

ShaEP

Notes on Usage

Matthew Holland
Brennan Group
Target Discovery Institute, University of Oxford
`matthew.holland@spc.ox.ac.uk`

27 July 2021

1 Introduction

ShaEP is a software tool developed by Dr Mikko Vainio and co-workers at Åbo Akademi University in Finland in 2009.[1] It aligns pairs of rigid, 3D molecules based on their shape and electrostatic potential, and returns a similarity index for each pair of molecules - one for the shape similarity, and one for the electrostatic potential similarity. It can be used to compare two molecules, or to screen a query molecule against a library of molecules.

2 Installation

ShaEP is downloaded as a binary file from <http://users.abo.fi/mivainio/shaep/download.php>. To download, read the license agreement, select your operating system and press 'Download'. A .zip file will download into your downloads folder. Once the download has complete, click on the downloaded folder to unzip it and an executable file called 'ShaEP' should appear in the downloads folder. I recommend moving it from the downloads folder to the applications folder, either by drag and drop, or through the command line.

To make life easier, it is then worth adding the path to the program to your bash profile. To do this, open up a terminal window, and type the following:

```
vi ~/.bash_profile
```

Then navigate to the bottom of the window, and press the 'i' key (standing for insert). On a new line, type the following:

```
export PATH="/Applications:$PATH"
```

and then press the 'esc' key, and type ':wq', (standing for write, then quit) then press enter. You should be back in the terminal window, into which you should now type

```
source ~/.bash_profile
```

and press enter.

3 Usage

ShaEP takes as its input a file containing the coordinates of all the atoms of the query molecule, and a file containing the coordinates of the atoms of all of the molecules in the target library to be searched. It is possible only to compare the shapes of the molecules, without comparing their electrostatic potentials (ESPs), and for this only atomic coordinates are needed, not partial charges. The following file formats are accepted by ShaEP:

- .sdf
- .xyz

However, to compare and align molecules by electrostatic potential, the partial charge of each atom is needed, along with its coordinates. Currently the only accepted file format that contains all of this information is the TRIPOS Mol2 format. Software to generate these files is hard to come by, however I have written a tool (malt) for generating them, which is available for free on my GitHub.

Make sure the files for the query molecule and the target molecule/library are stored in the same directory (folder on your computer), open a terminal and navigate to the directory where the files are stored as below:

```
cd path/to/directory
```

To compare a mol2 file with a query molecule, containing atomic coordinates and atomic partial charges (query.mol2) to a target library file (target_library.mol2), type the following into the terminal:

```
shaep --maxhits 100 -q query.mol2 --output-file similarity.txt  
target_library.mol2
```

These instructions will compare the molecule in query to each molecule in target_library, printing the top 100 molecules' similarity scores to the file similarity.txt, which will be saved into the same directory as the input files. It will also return a file called similarity_hits.txt, which contains all of the hits identified in similarity.txt, ranked in order of highest similarity, with the least similar molecule at the top, and the most similar molecule at the bottom of the list.

If you want to return an sdf file, which contains the co-ordinates of the aligned molecules, relative to those of the target molecule, then you need to include the following additional commands:

```
shaep --maxhits 5 -q query.mol2 --output-file similarity.txt --structures  
overlay.sdf --outputQuery target_library.mol2
```

This performs the exact same operation as the previous example, but also returns an sdf file of the atomic coordinates of the target molecules aligned to the query molecule. This is useful for visualising and comparing the alignments of the target molecules to the query.

References

- [1] Vainio, M., J.; Puranen, J., S.; Johnson, M., S., ShaEP: Molecular Overlay Based on Shape and Electrostatic Potential. *J. Chem. Inf. Model*, **2009**, *49*, 492-502

A Commands and Options

A.1 Accepted Input Files

- .sdf (no partial charges)
- .sdf v3 from Accelrys (no partial charges)
- Tripos Mol2 format (coordinates and partial charges)
- smi SMILES format (no coordinates or partial charges)
- log Gaussian98/03 log file (no coordinates or partial charges)
- pdb (no partial charges)
- vbf Visipoint binary format
- xyz (no partial charges)

A.2 Command Options

<code>-q [--query] arg</code>	File name for a structure with which each input structure is superimposed. Multiple query structures can be used. If not given, the first file in the input list is used.
<code>--output-file arg</code>	Name of the output file for the similarity indices. If not given, the last filename in the input list is taken as the output file name.
<code>--maxhits arg</code>	The maximum number of compounds taken in the hitlist. Zero imposes no limit. Default = 0.
<code>-s [--structures] arg</code>	Write the superimposed structures in the hitlist to a file in sdf format. arg is filename.
<code>--minHitSimilarity arg</code>	The minimum similarity index value required for a compound to make it to the hitlist. Value must be within range [0, 1]. Zero imposes no limit, and is the default value.
<code>--outputQuery</code>	Write the matched query structure to the structures file before the superimposed structures, effective only with the 'structures' keyword. Exclusion volume structures are not written out.