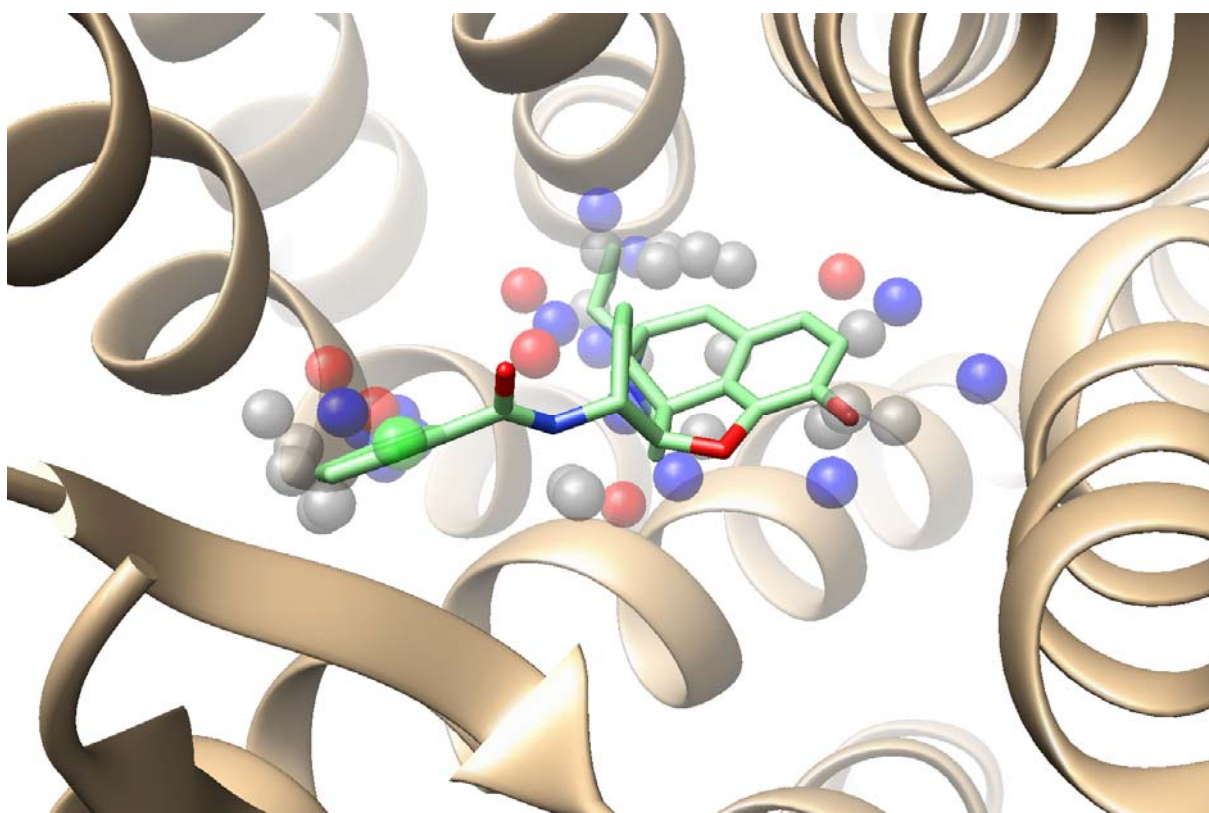


Shaper: Shape-based alignment of molecular objects



Jérémy DESAPHY, Franck DA SILVA, Guillaume BRET and Didier ROGNAN

Laboratoire d'Innovation Thérapeutique, UMR 7200 CNRS-Université de Strasbourg, F-67400 ILLKIRCH

Email: rognan@unistra.fr



Literature Corner

Please have a look at these articles for detailed information on the basic principles and concepts underlying Shaper usage

Desaphy, J., Azdimousa, K., and Rognan, D. (2012) Comparison and druggability prediction of protein-ligand binding pockets from pharmacophore-annotated shapes., *J. Chem. Inf. Model.*, **52**, 2287-2299



Tran-Nguyen, V.K., Da Silva, F., Bret, G. and Rognan, D. (2019) All in One: Cavity Detection, Druggability Estimate, Cavity-Based Pharmacophore Perception and Virtual Screening. *J. Chem. Inf. Model.*, doi: 10.1021/acs.jcim.8b00684



What's new ?

Shaper v.2 enables the alignment of ligand atoms to structure-based pharmacophores generated in IChem.

For more details about the procedure, please have a look at:

Tran-Nguyen VK, Da Silva F, Bret G, Rognan D.

J Chem Inf Model. 2018 Dec 18. doi: 10.1021/acs.jcim.8b00684.

[All in One: Cavity Detection, Druggability Estimate, Cavity-Based Pharmacophore Perception and Virtual Screening.](https://doi.org/10.1021/acs.jcim.8b00684)



Installation

Shaper is provided as a zipped archive file (**Shaper.tgz**) containing the following material:



- **test**: a directory containing some test input/output files
- **CavCav.cff**: force-field to align cavities (Shaper)
- **CavLig14.cff** : force-field for aligning ligands to pharmacophores (Shaper2)
- **RotaMole**: executable to rotate/translate a molecule
- **Shaper**: C++ executable to align cavities
- **Shaper2**: python script to align ligands to pharmacophores
- **User_Guide.pdf**: this manual



Please note that Shaper/Shaper2 requires a valid license for OEChem and Shape toolkits from OpenEye Scientific Software (<https://www.eyesopen.com>).

Test examples shown here have been run with the redhat-RHEL7-g++4.8-x64 distribution

Note to users:

Shaper commands with options/tool/arguments will be displayed in *italic* characters after the ">" prompt

Input/output filenames will be displayed in **bold violet** characters

Terminal output will be displayed with a **gray background**

A copy of all test input/output files is given in the test directory of the IChem distribution. Before using IChem, please take the time to read the description of the technology. Articles to read will be mentioned by the following icon:





Aligning cavities with Shaper

Shaper is a C++ routine enabling the alignment of cavities generated with IChem (see page 25 of the IChem v6.0 user guide)

IChem VolSite is a structure-based tool to automatically detect cavities at the surface of a target protein, and predict their ligandability (structural druggability)

For more information, see:



Desaphy J, Azdimousa K, Kellenberger E, Rognan D.
J Chem Inf Model. 2012 Aug 27;52(8):2287-99.

[Comparison and druggability prediction of protein-ligand binding sites from pharmacophore-annotated cavity shapes.](#)

The correct syntax for using Shaper is:

> Shaper -r refcav -c cavity [-f cffFile] [-o output] [--Debug] [-rn refcav_name] [-cn compcav_name]

Flags :

-r reference: Cavity6 file for better results. Used as the reference cavity. Allowed file format: same as ligand file

-c cavity: Cavity6 file for better results. Used as the comparison cavity. Allowed file format: same as ligand file

-f cffFile: ColorForceField file. If not defined, will automatically used the default one (see below)

-o output: Output file: The alignment of the two compared molecules will be written in the file. Allowed file format: same as ligand file. Preferred file format : sdf or mol2

--Debug: Activate the verbose mode

-rn: Name of the reference cavity. Useful only for the results.csv file

-cn: Name of the comparison ligand or cavity. Useful only for the results.csv file

In this tutorial, we will align two PDB entries (**2rh1**: human beta-2 adrenergic receptor in complex with carazolol; **2ycx**: turkey beta-1 adrenergic receptor in complex with cyanopindolol) by matching their ligand-binding cavities.

Atomic coordinates (MOL2 files) for proteins, ligands and cavities have been downloaded from the scPDB website (<http://bioinfo-pharma.u-strasbg.fr/scPDB/>)

Alternatively, you can generate your own cavities points with the IChem Volsite command (see IChem user guide) as follows:

```
> IChem volsite 2rh1_protein.mol2 2rh1_ligand.mol2  
> mv CAVITY_N1_6.mol2 2rh1_cavity.mol2
```

The two complexes are in different coordinate frames (**Fig.1**) and not aligned when considering the raw PDB atomic coordinates

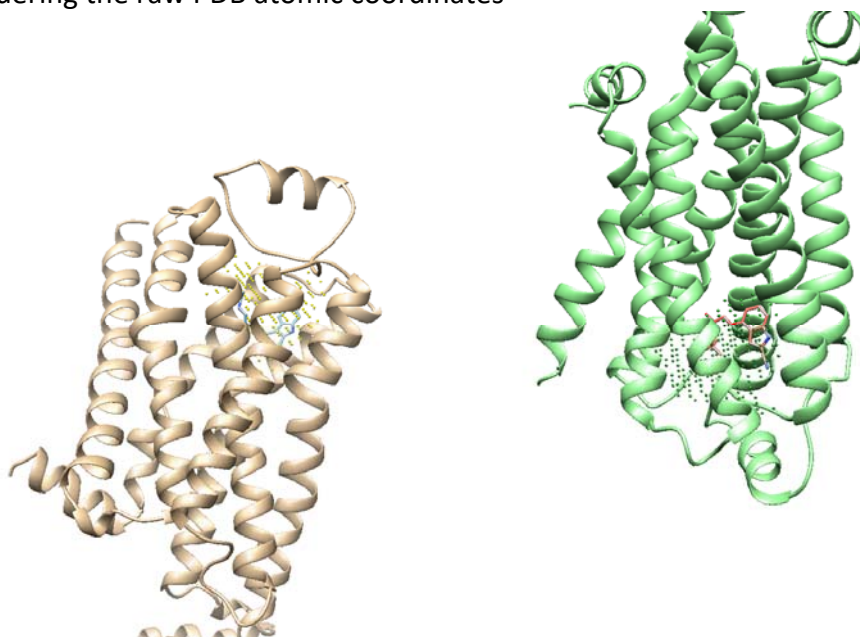


Fig.1. input PDB coordinates of 2rh1 and 2ycx entries (2rh1 protein: tan ribbons, 2rh1 ligand: cyan sticks, 2rh1 cavity: yellow dots, 2ycx protein: green ribbons, 2ycx ligand: orange sticks, 2ycx cavity: green dots)

1. Alignment of 1dah and 1a82 cavities

Input directory: `./test/cav-cav`

Input files:

- `2rh1_cavity.mol2` (2rh1 cavity coordinates)
- `2rh1_ligand.mol2` (2rh1 ligand coordinates)
- `2rh1_protein.mol2` (2rh1 protein coordinates)
- `2ycx_cavity.mol2` (2ycx cavity coordinates)
- `2ycx_ligand.mol2` (2ycx ligand coordinates)
- `2ycx_protein.mol2` (2ycx protein coordinates)
- `do.sh` (shell script to perform all operations)

Output files:

- `rot_2ycx_cavity.mol2` (2ycx cavity coordinates aligned that of 2rh1 cavity)
- `Shape_res.csv` (Shaper output)

Command:

```
> ../../Shaper -rn 2rh1 -cn 2ycx -r 2rh1_cavity.mol2 -c 2ycx_cavity.mol2 -o rot_2ycx_cavity.mol2
```

6 similarity values are outputted at the terminal

```
Reference file: 2rh1_cavity.mol2 ; Num Confs : 1
Comparison file: 2ycx_cavity.mol2
```

```
-----
          Tanim    FitTve    RefTve
COLOR:    0.386    0.543    0.571
FIT:      0.839    0.847    0.839
Combo:    1.23     1.39     1.23
```

COLOR: alignment of pharmacophoric features

FIT: alignment of shapes

Combo: COLOR + FIT

Tanim: Tanimoto similarity metric

FitTve: Fit Tversky similarity metric (normalized score with respect to the fit object)

RefTve: ref Tversky similarity metric (normalized score with respect to the reference object)

The Shape_res_csv output file contains a summary of the similarity scores focusing on the alignment of color (pharmacophoric) feature

```
RefName  Comp  Name  Simil  ColorTc  ColorRefTversky  ColorFitTversky
2rh1     2ycx   -1     0.3857  0.5707      0.5434
```

For details about the scores, please see Desaphy et al. (2012) Comparison and druggability prediction of protein-ligand binding pockets from pharmacophore-annotated shapes., *J. Chem. Inf. Model.*, **52**, 2287-2299

Some notes:

- Please use mol2 input files only
- **We recommend the usage of the ColorRefTversky similarity metric to distinguish similar from dissimilar cavities (threshold score > 0.45)**
- A default force-field (CavCav.cff), specifically validated for aligning protein cavities is used. Any user-tailored similar force field can be used although we do not recommend it.

2. Cavity-based alignment of 1dah and 1a82 protein-ligand coordinates

The above-described alignment has just generated cavity points of the 2ycx entries fitted to that of the 2rh1 reference. This output file can be used to reconstruct the rotation/translation matrix necessary to align protein-ligand coordinates accordingly.

Input directory: `./test/cav-cav/`

Input files: `2ycx_cavity.mol2` (2ycx cavity coordinates)
`2ycx_ligand.mol2` (2ycx ligand coordinates)
`2ycx_protein.mol2` (2ycx protein coordinates)
`rot_2ycx_cavity.mol2` (aligned 2ycx cavity points from previous step)

Output files: `rot_2ycx_ligand.mol2` (rotated coordinates of the 2rh1 ligand)
`rot_2ycx_protein.mol2` (aligned coordinates of the 2rh1 protein)

Command:

```
> ../../RotaMole -r rot_2ycx_cavity.mol2 -c 2ycx_cavity.mol2 -a 2ycx_protein.mol2  
2ycx_ligand.mol2
```

The RotaMole executable deduces the rotation/translation matrix to go from the initial to the aligned coordinates, and applies it to any molecule (-a option). The terminal outputs the x,y,z coordinates of the barycenter of the comparison and reference cavities and the root-mean square deviations of the barycenter after rotation/translation.

```
Reference Molecule barycenter: -31.1508 6.2931 8.33534  
Comparison Molecule barycenter: 27.3939 0.903644 -11.3826  
RMSD : 4.93792e-05
```

Atomic coordinates of the 2ycx entry are now aligned to that of the reference complex (2rh1) using the simple cavity-biased alignment procedure (**Fig. 2**).

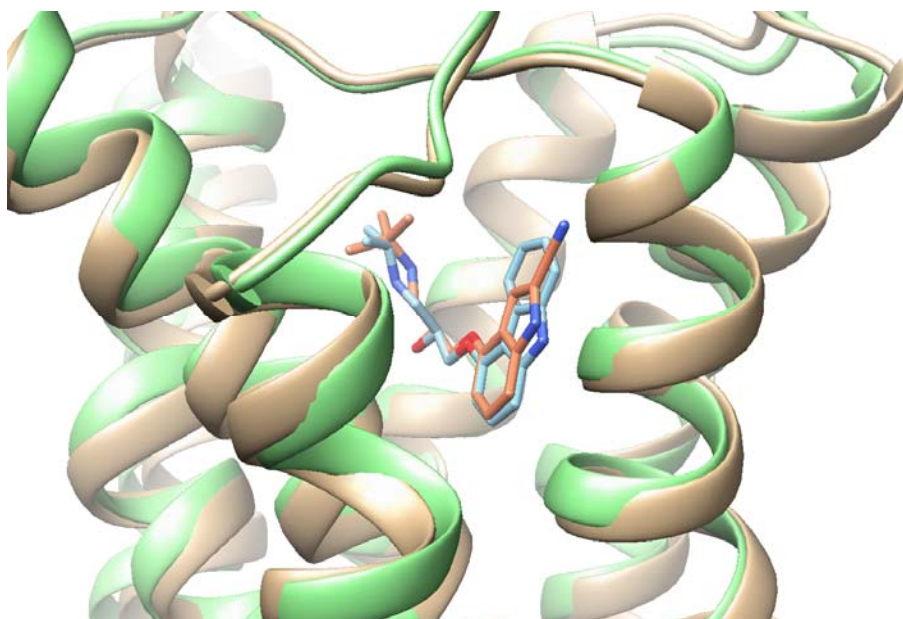


Fig. 2. Cavity-based alignment of 2rh1 and 2ycx PDB entries (2rh1 protein: tan ribbons, 2rh1 ligand: cyan sticks, 2ycx protein: green ribbons, 2ycx ligand: orange sticks)



Aligning ligands to pharmacophore with Shaper2

Shaper2 is a python script enabling the alignment of ligands to structure-based pharmacophores generated with IChem (see page 29 of the IChem v6.0 user guide)

For more information, see:



Tran-Nguyen VK, Da Silva F, Bret G, Rognan D.

J Chem Inf Model. 2018 Dec 18. doi: 10.1021/acs.jcim.8b00684.

[All in One: Cavity Detection, Druggability Estimate, Cavity-Based Pharmacophore Perception and Virtual Screening.](#)

The correct syntax for using Shaper2 is:

```
> Shaper2.py -r REF -c FIT -o OUTPUT -m OVERLAP -i START -s METRIC -a MAXSTEPS -n KEEPNB -v -u RADIUS (--version)
```

optional arguments:

```
-h or --help      show this help message and exit
-r or --ref       Path to reference <cavity.mol2> file
-c or --fit       Path to (multimol) <ligand.mol2> file
-o, --output      Path to overlap results <output.mol2> file
-t or --tsv       Path to results score <score.tsv> file
-f or --cff       Path to color force field file <FILE.cff>
-m or --method    Algorithm to calculate overlap (default="GridShapeGridColor")
{GridShape,GridColor,GridShapeGridColor,GridShapeAnalyticColor,GridShapeExactColor,AnalyticShape,Analytic
Color,AnalyticShapeAnalyticColor,AnalyticShapeGridColor,AnalyticShapeExactColor,ExactShape,ExactColor,Exac
tShapeExactColor,ExactShapeGridColor,ExactShapeAnalyticColor}
-i or --ini       Starting point of the optimization (default="Inertial")
{Inertial,AsIs,Random,InertialAtHeavyAtoms,Subrocs},
-s or --score     Scoring function for sorting overlays (default="TanimotoCombo")
{TanimotoCombo,Tanimoto,ColorTanimoto,ScaledColor,ComboScore,FitColorTversky,FitTversky,FitTverskyCombo,
Overlap,RefColorTversky,RefTversky,RefTverskyCombo},
-a or --maxstep   Set the maximum number of optimization iteration steps (default=200).
-n or --keepnb    Number of solutions to keep (default=1)
-v or --verbose   verbose (default=False)
-u or --radius    Set the radius for atom overlap (default=1.7). Only valid for "ExactShape"
--version         show program's version number and exit
```

For more details about all options, please look at the ShapeTK manual at <https://docs.eyesopen.com/toolkits/python/shapetk/index.html>

In the following tutorial, we will align 4-hydroxytamoxifen to a cavity-based pharmacophore generated from the structure of the human estrogen receptor α (PDB identifier 3ert). In a first example, a single conformer will be aligned. In the second example, multiple conformers will be aligned and scored with respect to two different strategies.

The cavity-based pharmacophore was generated with IChem VolSite (page 29 of the IChem v.6 user guide)

```
> IChem --pharm -pCA 0 -pCZ 1.56 -pO 0 -pN 0 -pOD1 0 -pNZ 0 -pZn 3.49 volsite
3ert_protein.mol2
```

1. Alignment of a single ligand conformer

Input directory: `./test/lig-cav/`

Input files: `3ert_ligand.mol2` (randomized orientation of ligand coordinates)
`3ert_ligand_Xtal.mol2` (X-ray pose of the ligand)
`3ert_pharmacophore.mol2` (cavity-based pharmacophore)
`3ert_protein.mol2` (3ert protein coordinates)
`do.sh` (shell script to perform the tutorial)

Output files: `aligned_3ert_ligand.mol2` (aligned coordinates of the 3ert ligand)
`aligned_3ert_ligand.tsv` (output of the alignment)

Command:

```
> ../../Shaper2.py -r 3ert_pharmacophore.mol2 -c 3ert_ligand.mol2 -o aligned_3ert_ligand.mol2 -t
aligned_3ert_ligand.tsv -f ../../CavLig14.cff -m ExactShapeExactColor -i Subrocs -s TanimotoCombo
-n 1
```

Two output files are generated: `aligned_3ert_ligand.mol2` (ligand aligned to the pharmacophore), `aligned_3ert_ligand.tsv` (alignment output scores).

The aligned pose exhibits a rmsd to the x-ray pose of 1.0 Å for a score (TanimotoCombo) of 0.437

2. Alignment of multiple conformers

Input directory: `./test/lig-cav/`

Input files: `3ert_ligand_confs.mol2` (104 OMEGA conformers of the ligand)
`3ert_ligand_Xtal.mol2` (X-ray pose of the ligand)
`3ert_pharmacophore.mol2` (cavity-based pharmacophore)
`3ert_protein.mol2` (3ert protein coordinates)
`do.sh` (shell script to perform the tutorial)

Output files: `aligned_3ert_ligand_confs.mol2` (aligned coordinates of the 3ert ligand)
`aligned_3ert_ligand_confs.tsv` (output of the alignment)

Command:

```
> ../../Shaper2.py -r 3ert_pharmacophore.mol2 -c 3ert_ligand_confs.mol2 -o
aligned_3ert_ligand_confs.mol2 -t aligned_3ert_ligand_confs.tsv -f ../../CavLig14.cff -m
ExactShapeExactColor -i Subrocs -s TanimotoCombo -n 10
```

Two output files are generated: aligned_3ert_ligand.mol2 (best 10 poses of the ligand aligned to the pharmacophore), aligned_3ert_ligand.tsv (alignment output scores).
More output poses can be generated by varying the $-n$ parameter (maximal value = number of input conformers)

The alignment scores (TanimotoCombo values) along with rmsd to the X-ray pose for the best 10 poses are :

Fit	TanCombo	rmsd, Å
3ert_OHT_1_ligand_73	0.511606	1.73
3ert_OHT_1_ligand_64	0.489066	1.88
3ert_OHT_1_ligand_48	0.486754	2.00
3ert_OHT_1_ligand_56	0.484928	2.49
3ert_OHT_1_ligand_102	0.470631	1.91
3ert_OHT_1_ligand_41	0.468782	5.11
3ert_OHT_1_ligand_75	0.467568	5.17
3ert_OHT_1_ligand_94	0.466532	2.04
3ert_OHT_1_ligand_76	0.466481	2.16
3ert_OHT_1_ligand_33	0.462975	1.33



Each pose can be rescored according to your favorite scoring function. For example, we minimized all protein-ligand coordinates with the MMFF94 force-field with Poisson-Boltzmann implicit treatment of electrostatics effects (available in OpenEye's Szybki tool) as described in Tran-Nguyen et al. JCIM, 2019. MMFF94 rescoring reduces the rmsd of the best pose (conformer 73) from 1.73 to 1.11 Å (**Fig.3**)

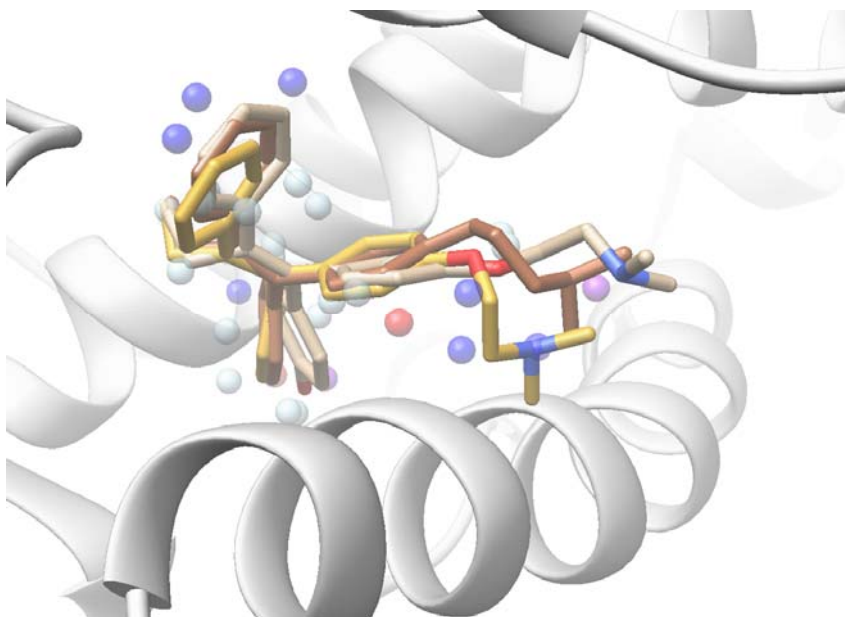


Fig.3. Alignment of 4-hydroxytamoxifen to the cavity-based pharmacophore (transparent spheres) of the human estrogen receptor α subtype (white ribbons). X-ray pose, tan; Shaper2 ranked-1 pose, gold; MMFF94 refined Shaper ranked-1 pose, sienna. Pharmacophoric features are color-coded as follows: hydrophobic, cyan; hydrogen-bond donor, blue; positive ionisable, purple; hydrogen-bond acceptor, red; negative ionisable, dark red.