Manual of
SPCI (structural and physico-chemical interpretation) software
version 0.1.3

Version (date)	Changes and comments
0.1.0 (02.05.2015)	Changes from alpha version: 1. More precise SMARTS patterns was added. 2. Cross-validation calculation was sped up and intermediate predictions are saved in a text file. 3. Compounds, which cause errors in calculation of atomic properties with Chemaxon cxcalc tool, are excluded from further modeling. 4. Intermediate results of fragments contributions calculation are saved to a text file.
0.1.1 (07.02.2015)	1. Fixed errors in text output of intermediate results of fragments contributions calculation. 2. Fixed error in loading of file with descriptors on 32-bit platforms.
0.1.2 (13.05.2015)	Fixed error in loading of file with descriptors on 32-bit platforms.
0.1.3 (22.07.2015)	Added two automatic fragmentation schemes: detection of i) all rings and ii) Murcko frameworks.

The SiRMS-SFI software was designed for (semi)automatic extraction of structural features and their contributions to an investigated property from chemical datasets. It's a bunch of Python scripts and it requires:

Python	>= 3.2
matplotlib	1.4.3
numpy	1.9.2
scipy	0.15.1
scikit-learn	0.16.1
setuptools	15.1
python-dateutil	2.4.2
pytz	2015.2
pyparsing	2.0.3
six	1.9.0
indigo toolkit	1.1.12

To simplify the usage a GUI was developed. It has a limited number of options but suits well for first time users and if there is no need in comprehensive analysis.

Installation and launch.

To download the software visit the page http://qsar4u.com/pages/sirms_qsar.php

To make life easier for Windows users WinPython distributions (32 and 64-bit) were prepared and zipped. They contain all required packages for immediate start.

Step 1. Visit the page given above, download a desired WinPython distribution and unzip the archive into a folder.

Step 2. Download sirms-sfi.zip and unpack into the same folder that WinPython (installation folder). You should obtain an installation folder with three items: sirms-gui and WinPython folders and start-sirms-gui.bat.

```
Program_installation_folder
|-sirms-gui
|-WinPython
|-start-sirms-sfi.bat
```

To launch the application choose start-sirms-sfi.bat.

Update:

To update the application just remove sirms-sfi folder from the installation folder and unzip sirms-sfi.zip in it.

Structural *(doesn't require Chemaxon)* and physico-chemical interpretation *(Chemaxon required)*

Structural interpretation returns only overall fragments contributions while physico-chemical interpretation can additionally estimate contributions of some physico-chemical factors (electrostatic, hydrophobic, hydrogen bonding and dispersive terms). Installed Chemaxon is required for physico-chemical interpretation.

If you have license for Chemaxon check whether they are installed and PATH variable (to launch standardize and cxcalc from command line) are correctly configured. If standardize or cxcalc command are not recognized from your command line, add JChem bin folder to PATH variable.

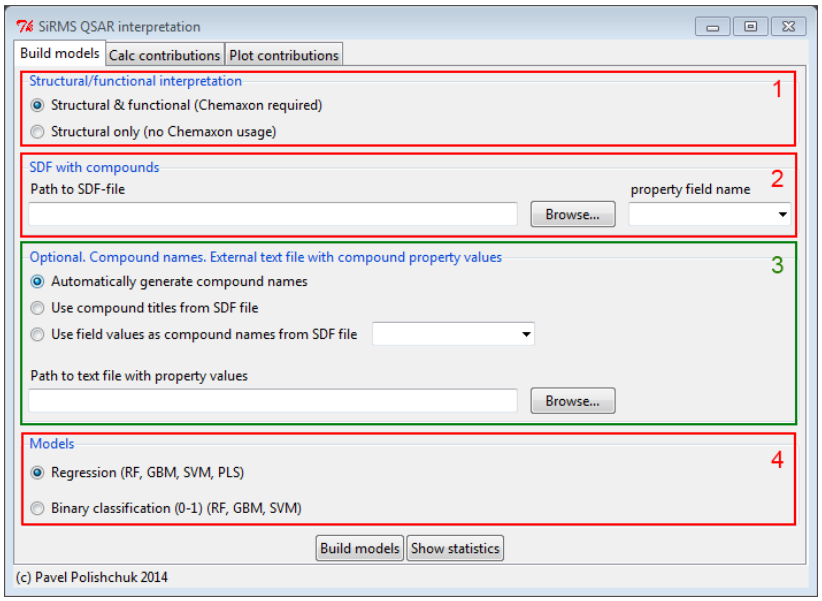
Workflow:

Step 0. Data preparation and project start.

For analysis a single sdf file is required, which contains multiple compounds' structures and corresponding property value as a field in sdf file. As an alternative, you may prepare separate text file with compounds' property values (file must have a header and two columns compound name and property value separated by tab).

Place the sdf file and an optional text file in a separate folder which will be a project folder.

Step 1. Build models.



Red fields are required.

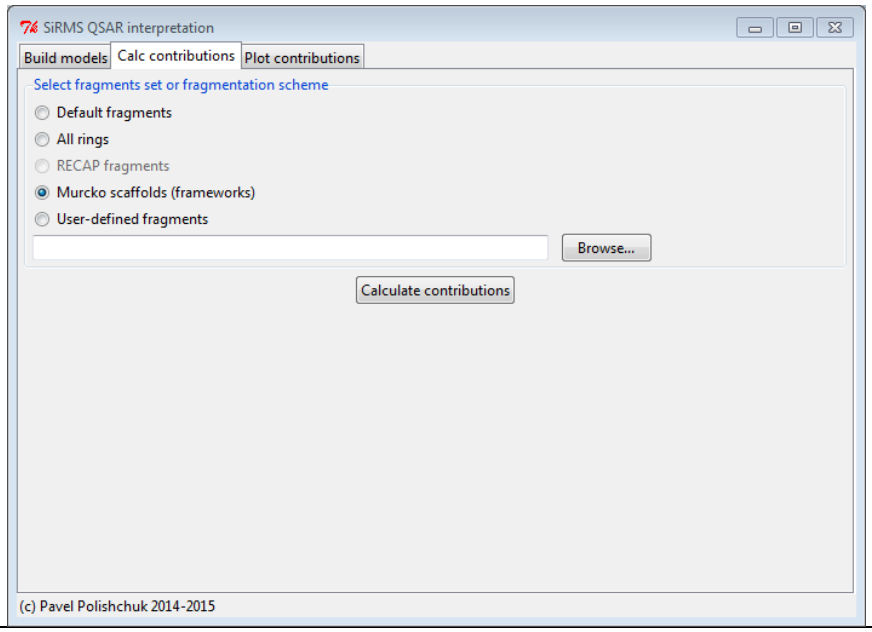
1. Choose the type of analysis to perform.

2. Specify path to sdf file with compounds and choose the name of the field containing property values.

3. Optional. Only required if property values is supplied in a separate file.

4. Choose the desired type of models. All listed models will be build with parameters optimization and 5-fold cross validation will be performed (you may not choose separate models for building in this simplified version of application). **All cores except one will be used automatically.**

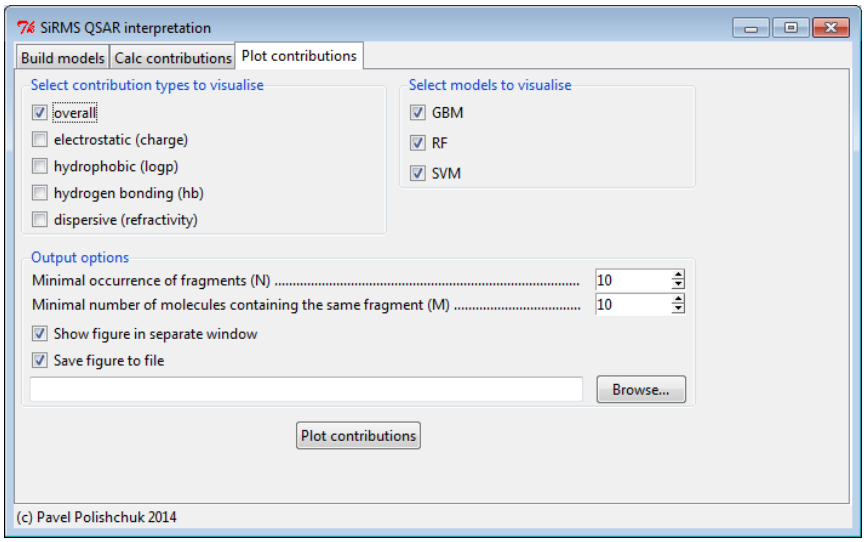
Step 2. Calculation of the fragments' contributions.



- default fragments – common functional groups and rings (specified in default.smarts file);
- all rings – automatically detect all separate single and polycyclic rings;
- Murcko frameworks – automatically detect Murcko frameworks;

user-defined fragments, which preferably should be in SMILES or SMARTS notation (use of SDF is also possible).

Step 3. Plot contributions and save plot in png file.



To plot contribution you need to specify initial sdf file and model type on the first tab.

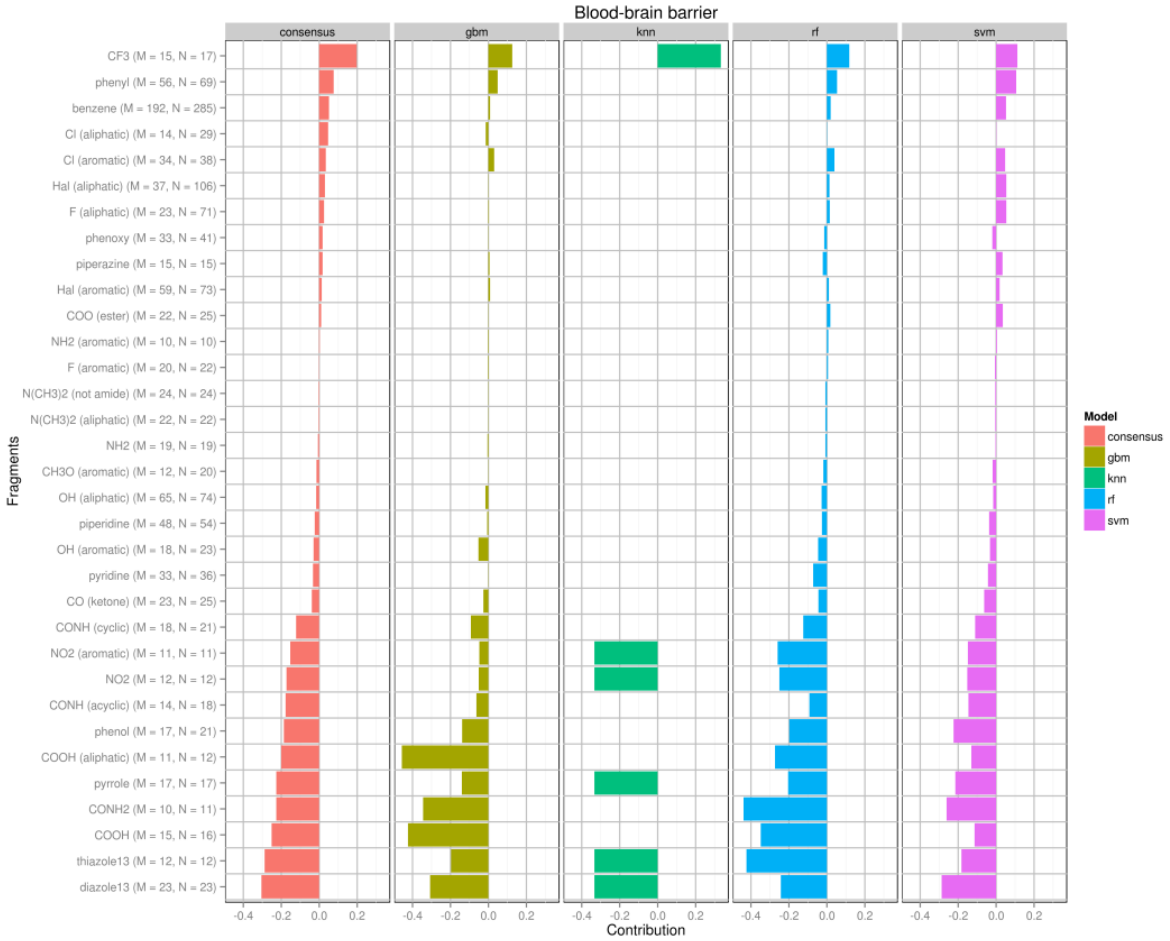
Choose contribution types and models that you want to plot. Overall type is recommended for structural interpretation. Other types is used for physico-chemical interpretation (to show the contribution of separate physico-chemical factors).

Alternative step 3.

To visualize and customize the plot you may use free web tool.

Full version - https://pavel.shinyapps.io/sirms-qsar-vis/ (to start just upload the file with calculated contributions, it is located in the project folder and named e.g. default_frag_contributions.txt)

Demo version - https://pavel.shinyapps.io/sirms-qsar-vis-demo (to play with pre-uploaded fragments contribution for blood-brain barrier dataset)


Blood-brain barrier
consensus
gbm
knn
rf
svm
CF3 (M = 15, N = 17)
phenyl (M = 56, N = 69)
benzene (M = 192, N = 285)
Cl (aliphatic) (M = 14, N = 29)
Cl (aromatic) (M = 34, N = 38)
Hal (aliphatic) (M = 37, N = 106)
F (aliphatic) (M = 23, N = 71)
phenoxy (M = 33, N = 41)
piperazine (M = 15, N = 15)
Hal (aromatic) (M = 59, N = 73)
COO (ester) (M = 22, N = 25)
NH2 (aromatic) (M = 10, N = 10)
F (aromatic) (M = 20, N = 22)
N(CH3)2 (not amide) (M = 24, N = 24)
N(CH3)2 (aliphatic) (M = 22, N = 22)
NH2 (M = 19, N = 19)
CH3O (aromatic) (M = 12, N = 20)
OH (aliphatic) (M = 65, N = 74)
piperidine (M = 48, N = 54)
OH (aromatic) (M = 18, N = 23)
pyridine (M = 33, N = 36)
CO (ketone) (M = 23, N = 25)
CONH (cyclic) (M = 18, N = 21)
NO2 (aromatic) (M = 11, N = 11)
NO2 (M = 12, N = 12)
CONH (acyclic) (M = 14, N = 18)
phenol (M = 17, N = 21)
COOH (aliphatic) (M = 11, N = 12)
pyrrole (M = 17, N = 17)
CONH2 (M = 10, N = 11)
COOH (M = 15, N = 16)
thiazole13 (M = 12, N = 12)
diazole13 (M = 23, N = 23)
Fragments
-0.4
-0.2
0.0
0.2
Contribution
Model
consensus
gbm
knn
rf
svm