Filetypes

SMSD matches various filetypes, to see a list of supported types, run SMSD with "-h". At the end of the help, the list of types is shown, along with a description for each type. The type is a short identifier ('MOL', 'PDB', etc) that is used to tell SMSD what to expect. The query and the target files can have different types, for example:

./SMSD -Q MOL -q molfile.mol -T PDB -t pdbfile.pdb

Where the uppercase flags (-Q and -T) give the types, and the lowercase flags (-q and -t) give the filenames. For 'string' types - such as SMILEs - the filename will be the data itself:

./SMSD -Q SMI -q "CCC" -T PDB -t pdbfile.pdb

Note that, while the quotes may not always be necessary, they will prevent problems with more complex SMILEs.

Types can also be used with the output subgraph. There is a corresponding pair of -o/-O flags for the output filepath, and output filetype, respectively. So, to write the subgraph to a molfile, write:

./SMSD -Q SMI -q "CCC" -T PDB -t ATP.pdb -O MOL -o subgraph.mol

For convenience, the output filepath can be given as the special name "--", which means "write to stdout". This is a quick way to see the subgraph, especially, if the output filetype is SMI.

./SMSD -Q SMI -q "CCC" -T SMI -t "CCN" -O SMI -o --

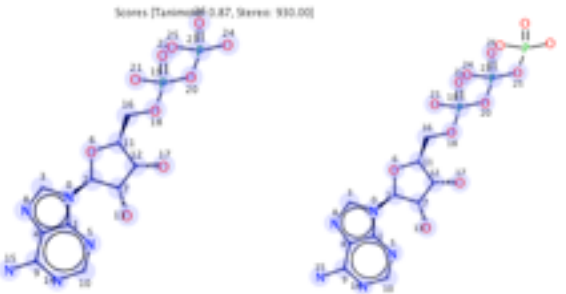
This will just print "CC", as that is the common subgraph of these two smiles.

Images

To generate an image of the isomorphism, use the -g flag, like this:

./SMSD -Q MOL -q ADP.mol -T MOL -t ATP.mol -g

This will generate an image named "ADP_ATP.png" looking something like:



Clearly the name of the image is generated using the names of the query and target input files. If string format molecules are used, the string will be used; for example, "CCC_CCCC.png". The size of the image can be changed with the -d flag, like : -d 400x200 to create an image with a width of 400 and a height of 200.

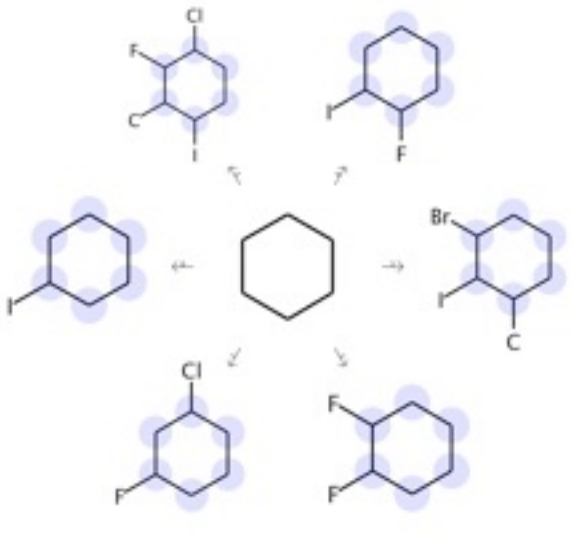
Image options can also be passed at the command-line using the syntax "-Iopt=value". For example, "-IdrawAromaticCircles=true". To get a list of the options, along with their default values, just use the -I flag without any arguments.

N-MCS

To get the maximum common subgraph (MCS) of a set of molecules, provide only a target file, which must be a multi-file format such as SDF. As an example:

./SMSD -T SDF -t arom.sdf -N -g -z

This will produce a 'hub-wheel' image of the MCS that looks like:



where the central (hub) molecule is the MCS of the molecules around the rim of the wheel.

Just use ./SMSD -I to list all the image options.

Usage and command line options:

-A	Appends output to existing files, else creates new files
-a	Add Hydrogen
-b	Match Bond types (Single, Double etc)
-d	Dimension of the image in pixels

-f	Default: 0, Stereo: 1, Stereo+Fragment: 2, Stereo+Fragment+Energy: 3
-g	create png of the mapping
-h,--help	Help page for command usage
-I	Image options
-m	Report all Mappings
-N	Do N-way MCS on the target SD file
-o	Output the substructure to a file
-O	Output type
-Q	Query type (MOL, SMI, etc)
-q	Query filename
-r	Remove Hydrogen
-s	SubStructure detection
-S	Add suffix to the files
-T	Target type (MOL, SMI, etc)
-t	Target filename
-z	Match Rings in the MCS/ Substructure search

Allowed types for single-molecules (query):

MOL	MDL V2000 format
ML2	MOL2 Tripos format
PDB	Protein Databank Format
CML	Chemical Markup Language
SMI	SMILES string format
SIG	Signature string format

Allowed types for multiple-molecules (targets only):

SDF	SD file format

NOTE: *Remove hydrogens before performing graph matching.*