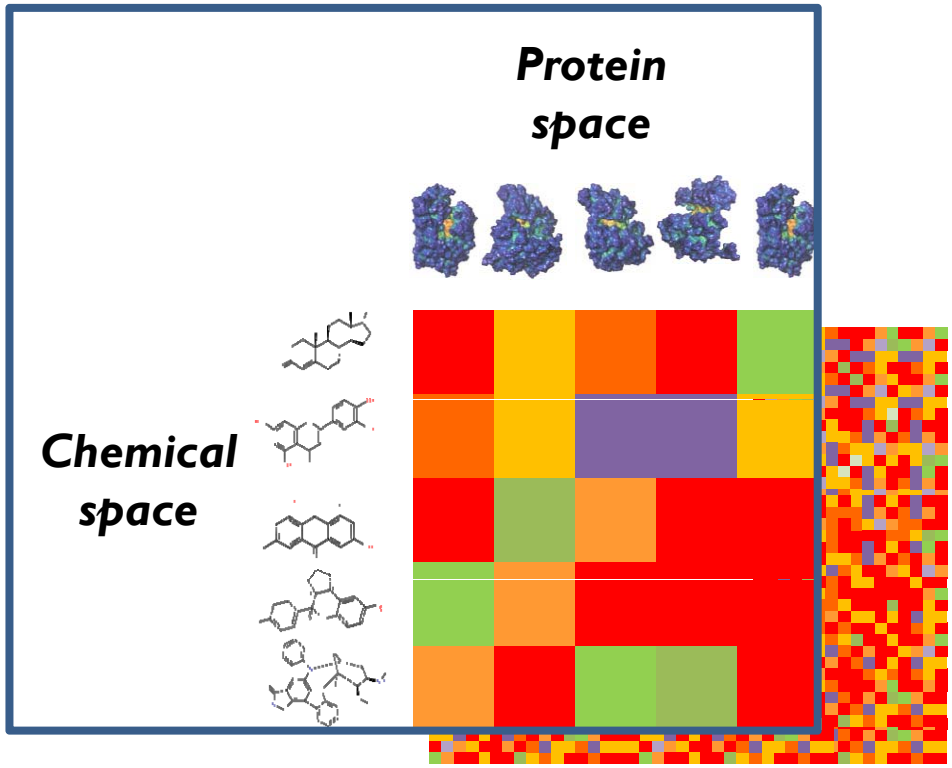
Compound profiling using similarity between protein binding sites and shape analysis of ligand

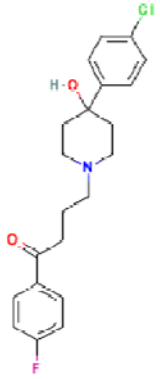
Chemogenomics or *in silico* poly-pharmacology



What are 3D-computational approaches to identify all possible ligands for all possible targets?

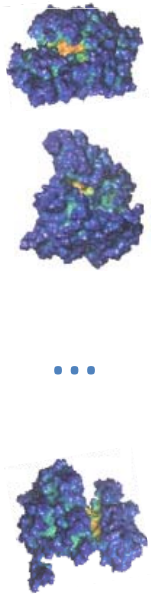
Compound profiling using **similarity between protein binding sites** and shape analysis of ligand

Protein-centric method



query ligand

putative targets in the PDB ?



protein 1

protein 2

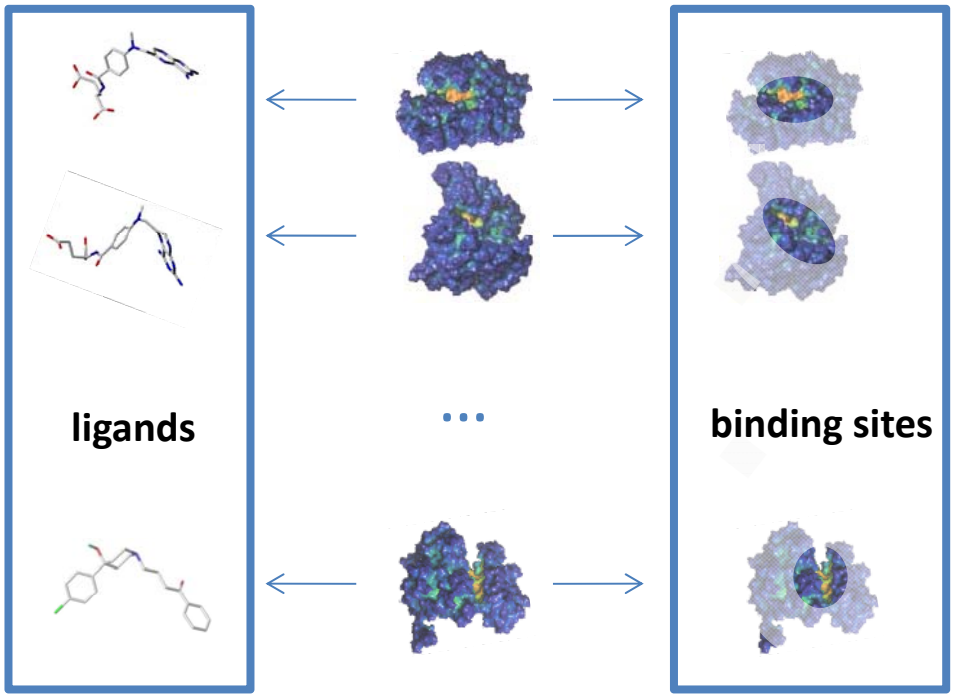
...

protein n

List for exp. validation

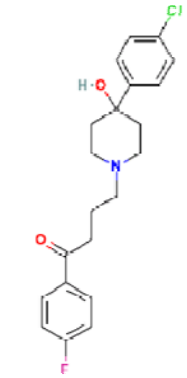
Compound profiling using **similarity between protein binding sites** and shape analysis of ligand

Searched database: 3D binding sites (sc-PDB)

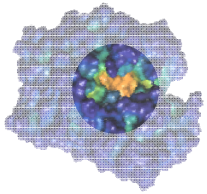


Compound profiling using **similarity between protein binding sites** and shape analysis of ligand

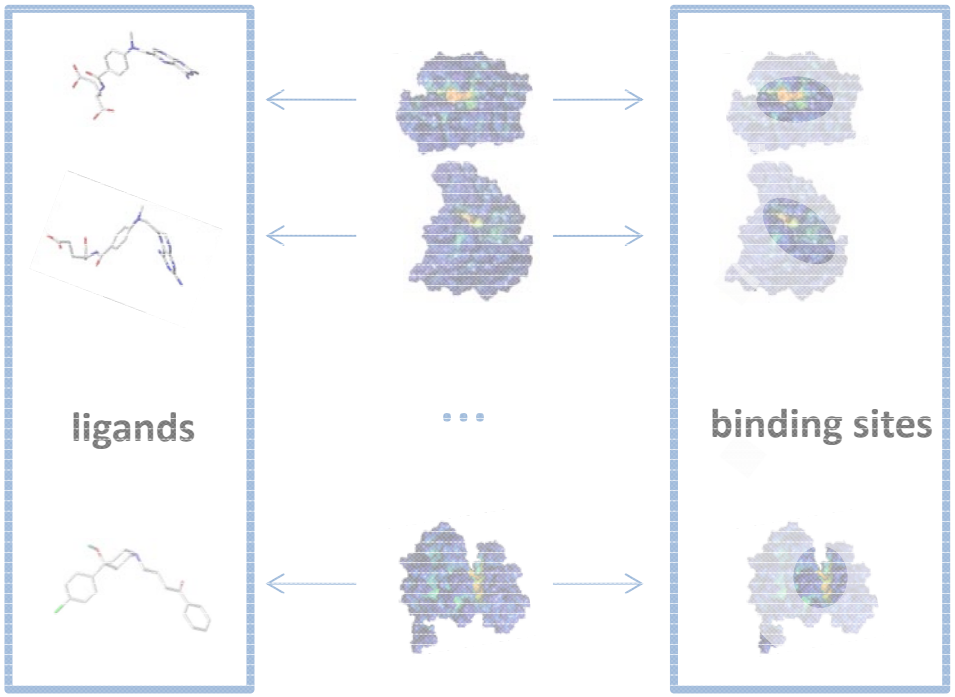
Query: ligand binding site of the primary target



query ligand

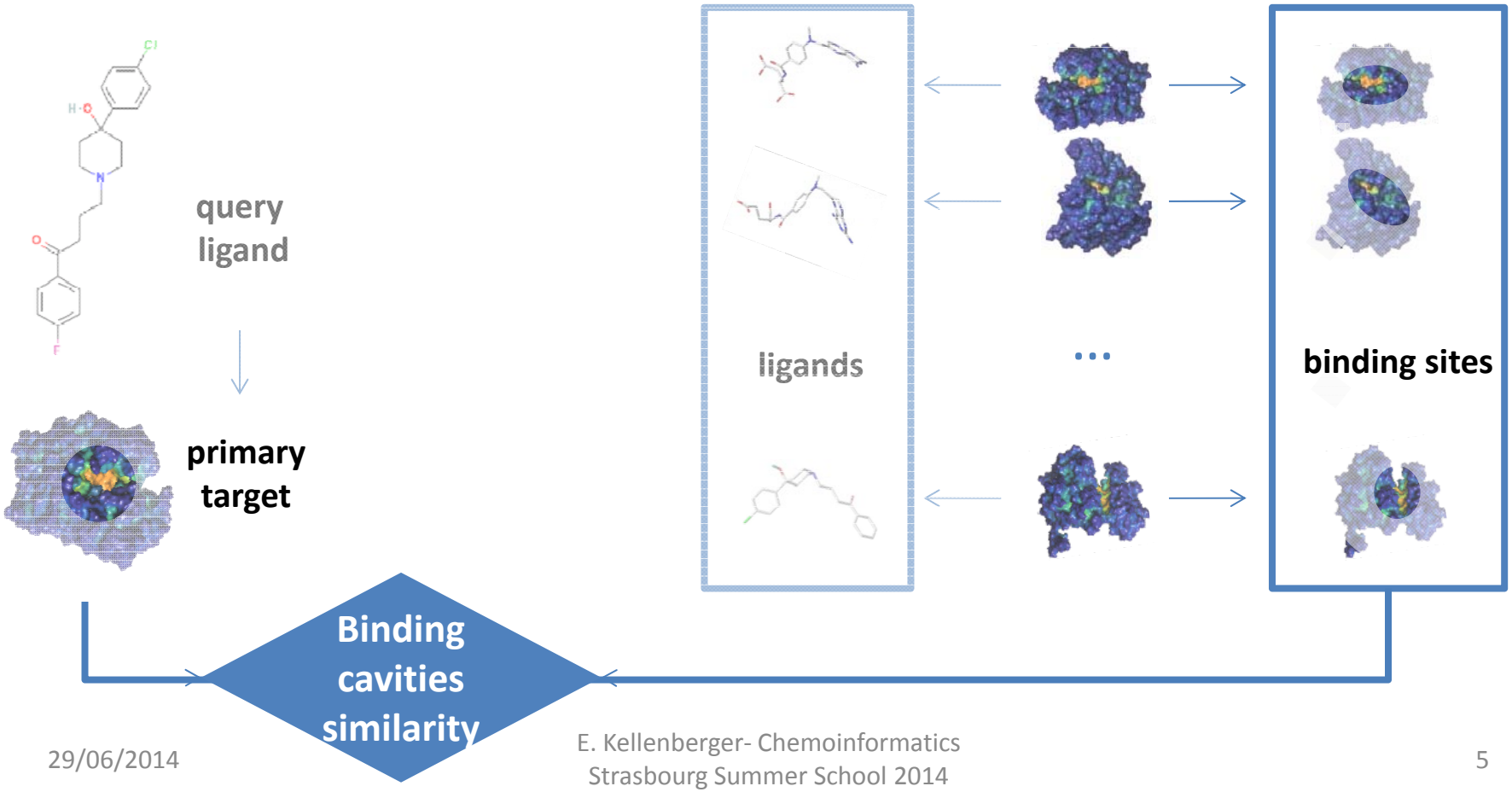


primary target



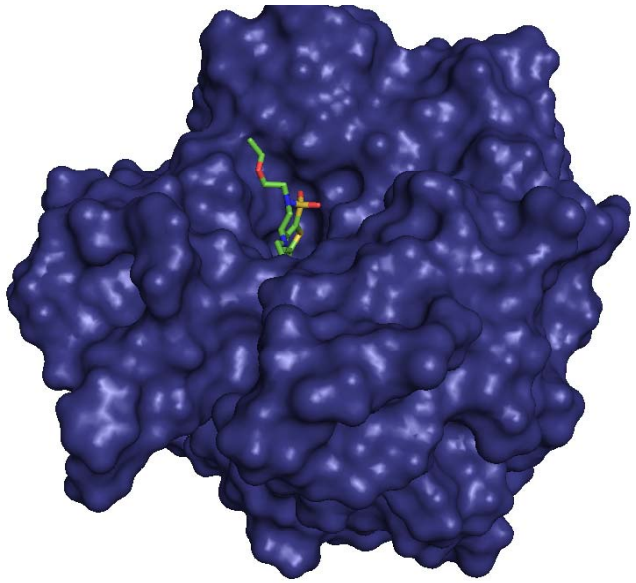
Compound profiling using similarity between protein binding sites and shape analysis of ligand

Method: comparison of molecular shape and properties

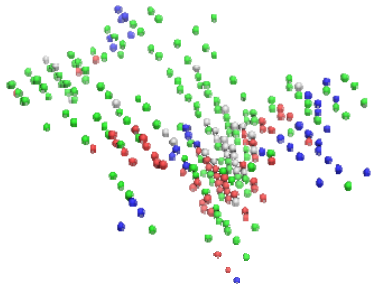


Compound profiling using **similarity between protein binding sites** and shape analysis of ligand

Method: 1- from ligand binding site to cavity



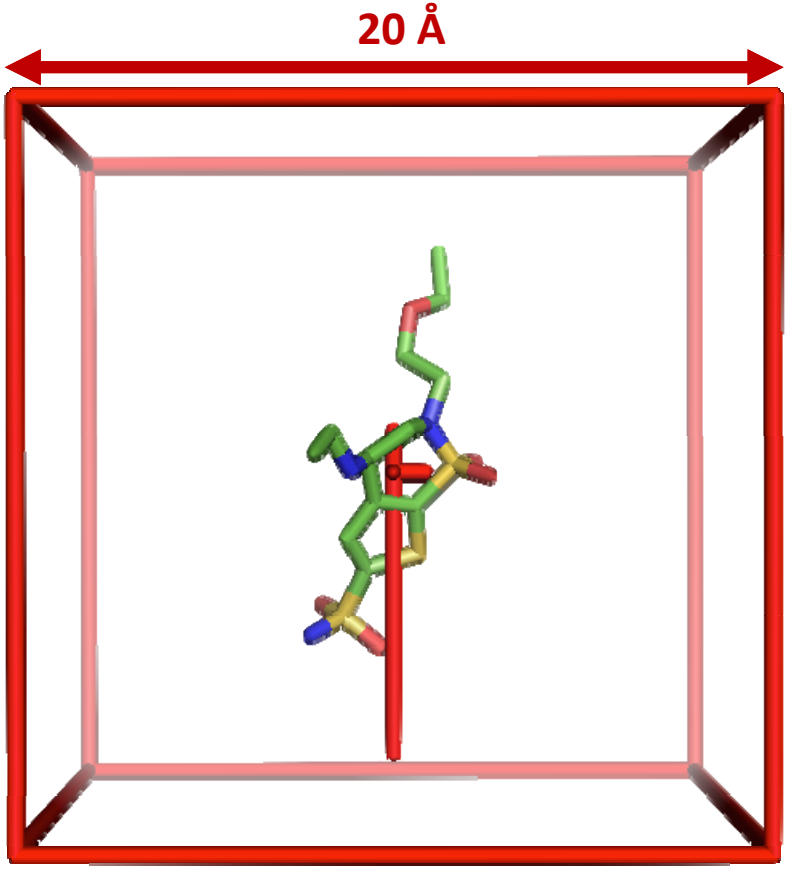
Volsite



Compound profiling using **similarity between protein binding sites** and shape analysis of ligand

Method: 1- from ligand binding site to cavity

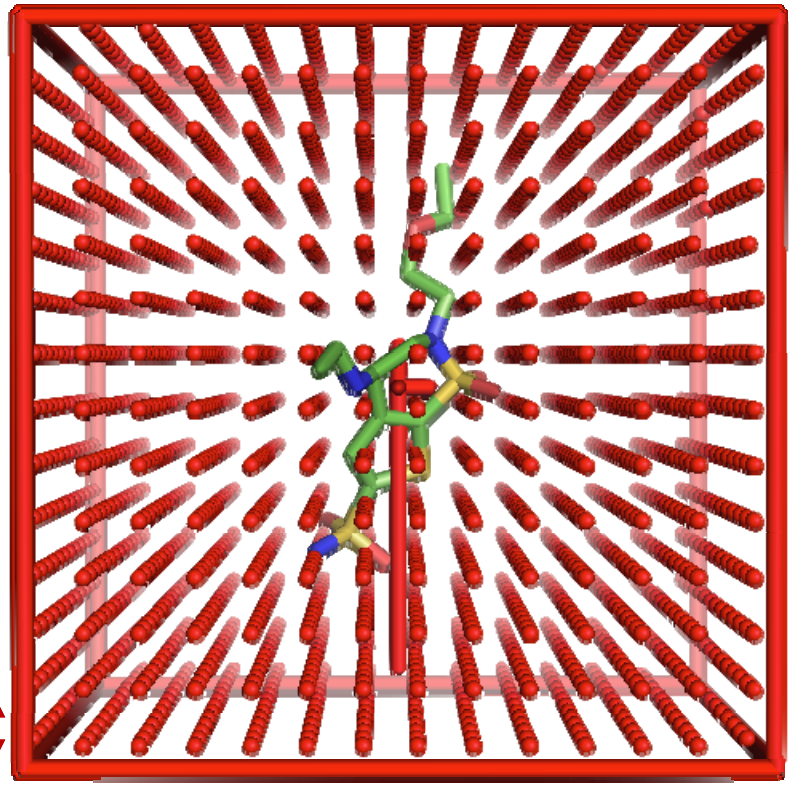
- box centered on the **ligand**



Compound profiling using **similarity between protein binding sites** and shape analysis of ligand

Method: 1- from ligand binding site to cavity

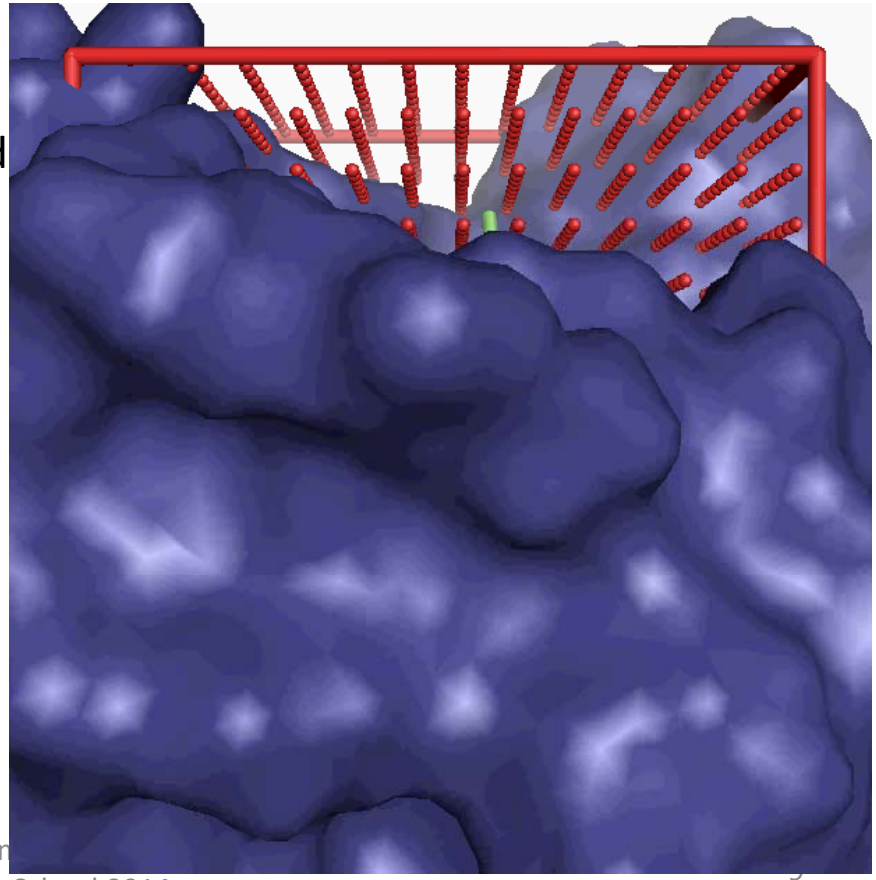
- box centered on the **ligand**
- box tranformed into a cubic grid



Compound profiling using **similarity between protein binding sites** and shape analysis of ligand

Method: 1- from ligand binding site to cavity

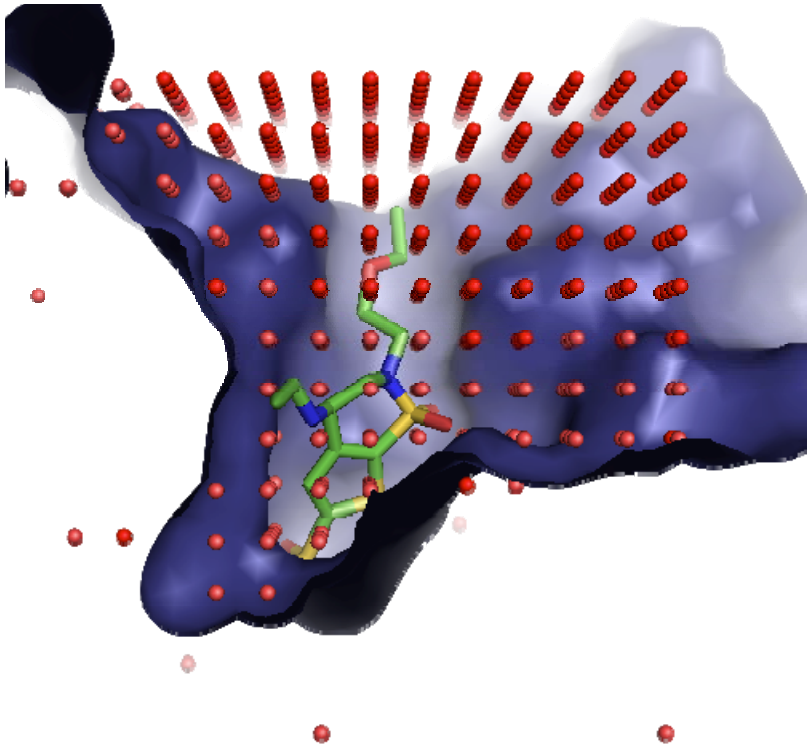
- box centered on the **ligand**
- box tranformed into cubic grid
- grid points superimposed to to **protein** atoms are removed



Compound profiling using **similarity between protein binding sites** and shape analysis of ligand

Method: 1- from ligand binding site to cavity

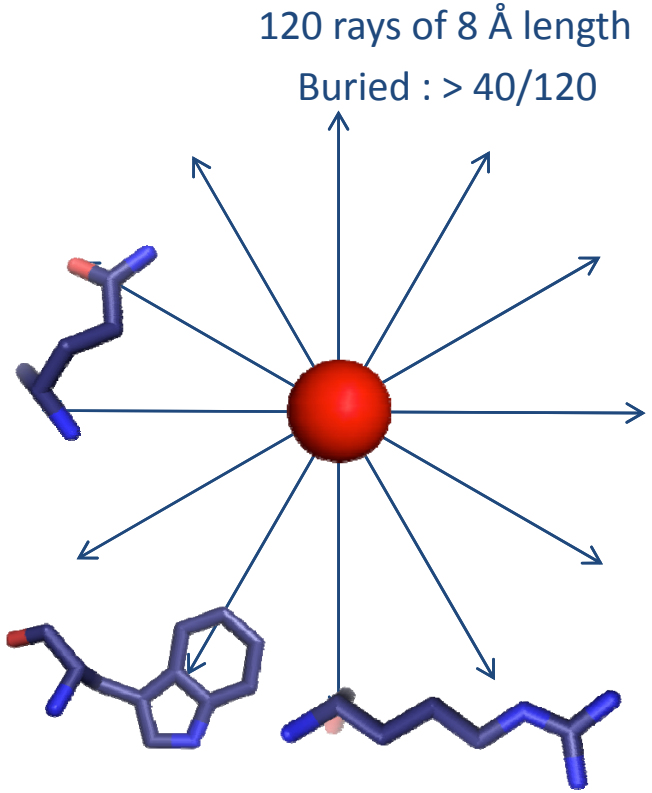
- box centered on the **ligand**
- box tranformed into cubic grid
- grid points superimposed to to **protein** atoms are removed



Compound profiling using **similarity between protein binding sites** and shape analysis of ligand

Method: 1- from ligand binding site to cavity

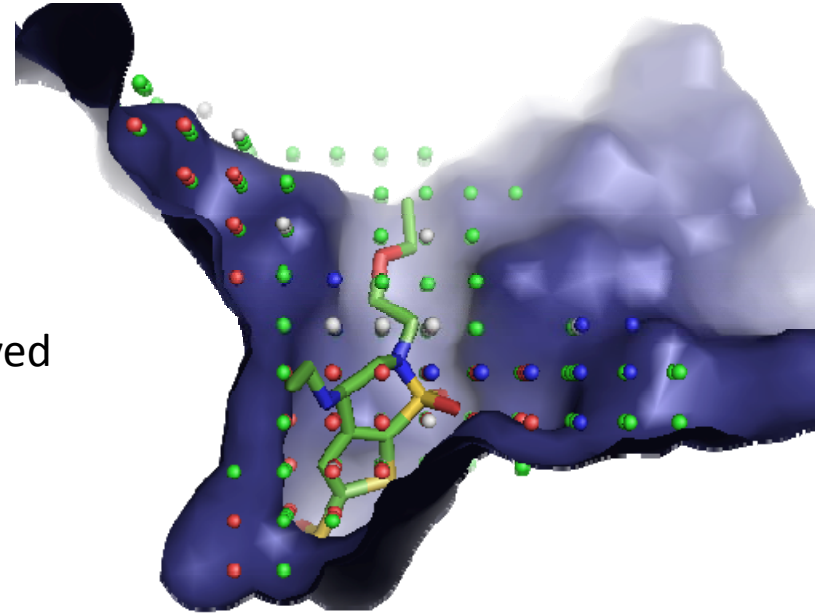
- box centered on the **ligand**
- box tranformed into cubic grid
- grid points superimposed to to **protein** atoms are removed
- Non-buried grid points are removed



Compound profiling using **similarity between protein binding sites** and shape analysis of ligand

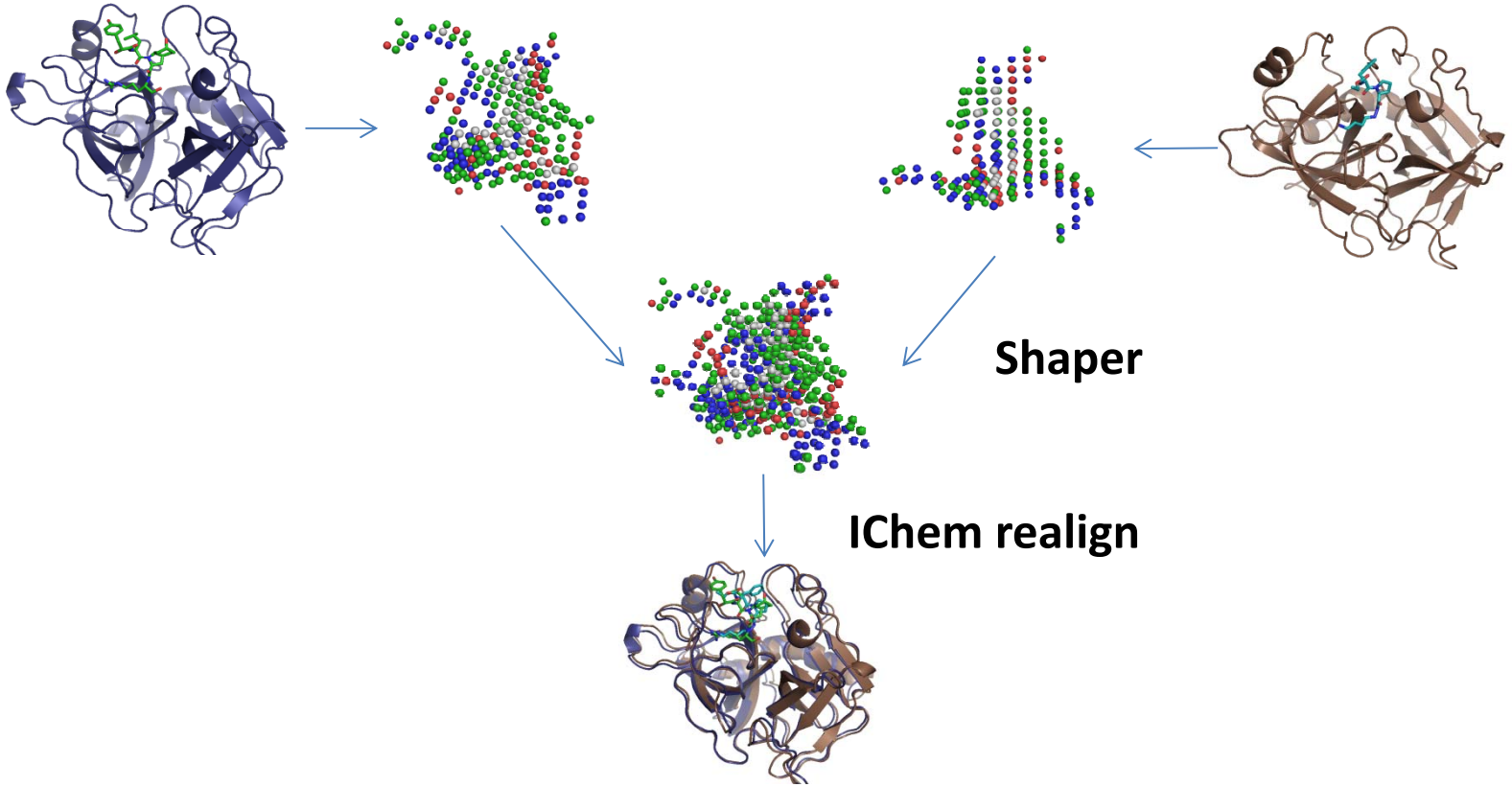
Method: 1- from ligand binding site to cavity

- box centered on the **ligand**
- box tranformed into cubic grid
- grid points superimposed to to **protein** atoms are removed
- Non-buried grid points are removed
- Grid points colored according to binding property of neighboring protein atoms (hydrophobic, aromatic, H-bond donor, H-bond acceptor, H-bond donor/acceptor, positive, negative, **null**)



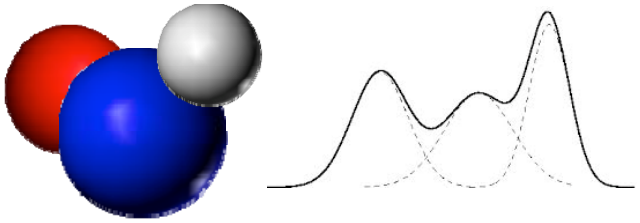
Compound profiling using **similarity between protein binding sites** and shape analysis of ligand

Method: 2- 3D alignment of two cavities

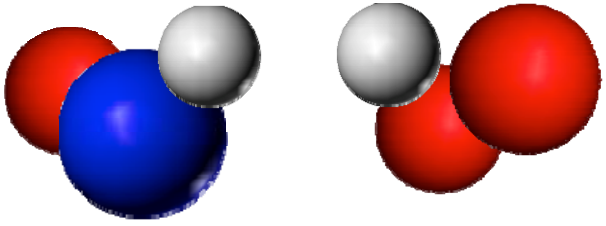


Compound profiling using **similarity between protein binding sites** and shape analysis of ligand

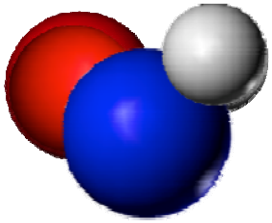
Method: 2- 3D alignment of two cavities



Molecule To Gaussians



Reference (R) Compared (C)

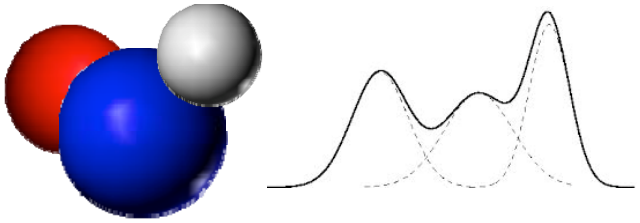


Rotation
Translation

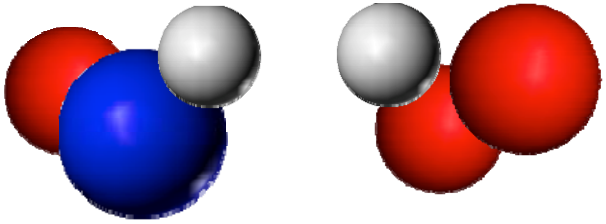
$$Tc = \frac{V_{R.C}}{V_R + V_C - 2V_{R.C}}$$

Compound profiling using **similarity between protein binding sites** and shape analysis of ligand

Method: 3- scoring the alignment

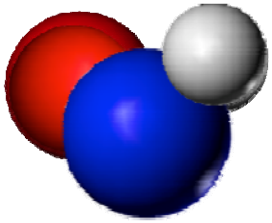


Molecule To Gaussians



Reference (R) Compared (C)

Rotation
Translation

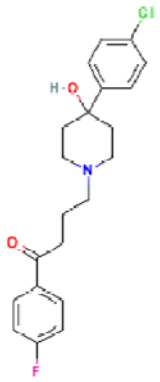


$$refTversky = \frac{O_{R.C}}{0.95I_R + 0.05I_C}$$

$$Tc = \frac{V_{R.C}}{V_R + V_C - 2V_{R.C}}$$

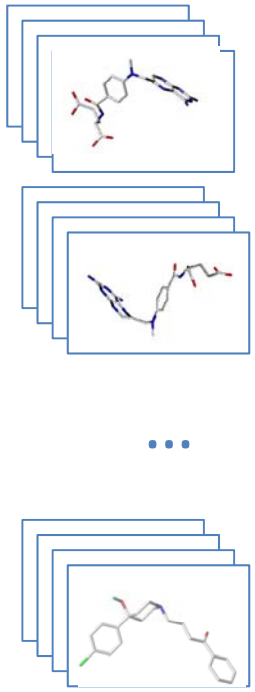
Compound profiling using similarity between protein binding sites and **shape analysis of ligand**

Ligand-centric method



query ligand

putative targets?



Ligands of protein 1

Ligands of protein 2

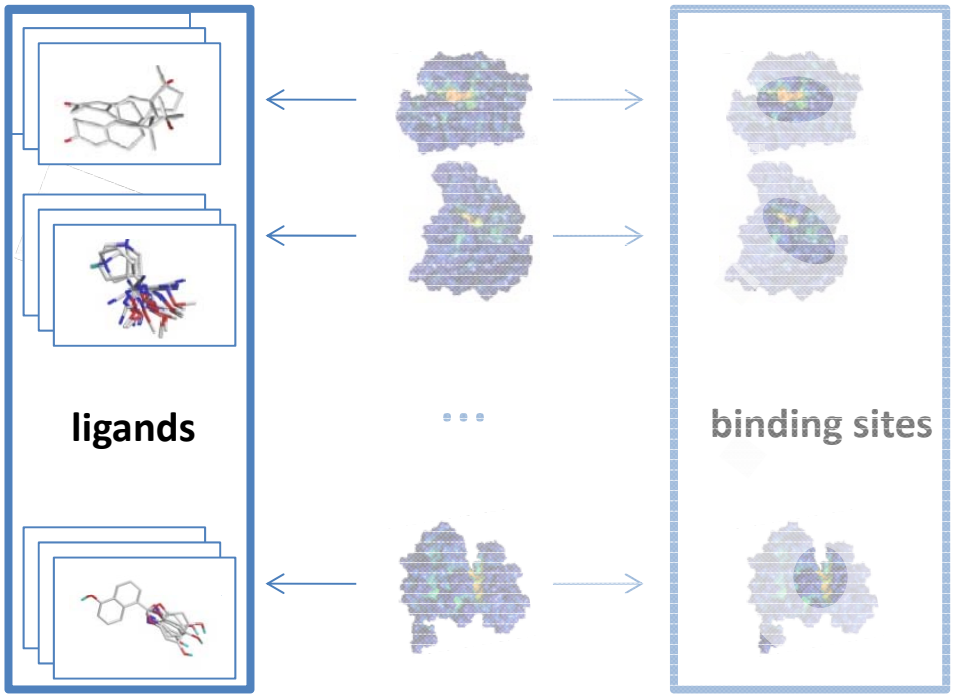
...

Ligands of protein n

List for exp. validation

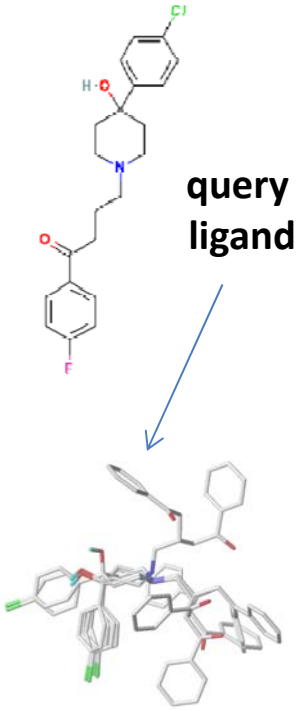
Compound profiling using similarity between protein binding sites and **shape analysis of ligand**

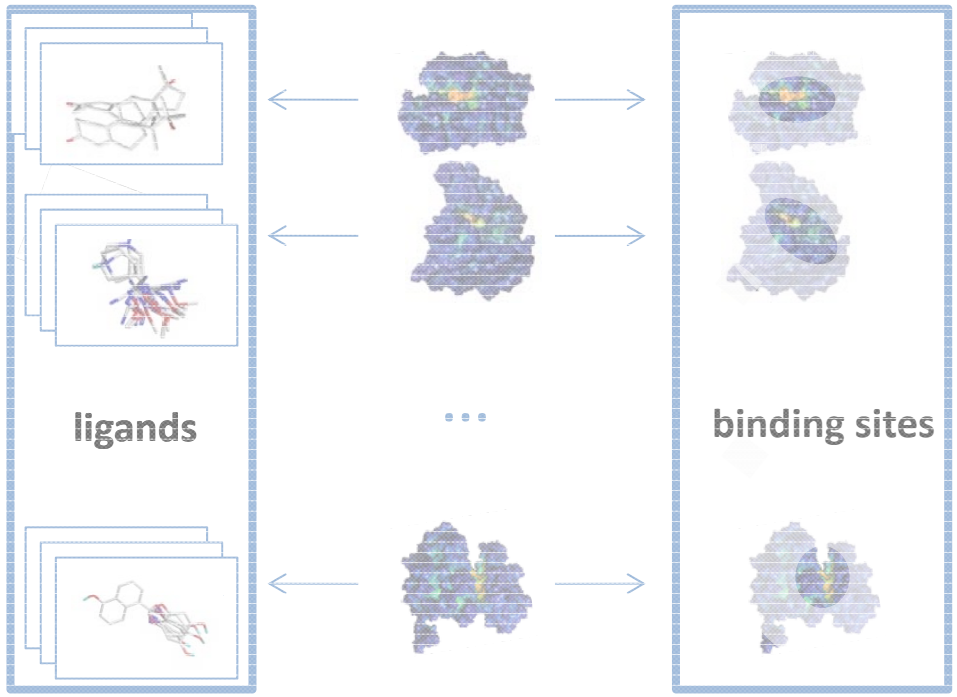
Searched database: different ligands per target, conformational ensemble



Compound profiling using similarity between protein binding sites and **shape analysis of ligand**

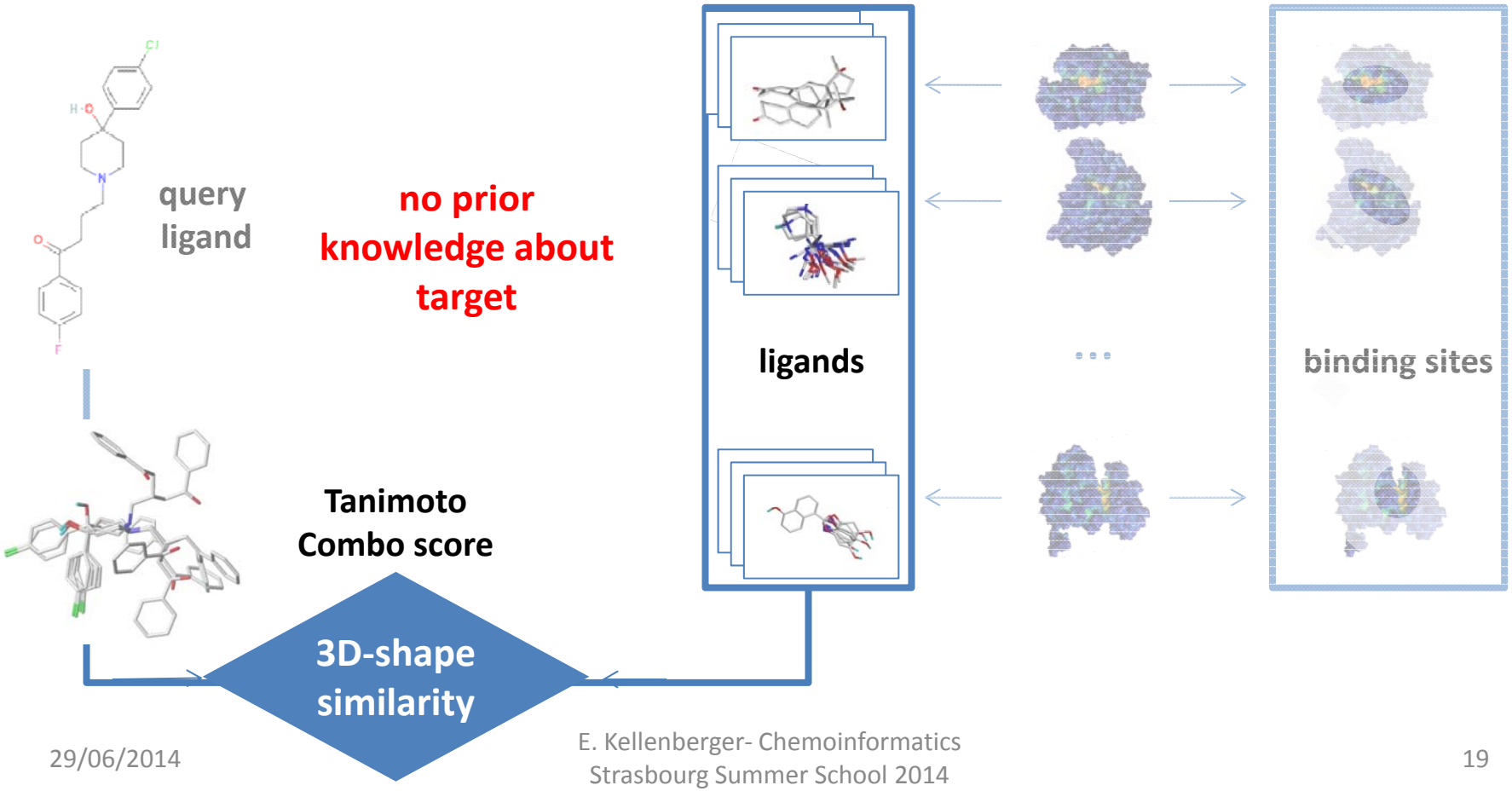
Query: conformational ensemble or selected conformer of the ligand





Compound profiling using similarity between protein binding sites and **shape analysis of ligand**

method: shape comparison using ROCS (OpenEye)



Tutorial : the test case

haloperidol

- antipsychotic drug, also used to control tics and vocal utterances that are part of Tourette's syndrome.
- approved by the FDA in 1967

Ship of Fools (XVe)
by Hieronymus Bosch

Tutorial : the test case

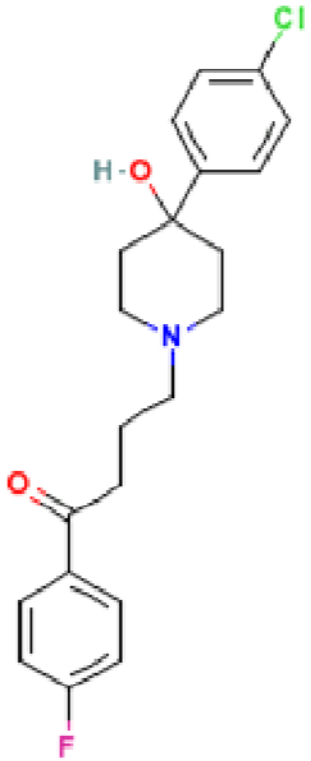
haloperidol

- interferes with the effects of neurotransmitters in the brain
- blocks **dopamine and serotonin receptors**

Primary targets

- Among common adverse effects:
  - constipation by competing with **acethylcholine**
  - Hypotension caused by the binding to **adrenergic receptors**

off targets



Tutorial : the test case

Dataset for protein-centric approach

query: human dopamine D3 receptor (3pbl)

Database:

15 proteins

27 sites

Functional class	Target Name	PDB
Class A GPCR	β1-adrenocept	2ycw, 2rh1
	Muscarinic M2 rec	3uon
	Chemokine receptor CCR5	4mbs
Class B GPCR	CRF1	4k5y
3-Ketosteroid receptor	Androgen receptor	2qpy, 3b5r
Estrogen receptor	ER beta	2fsz, 2j7x
heat shock protein	HSP90alpha	owd, 4efu
Carboxylic ester hydrolase	acethylcholinestera	zgc, 3i6m
Protein-serine/threonine kinase	CDCK2	1gij, 1w0x
	Aurora kinase	2np8, 2x81
Aspartic endopeptidase	Beta secretase	2fdp, 4djv
	renin	2g1o, 3vye
Serine endopeptidase	Thrombin	3rlw, 3sv2
Methyltransferase	Thymidylate synthase	1ci7, 3uwl
Carbon-oxygen lyase	Carbonic anhydrase	1a42, 3mhm

off targets

off target

Tutorial : the test case

Dataset for ligand-centric approach

query: haloperidol, low energy conformer

- 3D coordinates generated using corina (Molecular Network)
- 590 conformers generated using Omega (OpenEye)

Database: foreach of the15 proteins

10 different ligands

- from sc-PDB (ligand co-crystallized with the protein)
- from chEMBL (drug or ligand with IC_{50} <50 nM), 3D coordinates generated using corina

200 conformers generated using Omega (OpenEye)

Tutorial : demo 1

protein-centric approach: cavity detection

```
Desk21@dhcp-41-195:/tmp/CS3-3D$ Ichem volsite
REF/D3receptor-3pbl_protein.mol2 REF/D3receptor-
3pbl_ligand.mol2

Protein file : REF/D3receptor-3pbl_protein.mol2
Ligand file :  REF/D3receptor-3pbl_ligand.mol2
Assigning properties to protein atoms
Protein name : Unknown
Druggability : 1.17
Prediction :   Is Druggable

Desk21@dhcp-41-195:/tmp/CS3-3D$ ls
CAVITY_N1_12.mol2  CAVITY_N1_6.mol2  CAVITY_N1_ALL.mol2
CAVITY_N2_4.mol2   CAVITY_N2_8.mol2
CAVITY_N1_4.mol2   CAVITY_N1_8.mol2  CAVITY_N2_12.mol2
CAVITY_N2_6.mol2   CAVITY_N2_ALL.mol2 VolSite_stat.csv
```

Tutorial : demo 1

protein-centric approach: cavity comparison

```
Desk21@dhcp-41-195:/tmp/CS3-3D$ Shaper -r
REF/D3receptor-3pbl_cavity6.mol2 -c CAVITY/CCR5-
4mbs_cavity6.mol2 -o D3receptor-3pbl_CCR5-4mbs.pdb -rn
D3receptor -cn CCR5-4mbs
```

```
Reference file :  REF/D3receptor-3pbl_cavity6.mol2 ;
Num Confs : 1
Comparison file : CAVITY/CCR5-4mbs_cavity6.mol2
-------------------------------------------------------
-------------------------------------------------------
        Tanim   FitTve  RefTve
COLOR:  0.186   0.292   0.337
FIT:    0.497   0.634   0.497
Combo:  0.682   0.926   0.682
```

Tutorial : demo 1

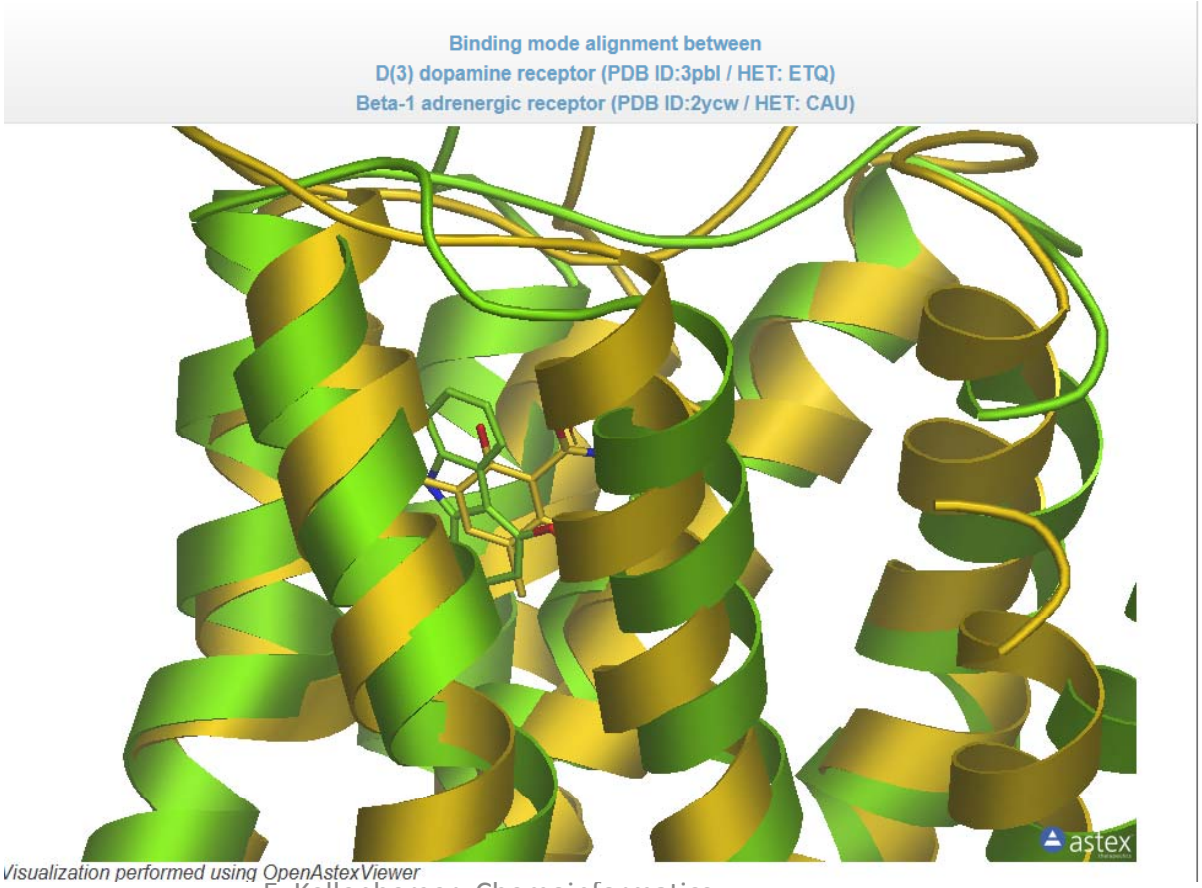
protein-centric approach: results

```
Desk1@dhcp-41-210:/tmp/CS3-3D$ sort -r -k5 Shape_res.csv
Ref Name        Comp Name                Simil   ColorTc ColorRefTvColorFitTv

D3      acethylcholinesterase-3i6m      -1      0.2487  0.4571  0.3529
D3      adrenoreceptor-2ycw             -1      0.3022  0.4466  0.4830
D3      adrenoreceptor-2rh1             -1      0.2434  0.4298  0.3595
D3      acethylcholinesterase-1zgc      -1      0.2026  0.4066  0.2877
D3      renin-3vye                      -1      0.1652  0.3732  0.2287
D3      renin-2g1o                      -1      0.1597  0.3728  0.2183
D3      betasecretase-2fdp              -1      0.1796  0.3634  0.2620
D3      betasecretase-4djv              -1      0.1842  0.3519  0.2788
D3      thrombin-3rlw                   -1      0.2125  0.3445  0.3567
D3      M2receptor-3uon                 -1      0.2408  0.3439  0.4453
/.../
D3      ERbeta-2j7x                     -1      0.1763  0.2252  0.4481
D3      aurorakinase-2x81               -1      0.1562  0.2165  0.3592
```

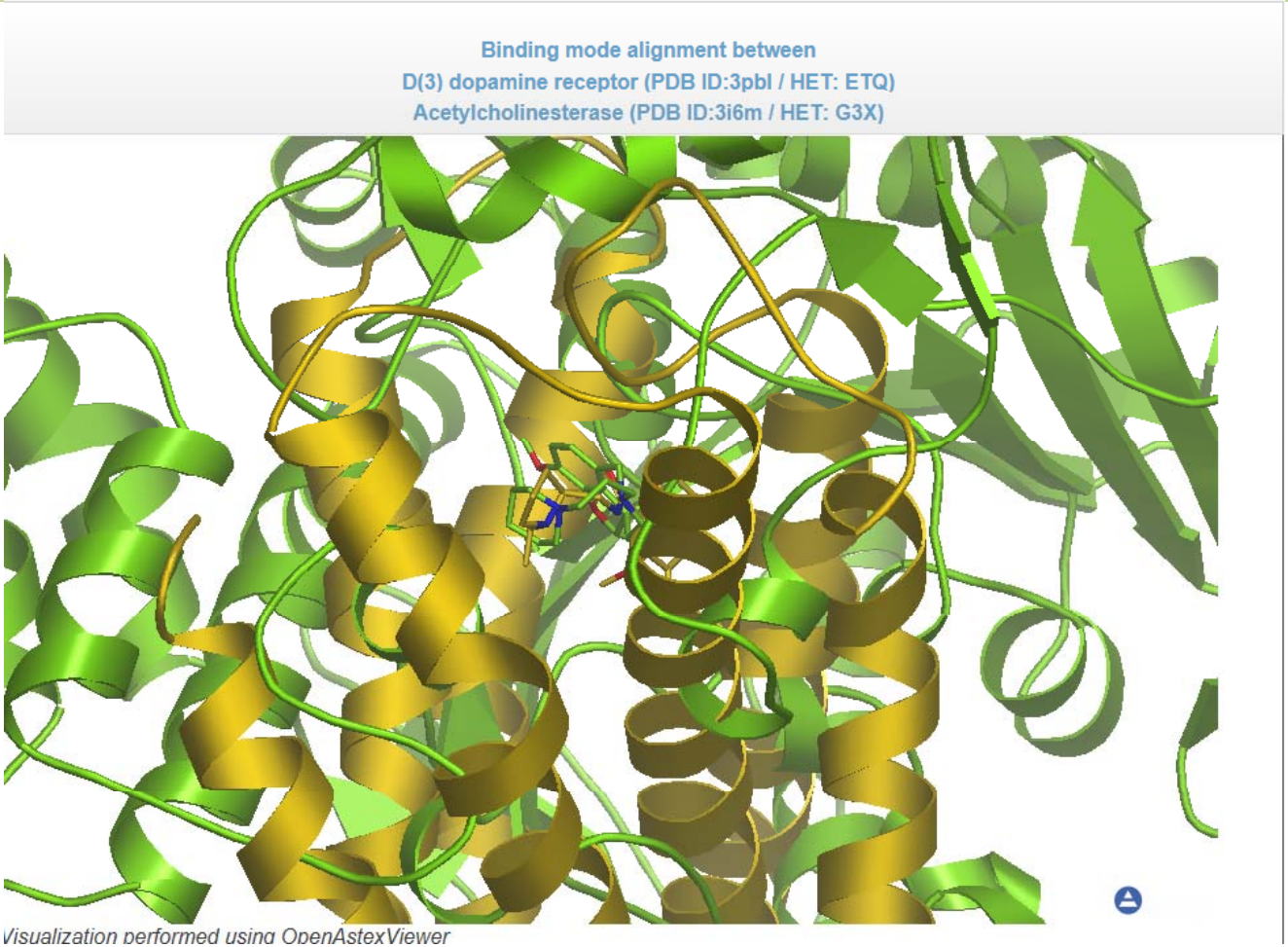
Tutorial : demo 1

protein-centric approach: common fold, conserved amino acids



Tutorial : demo 1

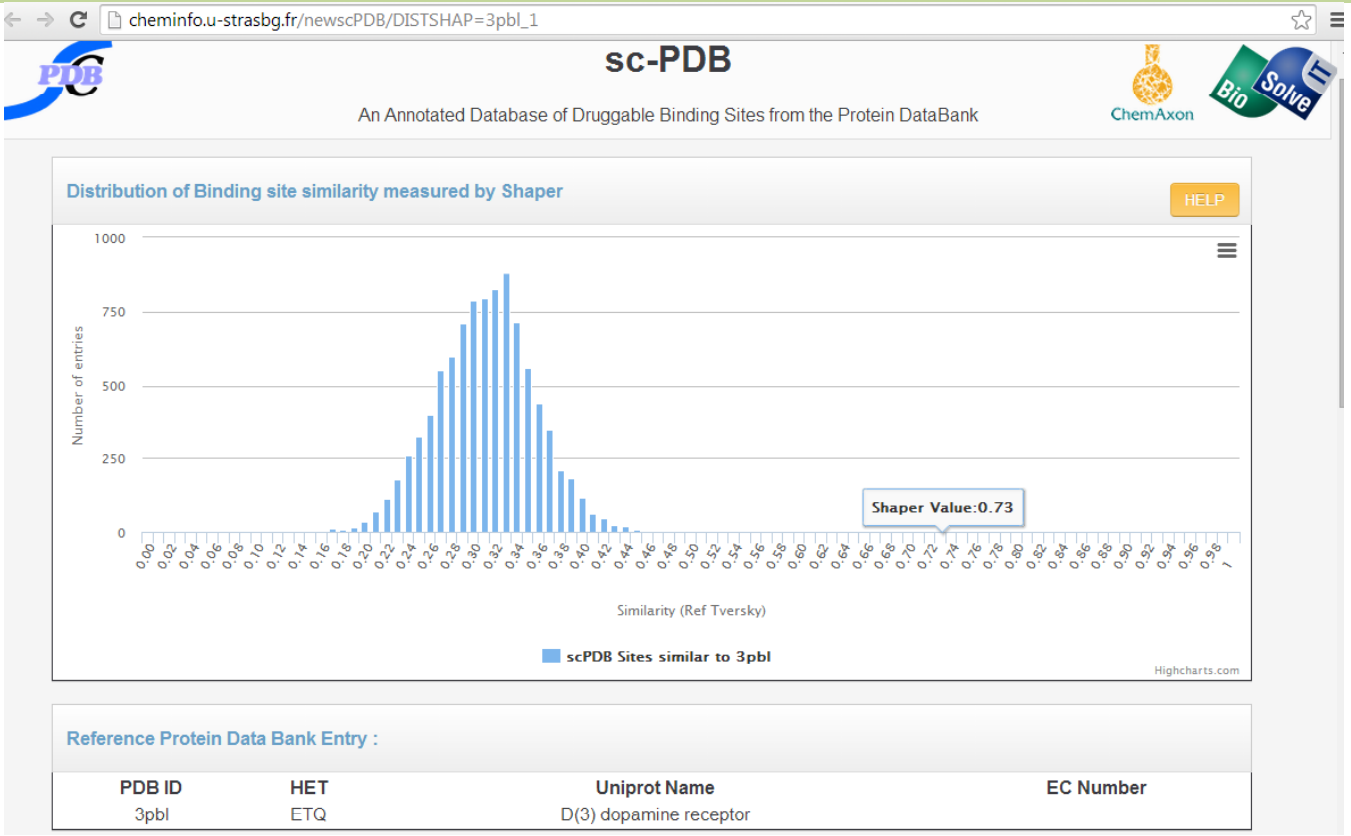
protein-centric approach: different fold, conserved amino acids



Tutorial : demo 1

protein-centric approach: searching the PDB

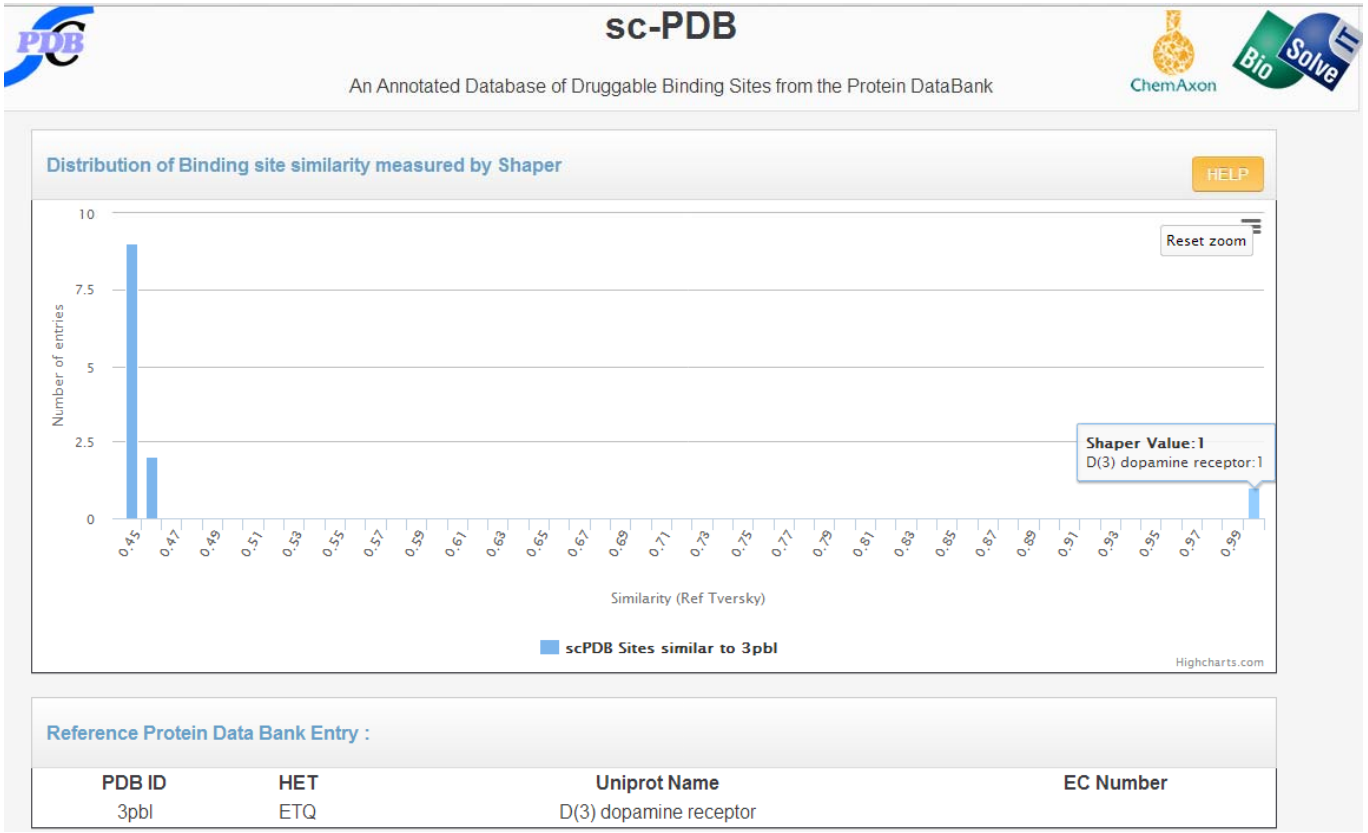
3pbl
vs
sc-PDB
(3 678 proteins,
9 283 PDB entries)



Tutorial : demo 1

protein-centric approach

3pbl
vs
sc-PDB
(3 678 proteins,
9 283 PDB entries)



Tutorial : demo 1

protein-centric approach

3pbl
vs
sc-PDB
(3 678 proteins,
9 283 PDB entries)

cheminfo.u-strasbg.fr/newscPDB/DISTSHAP=3pbl_1

how 25 entries Search:

PDB ID	HET	Uniprot Name	EC Number	Binding Site Similarity	Align
3pbl	ETQ	D(3) dopamine receptor	/	1.000	ALIGN
1odm	ASV	Isopenicillin N synthase	1.21.3.1	0.460	ALIGN
3i6m	G3X	Acetylcholinesterase	3.1.1.7	0.460	ALIGN
1w3x	W2X	Isopenicillin N synthase	1.21.3.1	0.450	ALIGN
1x8j	AE2	Retinol dehydratase	/	0.450	ALIGN
2isc	223	Purine nucleoside phosphorylase, putative	/	0.450	ALIGN
2ycw	CAU	Beta-1 adrenergic receptor	/	0.450	ALIGN
1h22	E10	Acetylcholinesterase	3.1.1.7	0.450	ALIGN
3ju8	NAD	N-succinylglutamate 5-semialdehyde dehydrogenase	1.2.1.71	0.450	ALIGN
2bu9	HFV	Isopenicillin N synthase	1.21.3.1	0.450	ALIGN
2q3j	H02	Ferrochelatase	4.99.1.1	0.450	ALIGN
4jlj	1NM	Deoxycytidine kinase	2.7.1.74	0.450	ALIGN

Tutorial : demo 2

Ligand-centric approach: rocs

```
Desk21@dhcp-41-195:/tmp/CS3-3D$ rocs -query
REF/haloperidol.mol2 -dbase LIGAND/CCR5-
4mbs_ligandmulticonf.mol2 -prefix CCR5-4mbs -oformat
mol2 -maxhits 1
                :
/.../
ROCS 3.2.0.4: OpenEye Scientific Software, Santa Fe, NM.
/.../
Query being read from:
        REF/haloperidol.mol2
File prefix is:                         CCR5-4mbs
/.../
Log file will be written to:            CCR5-4mbs.log
Statistics will be written to           CCR5-4mbs_1.rpt
Hit structures will written to:         CCR5-4mbs_hits_1.mol2
Status file will be written to:         CCR5-4mbs 1.status
```

Tutorial : demo 2

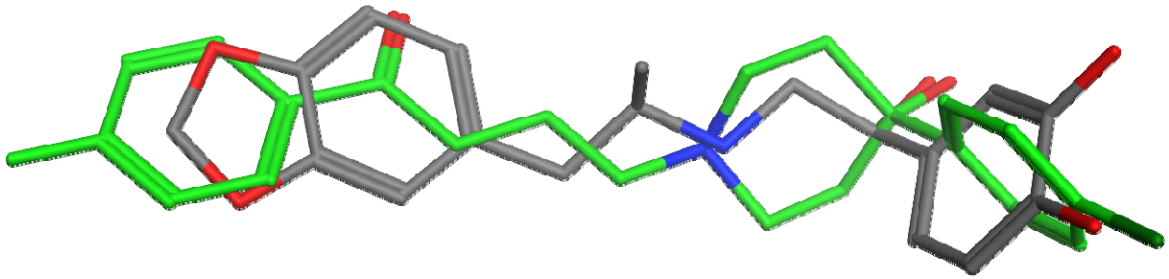
Ligand-centric approach: rocs

```
Desk21@dhcp-41-195:/tmp/CS3-3D$ rocs -query
REF/haloperidol.mol2 -dbase LIGAND/CCR5-
4mbs_ligandmulticonf.mol2 -prefix CCR5-4mbs -oformat
mol2 -maxhits 1
/.../
Query(#1) molecule written to    CCR5-4mbs_hits_1.mol2
Query(#1): haloperidol_1 has 1 conformer(s)
Database 1 of 1:LIGAND/CCR5-4mbs_ligandmulticonf.mol2
1 molecules in 1 seconds -> 1.0 molecules/sec
                                200 overlays/sec
1 hits found
================================================
                  writing
hits|**************************************|100.00%
Molecule read failures: 0
#warnings              : 0
#errors                : 0
#queries processed     : 1
```

Tutorial : demo 2

Ligand-centric approach: virtual screening using rocs query with single conformer

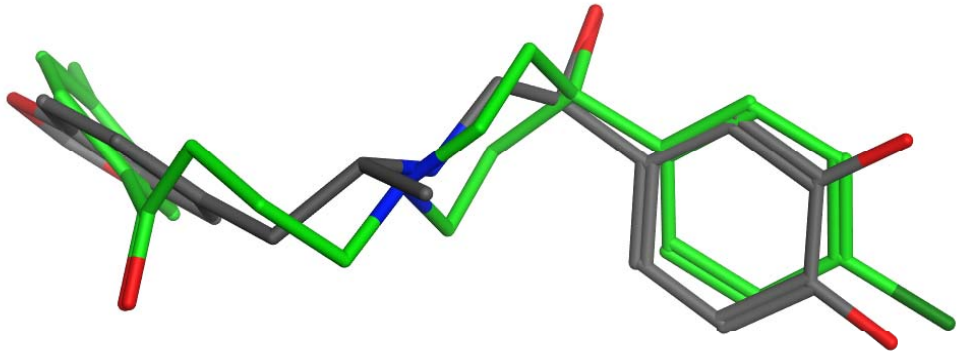
```
Desk21@dhcp-41-210:/tmp/CS3-3D$ sort -r  -k5 rocs_res.csv
File                                ShapeQuery      Rank     TcCombo
adrenoreceptor-protokylol           haloperidol_1   1        1.033
acethylcholinesterase-1eve          haloperidol_1   1        1.005
CCR5-CHEMBL392659                   haloperidol 1   1        0.951
/..../
HSP90alpha-1yet                     haloperidol_1   1        0.465
```



Tutorial : demo 2

Ligand-centric approach: virtual screening using rocs query with multiple conformers

```
Desk21@dhcp-41-210:/tmp/CS3-3D$ sort -r  -k5 rocs_res.csv
File                              ShapeQuery              Rank    TcCombo
CCR5-CHEMBL392659                 haloperidol_475         1         1.184
adrenoreceptor-protokylol         haloperidol_572         1         1.126
CCR5-CHEMBL2178576                haloperidol_56          1         1.096
CCR5-CHEMBL322251                 haloperidol_565         1         1.067
renin-3vye                        haloperidol_535         1         1.064
acethylcholinesterase-1eve        haloperidol_241         1         1.063
```



conclusion

3D methods are suitable to the fast profiling of haloperidol

- Protein- and ligand-centric methods identified the same two targets, which are likely true positives
- Protein-centric approach requires the 3D-stucture of the primary target and of tested proteins
- Ligand-centric approach requires for each of the tested protein several ligands (ideally of different chemotypes)