor blockade is another example of haloperidol promiscuous binding.

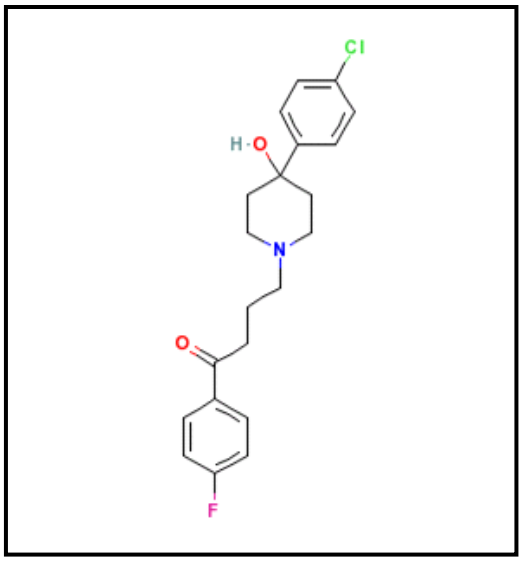

Figure 2: Chemical structure of haloperidol

2. Material: query and searched database

The reference in the ligand-centric method is haloperidol. The reference in the protein-centric method is human dopamine D3 receptor (Table 1).

Functional class[a]	Target			Ligand
	Name[a]	organism	PDB ID[b]	
Class A G-protein coupled receptor: Dopamine receptor	dopamine D3 receptor	human	3pbl	haloperidol

Table 1: Description of the reference entries

[a] as defined in IUPAR-DB (www.iuphar-db.org/), [b] identifier in the Protein Databank (www.rcsb.org/pdb)

Haloperidol is profiled against fifteen druggable targets in the Protein Databank. The name and functional class of targets are given in Table 2. For each target, ten different ligands were collected from the sc-PDB (ligand co-crystallized with the protein) [8] and from chEMBL database (drug targeting the protein, or ligand with IC_{50} <50 nM).

Functional class[a]	Target		Database ID[b]	
	Name[a]	Organism	Protein	Ligand
Class A G-protein coupled receptor: adrenoceptor	β1-adrenoceptor	turkey	2ycw	2rh1,2ycw, atenolol, oxprenolol, penbutolol, protokylol, timolol, CHEMBL1788270, CHEMBL276659, CHEMBL51667
		human	2rh1	
Class A G-protein coupled receptor: acethylcholine receptors (muscarinic)	M2 receptor	human	3uon	3uon, CHEMBL10272, CHEMBL135645, CHEMBL1779036, CHEMBL194837, CHEMBL495531, procyclidine, scopolamine, tolterodine, trospium
Class A G-protein coupled receptor: Chemokine receptor	CCR5	human	4mbs	4mbs, CHEMBL1178786, CHEMBL182940, CHEMBL207004, CHEMBL2178576, CHEMBL322251, CHEMBL322693, CHEMBL392659, CHEMBL481068, CHEMBL540366
Class B G-protein coupled receptor: corticotropin-releasing factor receptor	CRF1	human	4k5y	4k5y, CHEMBL115142, CHEMBL1819077, CHEMBL1939593, CHEMBL2087552, CHEMBL482950, CHEMBL484158, CHEMBL497653, CHEMBL525716, CHEMBL573978
3-Ketosteroid receptor	Androgen receptor	mouse	2qpy,	1e3g, 1gs4, 1t7r, 1z95, 2ax6, 2axa, 2hvc, 2pnu, 2qpy, 3b5r
		human	3b5r	
Estrogen receptor	ER beta	human	2fsz	1hj1, 1l2j, 1nde, 1qkn, 1u3r, 1u3s, 1x76, 1x78, 2fsz, 2j7x
		rat	2j7x	
heat shock protein	HSP90alpha	human	3owd, 4efu	1yet, 2bz5, 2qf6, 2qg0, 2qg2, 2vcj, 2xhr, 3ekr, 3owd, 4efu
Carboxylic ester hydrolase	acethylcholinesterase	electric ray	1zgc, 3i6m	1e66, 1eve, 1qon, 1zgc , 2gyw, 2ha6, 2xi4, 3i6m, 4arb, 4b7z
Protein-serine/threonine kinase	Cyclin-dependent kinase 2 (CDCK2)	human	1gij, 1w0x	1di8, 1dm2, 1e9h, 1gij, 1ke8, 1oit, 1p2a, 1w0x, 2bts, 2c5x
Protein-serine/threonine kinase	Aurora kinase	human	2np8, 2x81	2np8, 2x81, 3d14, 3dj5, 3lau, 3myg, 3o50, 3p9j, 3r21, 3unz
Aspartic endopeptidase	Beta secretase	human	2fdp, 4djv	2b8v, 2f3f, 2fdp, 2oah, 2q15, 2qu3, 2vij, 3exo, 3pi5, 4djv
Aspartic endopeptidase	renin	human	2g1o, 3vye	2bks, 2bkt, 2g1n, 2g1o, 2g1r, 2g1s, 2g1y, 2g20, 2g24, 3vye
Serine endopeptidase	Thrombin	human	3rlw, 3sv2	3da9, 3p17, 3qwc, 3rlw, 3rml, 3shc, 3sv2, 3u98, 3utu, 4bah
Methyltransferase	Thymidylate synthase	pneumocystis carinii	1ci7, 3uwl	1axw, 1ci7, 1f28, 1f4g, 1jtq, 2aaz, 2fto, 3uwl, 4fog, 4lrr
Carbon-oxygen lyase	Carbonic anhydrase	human	1a42, 3mhm	1a42, 2nnv, 3bet, 3dbu, 3f4x, 3ffp, 3k2f, 3m67, 3mhm, 3n0n

Table 2: Description of the compared entries

[a] as defined in ENZYME for enzymes (enzyme.expasy.org/), IUPAR-DB for non-enzymatic receptors (www.iuphar-db.org/), [b] identifier in Protein Databank (www.rcsb.org/pdb), drug name, or identifier in chEMBL (www.ebi.ac.uk/chembl/)

3. Methods

Site comparison will be performed using Volsite and Shaper (UMR-7200).[9] Ligand shape analysis will be performed using ROCS. [7] A brief overview of the methods is given below.

VolSite and Shaper

The program VolSite detects cavities in a protein as illustrated in Figure 3; *step1*, the protein is placed into a cubic grid; *step2*, grid is pruned according to the protein atomic coordinates; *step3*, grid is further pruned in order to discard non-buried points (and optionally the points sitting too far from any ligand atom). Remaining grid points are colored according to the pharmacophoric properties of nearest protein atoms (hydrophobic, aromatic, H-Bond acceptor, negative ionizable, H-Bond acceptor/donor, H-Bond donor, positive ionizable, null); *step4*, each ensemble of contiguous cell defines one cavity, adjacent cavities are merged. By default, the grid is centered on the ligand center, the grid edge is 20Å and the grid resolution is 1.5Å.

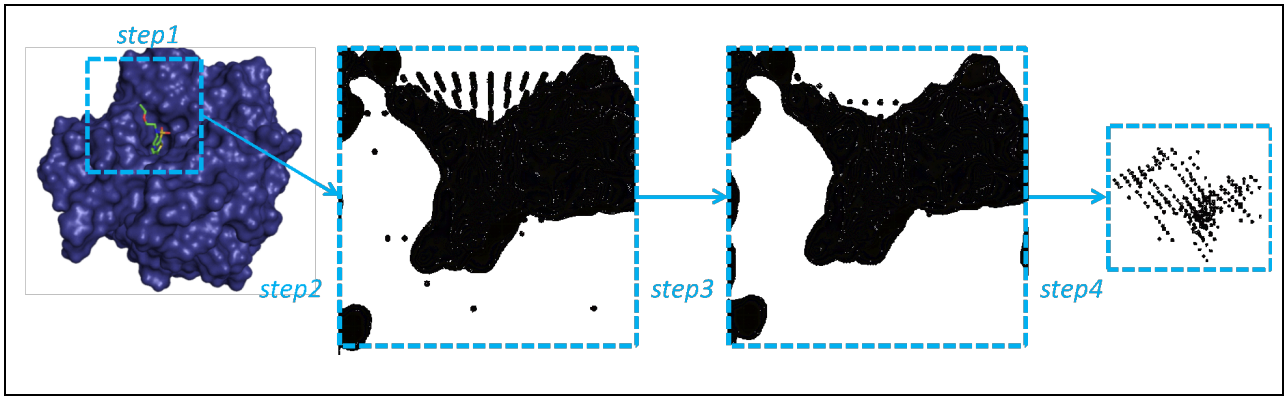

Figure 3: Principle of cavity detection in VolSite

The negative images of binding sites created with VolSite are input for the 3D-alignment program Shaper. Shaper superimposes a query site to a compared site by maximizing the

geometric overlap of corresponding pruned colored grids. Best geometric solutions are ranked depending on the match of properties of overlaid grid points, using a *refTversky* score as defined in equation 1.

$$\frac{O_{r,c}}{0.95\, I_r + \ 0.05\, I_c} \qquad \text{(eq 1)}$$

where $O_{r,c}$ represents the grid overlap, I_r the reference grid and I_c the compared grid.

ROCS

ROCS is a fast shape comparison application. It represents the molecular volume as atom-centered Gaussian functions. It maximizes the shared volume between the query molecule and the compared molecule by optimizing the overlap of their Gaussian functions (rigid body motions).

4. Programs and input data

Volsite, Shaper and ROCS are installed in the computer rooms available during the summer school. Volsite and Shaper are available upon request to Didier Rognan (rognan@unistra.fr). They are freely available for academic purposes. Shaper requires licensing for OEChem TK. OEChem TK and ROCS are developed and distributed by OpenEye Scientific Software (www.eyesopen.com).

The three programs are called *via* command lines. To do the exercises, open a Linux session and launch a terminal. Three-dimensional structures will be analyzed using the graphical interface of Molecular Operating Environment (MOE, Chemical Computing Group, Inc.).

Haloperidol reference structure was prepared as follows: the 2D structure was downloaded from chEMBL, ionized using filter v4.40 (OpenEye Scientific Software, Inc.) and folded using

Corina v3.40 (Molecular Network, GmbH). Conformers were generated using default settings of omega2 v2.4.6 (OpenEye Scientific Software, Inc.). Default parameters set the maximal number of conformers to 200. No limits were fixed in the case of haloperidol, yielding 590 conformers. The compared ligands (Table 1) were retrieved from sc-PDB (ligand.mol2 files) or downloaded from chEMBL and ionized using filter, then folded using Corina and submitted to conformational sampling using omega2 (default settings).

The three-dimensional structures of proteins were downloaded from sc-PDB.

The name of and path to input files are detailed in Table 3. For the sake of time, output files are given too.

File description	Path from working directory ($WORKDIR)	File name
Input: Reference 3D protein structure		
Crystal structure of dopamine receptor	REF	*D3receptor-3pdbl_protein.mol2*
Binding cavity in dopamine receptor	REF	*D3receptor-3pdbl_cavity6.mol2*
Ligand co-crystallized with dopamine receptor	REF	*D3receptor-3pdbl_ligand.mol2*
Input: Reference 3D ligand structure		
Lowest energy conformer of haloperidol	REF	*haloperidol.mol2*
590 conformers of haloperidol	REF	*haloperidol_multiconf.mol2*
Input: Compared 3D protein structures		
25 crystal structures of 15 proteins	TARGET	*${Name}-${ ID}_protein.mol2*
25 binding cavities in 15 proteins	CAVITY	*${Name}-${ ID}_cavity.mol2*
Input: Compared 3D ligands		
diastereoisomers of 10 X 15 ligands	LIGAND	*${Name}-${ ID}_ligand.mol2*
Conformers of 10 X 15 ligands (max. 200 conformers per ligand)	LIGAND	*${Name}-${ ID}_ligandmulticonf.mol2*
Execution: scripts		
Screen by binding site similarity	EXE	*shaper.bash*
Screen by ligand shape	EXE	*rocs.bash*
Output: binding site comparison		
Ranked list of target proteins	SHAPER	*Shape_res.csv*
Compared cavity 3D-aligned to reference cavity	SHAPER	*3Dreceptor-3pdb_${Name}-${ ID}.pdb*
Output: ligand shape analysis		
Log file	ROCS	*${Name}-${ ID}_1.log*
Parameter	ROCS	*${Name}-${ ID}_1.param*
Job summary	ROCS	*${Name}-${ ID}_1. status*
Overlay of reference and compared ligands	ROCS	*${Name}-${ ID}_1.mol2*

Ranked list of ligands	ROCS	*rocs_res.csv*

Table 3: Description of input and output files. See Table 1 for the list of *Name* and *ID*

5. Exercise 1: Binding site comparison

Haloperidol is profiled by comparing its primary target, namely the D3 dopamine receptor, to fifteen protein targets.

How to run Volsite?

`cd /tmp/CS3-3D` `source CS3.bash`	In linux terminal, go to working directory then define environment variables
`IChem`	Get the instructions to execute the program

`IChem volsite REF/D3receptor-3pbl_protein.mol2`
`REF/D3receptor-3pbl_ligand.mol2`
 	
`/usr/local/chemo/moe2013/bin/moe`
`Open`
`REF/D3receptor-3pbl_protein.mol2`
`REF/D3receptor-3pbl_ligand.mol2`
`CAVITY_N1_6.mol2`
`SiteView`
	Compute the cavity around the bound ligand in the protein. Here exemplified on the reference binding site.
	
	Visualize cavity, ligand and protein in MOE (open the three files and select the "site" view)

In practice, all cavities have been pre-computed.

How to run Shaper?

`cd $WORKDIR`	In linux terminal, go to working directory
`Shaper`	Get the instructions to execute the program

`Shaper -r REF/D3receptor-3pbl_cavity6.mol2`
`-c CAVITY/CCR5-4mbs_caviy6.mol2`
`-o D3receptor-3pbl_CCR5-4mbs.pdb`
`-rn D3receptor`
`-cn CCR5-4mbs`
	Compare two cavities. Here exemplified on dopamine D3 receptor and CCR5

To screen the full dataset, execute a script which iterates the command line over the 25 entries.

`cd $WORKDIR`	In linux terminal, go to working directory
`rm Shape_res.csv`	Remove the result file from previous calculations (append mode)
`EXE/shaper.sh`	Execute the script
`Sort –k5 Shape_res.csv`	Visualize output file

Analyze results.

To that end, consider that ColorRefTversky above 0.45 is significant (0.45 represent three times the standard deviation added to the mean score of the normal distribution obtained for the comparisons of the reference binding site to the 9 427 entries of sc-PDB, release 2013).

Two proteins have passed the score threshold: adrenergic receptor β1 (2ycw) and acethylcholinesterase (3i6m). The literature supports the cross binding of the ligands of the dopamine receptors to adrenergic receptors. Haloperidol is known to bind receptors of acethylcholine. Here we suggest that an enzyme whose substrate is acetylcholine can also be a target of haloperidol.

Going further…Observe aligned binding sites

`cd $WORKDIR`	In linux terminal, go to working directory
`cd ALIGN`	Go to the ALIGN directory

`IChem realign`
`../SHAPER/D3receptor-3pbl_\`
`adrenoreceptor-2ycw.pdb`
`../CAVITY/adrenoreceptor-\`
`2ycw_cavity6.mol2`
`../TARGET/adrenoreceptor-\`
`2ycw_protein.mol2`
`../LIGAND/adrenoreceptor-\`
`2ycw_ligand.mol2`
	Create in the current directory aligned files:
	`rot_adrenoreceptor-2ycw_protein.mol2`
	`rot_adrenoreceptor-2ycw_ligand.mol2`

Visualize *rot_adenoreceptor-2ycw_protein.mol2* and *rot_adenoreceptor-2ycw_ligand.mol2* files using MOE. Compare aligned structures to reference files (*D3receptor-3pbl_protein.mol2* and *D3receptor-3pbl_ligand.mol2*).

`/usr/local/chemo/moe2013/bin/moe`
`Open`
`REF/D3receptor-3pbl_protein.mol2`
`REF/D3receptor-3pbl_ligand.mol2`
`rot_adrenoreceptor-\`
`2ycw_protein.mol2`
`rot_adrenoreceptor-\`
`2ycw_ligand.mol2`
`SiteView`
	Visualize cavity, ligand and protein in MOE (open the four files and select the "site" view)

6. Exercise 2: Ligand shape analysis

Haloperidol is profiled by direct comparison of computed 3D-structures. Haloperidol is the reference, ten ligands are considered for each of the fifteen protein targets.

How to run ROCS?

`cd $WORKDIR`	In linux terminal, go to working directory
`rocs --help simple`	Get the instructions to execute the program

`rocs -query REF/haloperidol.mol2`
`-dbase LIGAND/CCR5-4mbs_ligand\`
`multiconf.mol2`
`-prefix CCR5-4mbs`
`-oformat mol2`
`-maxhits 1`
	Compare two ligands.
	Here the reference is the lowest energy conformer of haloperidol, and the compared ligand is represented with a conformation ensemble.

`more CCR5-4mbs_1.rpt`	Display alignment scores

To screen the full dataset, execute a script which iterates the command line over the 150 entries.

`cd $WORKDIR`	In linux terminal, go to working directory
`rm rocs_res.csv`	Remove the result file from previous calculations
`EXE/rocs.sh`	Execute the script
`cat rocs_res.csv`	Visualize output file

Analyze results.

To that end, consider the highest TanimotoCombo scores. The scores of two proteins have exceeded 1.0: adrenergic receptor β1 (with compared ligand protokylol) and acethylcholinesterase (with compared ligand from 1eve PDB entry). These two proteins were also retrieved by the protein-centric approach to profiling.

Visualize *adenoreceptor-protokylol_hits_1.mol2* and *acethylcholine_1eve_hits_1.mol2* files using MOE. Compare aligned structures to the reference file (*haloperidol.mol2*).

Note that the screening was based on a single conformer for the reference and an ensemble of conformers for each compared ligand. Observe results obtained using multiple conformers for the reference (*haloperidol_multiconf.mol2* file, -mcquery true option in rocs), and conclude that the top ranked proteins are the same, yet overall there was global increase of scores.

7. Conclusions

We have demonstrated that 3D computing methods are suitable to the profiling of haloperidol. Similarity between protein binding sites and similarity between ligands yielded the same results and identified two potential secondary targets that are likely true positives, as supported by experimental evidences. Last ligand-centric approach required several ligands (in our example the ligands co-crystallized with the target did not well superimposed to haloperidol).

8. Bibliography

1. Peters, J.-U., Polypharmacology - Foe or Friend? *Journal of Medicinal Chemistry* **2013,** 56, (22), 8955-8971.

2. Koutsoukas, A.; Simms, B.; Kirchmair, J.; Bond, P. J.; Whitmore, A. V.; Zimmer, S.; Young, M. P.; Jenkins, J. L.; Glick, M.; Glen, R. C.; Bender, A., From in silico target prediction to multi-target drug design: Current databases, methods and applications. *Journal of Proteomics* **2011,** 74, (12), 2554-2574.
3. Meslamani, J.; Bhajun, R.; Martz, F.; Rognan, D., Computational Profiling of Bioactive Compounds Using a Target-Dependent Composite Workflow. *Journal of Chemical Information and Modeling* **2013,** 53, (9), 2322-2333.
4. Reddy, A. S.; Zhang, S., Polypharmacology: drug discovery for the future. *Expert Review of Clinical Pharmacology* **2013,** 6, (1), 41-47.
5. Rognan, D., Chemogenomic approaches to rational drug design. *Br J Pharmacol*. **2007,** 152, (1), 38–52.
6. Vulpetti, A.; Kalliokoski, T.; Milletti, F., Chemogenomics in drug discovery: computational methods based on the comparison of binding sites. *Future Medicinal Chemistry* **2012,** 4, (15), 1971-1979.
7. Hawkins, P. C. D.; Skillman, A. G.; Nicholls, A., Comparison of Shape-Matching and Docking as Virtual Screening Tools. *Journal of Medicinal Chemistry* **2006,** 50, (1), 74-82.
8. Kellenberger, E.; Muller, P.; Schalon, C.; Bret, G.; Foata, N.; Rognan, D., sc-PDB: an Annotated Database of Druggable Binding Sites from the Protein Data Bank. *Journal of Chemical Information and Modeling* **2006,** 46, (2), 717-727.
9. Desaphy, J.; Azdimousa, K.; Kellenberger, E.; Rognan, D., Comparison and druggability prediction of protein-ligand binding sites from pharmacophore-annotated cavity shapes. *Journal of Chemical Information and Modeling* **2012,** 52, (8), 2287-99.