SARANEA

SARANEA is a free Java program that allows the visualization and interactive exploration of molecular datasets annotated with biological activity information. It relates compound activity or selectivity and structural similarity of molecules to each other. Thus it can aid in the elucidation of structure-activity and structure-selectivity relationships. Information on multiple targets can be compared with each other and combined in order to assess the activity profile of a compound identified as an activity/selectivity cliff marker for one particular target.

Input format

The input for SARANEA consists of a list of molecules for each target including potency values. In addition, molecule sets can be provided that do not have a potency associated with them but serve as marker compounds, for example to identify potentially toxic compounds.

The single input file to be opened with SARANEA is a zip folder containing three subfolders, named *SARSets*, *ADDSets*, and *Settings*. Each of these subfolders contains comma separated value (.csv) files that list the compounds and additional information:

SARSets

This folder contains compounds with potency information. For each target, a .csv file that has the name of the target (e.g. *Thrombin.csv*) should be deposited in this folder. The first line of each file should contain field names. An arbitrary number of user-defined descriptors can be added to the file. However, the following fields MUST be present in order for SARANEA to work properly:

SMILES	Molecule in canonical SMILES format. They must be unique. SARANEA uses SMILES to identify identical compounds in different sets.
ID	An identifier (text). This should, but does not need to be unique.
NAME	A name that will be displayed for the molecule.
POT	Potency of the molecule in nM.
FP	Molecular fingerprint (space separated feature list).

ADDSets

This subfolder contains compounds that do not have a potency annotation. They can be used as marker molecules to identify bioactive compounds that have structural similarity to these markers. For example, toxic compounds could serve as markers. For each additional set of marker compounds, a .csv file has to be provided that has a unique name identifying the marker set (e.g. Toxic.csv). These sets appear in the target list in SARANEA as red labels. The first line of each file should contain field names. An arbitrary number of user-defined descriptors can be added to the file. However, the following fields MUST be present in order for SARANEA to work properly:

SMILES	Molecule in canonical SMILES format. They must be unique.
ID	An identifier (text). This should, but does not need to be unique.
NAME	A name that will be displayed for the molecule.
FP	Molecular fingerprint (space separated feature list).

The ADDSets folder can be empty.

Settings

This folder contains a .csv file for each applicable similarity threshold. For example, if the similarity threshold to be used is Tc > 0.65, the settings for this threshold are provided in the file *065.csv*. Each file contains two columns. In the first column, the name of the parameter is given. The second column provides the value for the parameter.

These files serve normalization purposes for each fingerprint and should not be changed. However, given a valid SARANEA zip folder, the normalization utility (see below) can be used to generate a setting file for a given similarity threshold. Thus, normalized scores can be calculated for arbitrary reference sets, similarity thresholds, and fingerprints.

Starting SARANEA

Double-click on the SARANEA.bat provided in the download package. Make sure you have the Java Runtime Engine 1.6 or later installed.

In the opening window, click on the menu File/Open. A dialogue will show up where you can select the SARANEA zip folder you want to open. After clicking "OK", a dialogue will appear where you have to select a similarity threshold. Available thresholds are provided in the Settings subfolder (as discussed in **Input format**). Select the threshold of your choice from the drop-down list and click ok.

Console

On the console (opened by starting SARANEA.bat) information about the current program status, along with some debug output is printed. You will usually not need to look at this output. However, if you have the impression that the program is "frozen" it might be actually due to a calculation that is in progress or because a file is being read. In these cases a look at the console often provides information about what the program is doing at a particular moment.

Main Controls

If the file has been read successfully, the message "Settings, Molecules, and Targets read." will appear on the console. In the main window, a panel will be shown that contains the main controls that are used to select molecules and/or targets that can then be analyzed in further detail.

Show Molecule Info	Opens an Info Panel for the molecule selected in the top drop-down list.
Add to current set	Adds the target or marker set selected in the middle drop-down list to the current combination of targets and marker sets.
Show Graph	Calculates and shows an interactive view of the network-like similarity graph (NSG) for the selected targets.
Write CSV	Calculates the NSG and writes it to a .csv file without interactive display.

Selecting Targets

In order to analyze structure-activity relationships, select the target of interest from the target drop-down list and click "Add to current set". Notice how the target is added to the list below the button. In order to remove it from the list, select it and press *Delete* on your keyboard.

When you add targets to the set list, the target drop-down list is automatically updated. It contains all targets that share at least five molecules with the selected target. In order to analyze structure-selectivity relationships, a second target can be added to the list.

Selection of marker sets

If any marker sets were provided in the *ADDSets* folder, they are also listed in the target drop-down list, but are highlighted in red. If a marker set was selected and the "Add to current set" button is clicked, a color dialogue appears. Each marker set can be assigned an individual color that will identify compounds belonging to the set. When a color has been chosen, a second dialogue appears and asks to provide a similarity threshold. Marker compounds are only included in the NSG if their

similarity value to at least one active molecule exceeds the predefined threshold or if they are active themselves. Moreover, edges between marker molecules are not shown.

Showing the graph

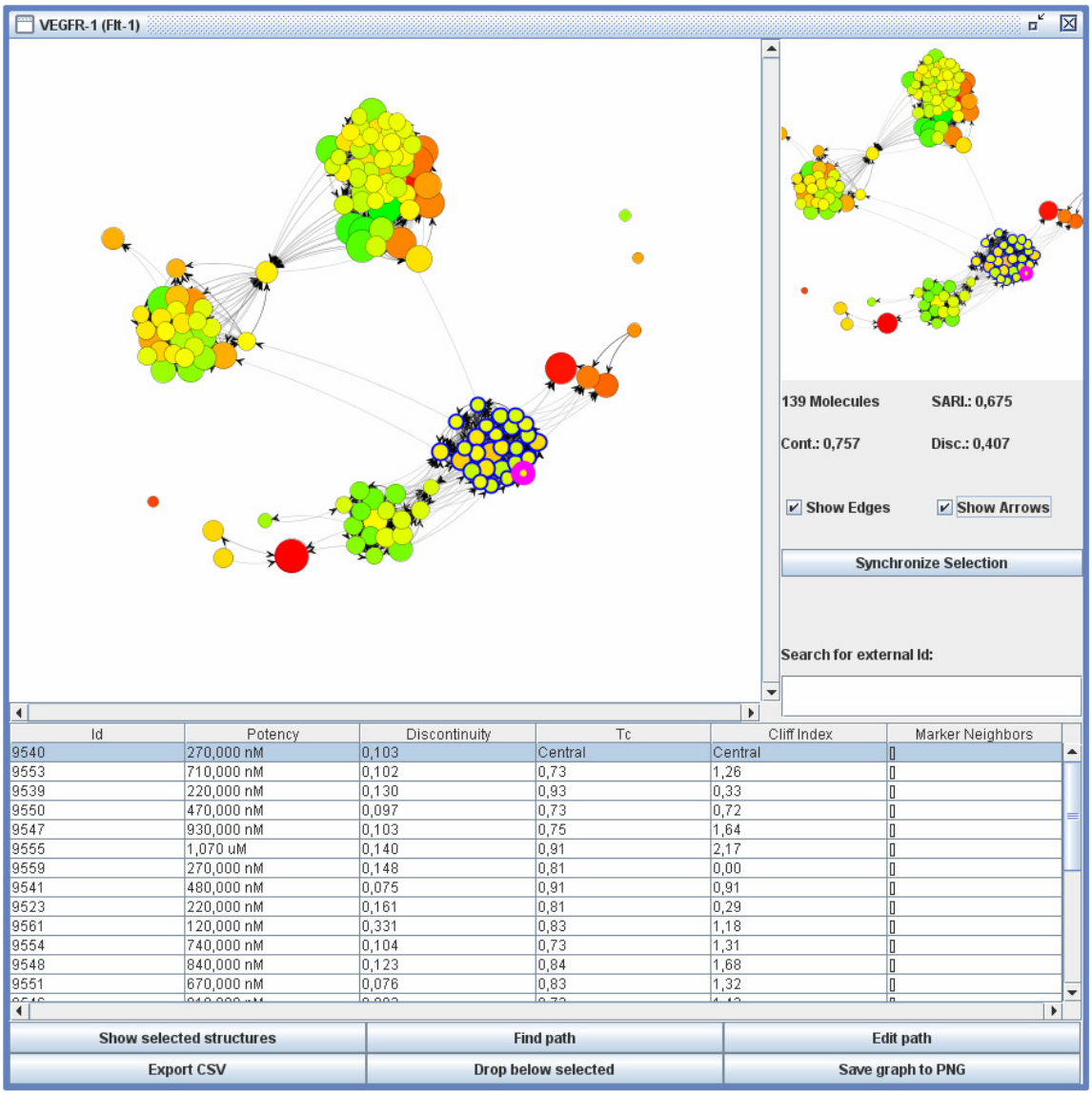
The central window of SARANEA is the interactive graph window that shows the currently assembled set. In order to create such a window, add the targets and markers of interest to the current set and click on the “Show graph” button.

The NSG will be calculated and then the graph window is shown. Depending on the size of the set and the speed of your computer, this may take some time. You can observe the progress in the console.

As soon as the NSG has been calculated and the molecule depictions have been preloaded, an internal frame will appear in the white field left to the panel (if you do not see a white field, try resizing the main window).

If the set is sufficiently large, you will probably see the layout algorithm repositioning the nodes. You can abort the layout algorithm by clicking anywhere inside the graph and pressing *Esc* on your keyboard.

The NSG window



Id	Potency	Discontinuity	Tc	Cliff Index	Marker Neighbors
9540	270,000 nM	0,103	Central	Central	[]
9553	710,000 nM	0,102	0,73	1,26	[]
9539	220,000 nM	0,130	0,93	0,33	[]
9550	470,000 nM	0,097	0,73	0,72	[]
9547	930,000 nM	0,103	0,75	1,64	[]
9555	1,070 uM	0,140	0,91	2,17	[]
9559	270,000 nM	0,148	0,81	0,00	[]
9541	480,000 nM	0,075	0,91	0,91	[]
9523	220,000 nM	0,161	0,81	0,29	[]
9561	120,000 nM	0,331	0,83	1,18	[]
9554	740,000 nM	0,104	0,73	1,31	[]
9548	840,000 nM	0,123	0,84	1,68	[]
9551	670,000 nM	0,076	0,83	1,32	[]

The large view shows the graph. Each node represents a molecule and a tooltip with the molecule structure is shown when hovering over a node. The upper right graph view is a satellite view and allows re-positioning the viewport if the main view shows a zoomed-in section of the graph.

Selected nodes have a blue circle around them and properties of the respective molecules appear in the table view below the main view. The molecule selected in the table is represented by a node with a wide, magenta circle.

Interacting with the graph

Inside the main graph view of the NSG window, the following mouse and key controls apply:

Mouse controls	
Left-click and drag outside node	Draw selection box
Left-click and drag inside node	Select and drag node
Left-double-click on node	Select neighbors of node
Right-click on node	Open node context menu
Mouse wheel	Zoom graph
Hold mouse wheel down and drag	Translate (move) graph

Key controls	
Shift down while clicking	Add / remove nodes from selection
Escape	Clear selection, abort layout algorithm
Shift+P	Find Path
Shift+T	SAR Tree Viewer
Shift+B	Edit path

In the small graph view, the left mouse button is used to drag the view port and the mouse wheel to zoom in or out.

Interpretation of the graph

Each node represents a molecule. Two molecules are connected if their molecular similarity, calculated as the Tanimoto coefficient of the two fingerprints, is above the current similarity threshold. Edge shading reflects the respective similarity value. Dark edges connect highly similar compounds, whereas light edges connect pairs of molecules with similarity close to the threshold. Arrowheads point from weakly active to strongly active compounds. In selectivity sets, they point from selectivity for the first target to selectivity for the second target.

Nodes are colored according to compound potency or selectivity. If colored according to potency, green means weakly, red means highly potent. The color spectrum is

adjusted to the minimum and maximum of the current set. If the set combines two targets, green means selectivity for one target, red means selectivity for the other target. Yellow nodes represent non-selective compounds. The potency and selectivity values are provided in the respective tooltips in the bottom left corner.
Black nodes represent marker compounds from additional sets.
The nodes are scaled according to their local discontinuity score: large nodes represent activity / selectivity cliff markers, whereas small nodes represent compounds within a smooth activity / selectivity landscape.

The info panel

In the right info panel, four parameters describing the current set are given. The number of molecules is the total number of compounds in the graph. The three scores Discontinuity, Continuity, and SARI account for the nature of the structure-activity/selectivity relationship within the current set. They range from 0 to 1. A high discontinuity score means that the set contains many cliffs, whereas a high continuity score accounts for the presence of many structurally distinct compounds with similar activity/selectivity. The SAR Index (SARI) categorizes the set with respect to its SAR type. Possible values range from 0 to 1 and low, intermediate and high values reflect the three general types of SAR, discontinuous, heterogeneous and continuous, respectively.

Hiding edges and arrows

In order to make the interaction with the graph more fluent, edges and arrowheads can be hidden by deselecting the respective checkboxes. If the set contains more than 100 molecules, edges are hidden automatically to facilitate the layout. However, they can be switched on at any time.

Synchronizing the selection

If multiple NSG windows for different sets are open, the selection of compounds between all windows can be synchronized by clicking on the “Synchronize selection” button. Compounds that are selected in the currently active window are then selected in all other windows, provided they are present in the respective set.

Searching for a molecule

In order to select a molecule with a particular ID, the ID can be typed into the respective text field. After pressing enter, the molecule will be selected in the graph. All other molecules will be deselected! If the provided ID was not found, an error message will appear in the console.

Node context menu

Right-clicking on a node opens a context menu with the following entries:

Molecule Info	Shows molecule info for the node in a new window.
Select Neighbors	Selects the node's neighbors (same as doubleclick)
Show SAR Tree	Opens the SAR Tree view for the particular node.

The table

The table at the bottom of the NSG window lists currently selected molecules in the graph. It provides six columns plus additional columns for every user-defined property that was provided in the respective .csv files. Columns can be sorted by clicking on the column header.

ID	The ID of the molecule
Potency / Selectivity	The potency or selectivity of the compound.
Discontinuity	A compound discontinuity score. High values represent cliff markers
Tc	The Tanimoto coefficient between the "Central" compound and this compound.
Cliff Index	A value that can be used to identify cliffs. High values represent similar molecules to the central molecule with a different potency / selectivity.
Marker Sets	A list of marker sets the compound is part of.

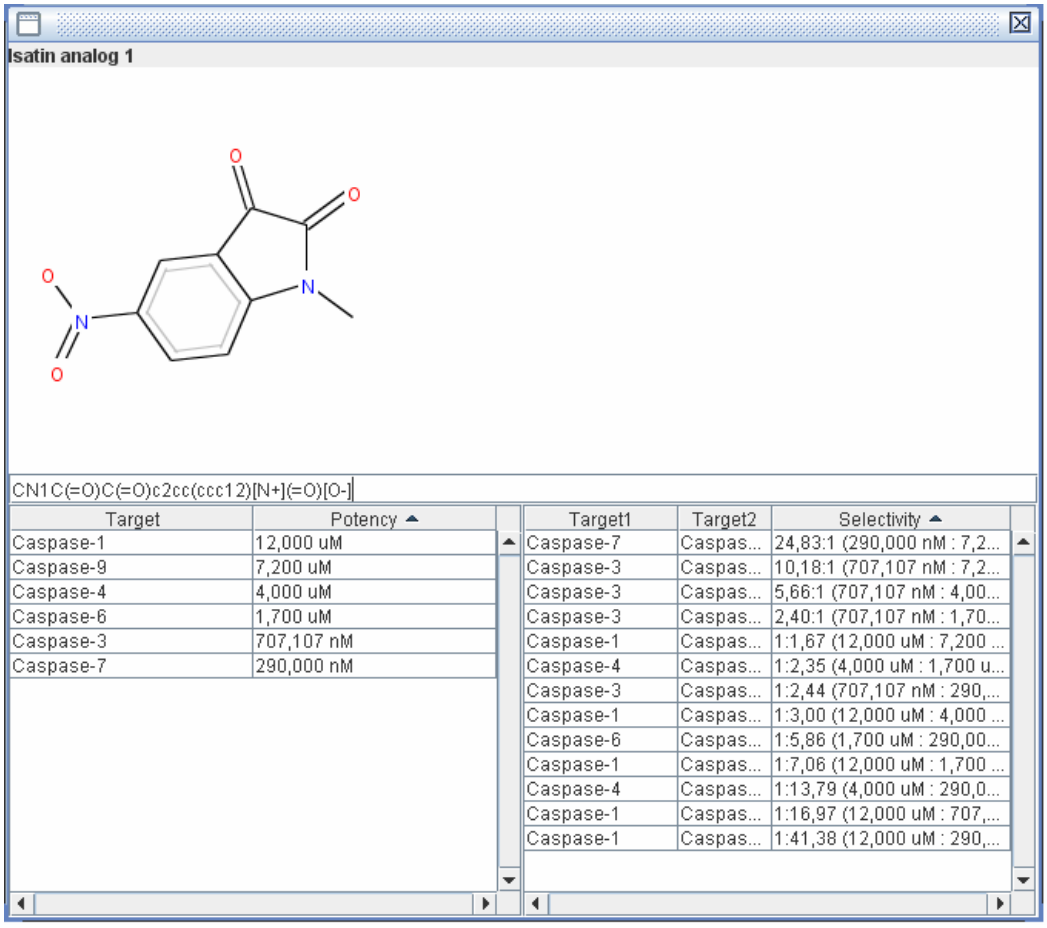
The compound marked as "central" is the compound selected first. If only one compound is selected, it is marked as "central".

In order to select only a subset of compounds, the button “Drop below selected” can be clicked. This will deselect all compounds that are reported below the selected compound in the (potentially sorted) table. This feature can be used in order to focus on highly potent / selective compounds or activity cliffs.

Export CSV

The current table can be saved in a .csv file. The button opens a “Save File…” dialogue where a destination .csv can be specified. After clicking “Save”, the table is written to that file. In the csv version, the potency and selectivity are reported as pKi values or the difference of pKi values, respectively.

Molecule Info



The molecule info frame can be opened by selecting “Show Molecule Info” in the main window or from the node context menu. It provides information about a

particular molecule and all targets. The left table shows potency, the right table shows selectivity values. By clicking on the column headings the tables can be sorted.
Double-clicking on a table entry starts NSG calculation and opens the respective NSG window. Watch the console for progress updates! The molecule depiction shows the name of the compound and a text field where the SMILES string can be copied from.

Show selected structures

By clicking on this button a new window opens that shows the currently selected molecules similar to the molecule info panel in a list. Clicking on a molecule selects it in the table. In all graphs it is automatically shown with a big magenta circle around it.
The list updates any time the table is updated. In order to show the updated list, it has to be activated (by clicking on it or the respective button).

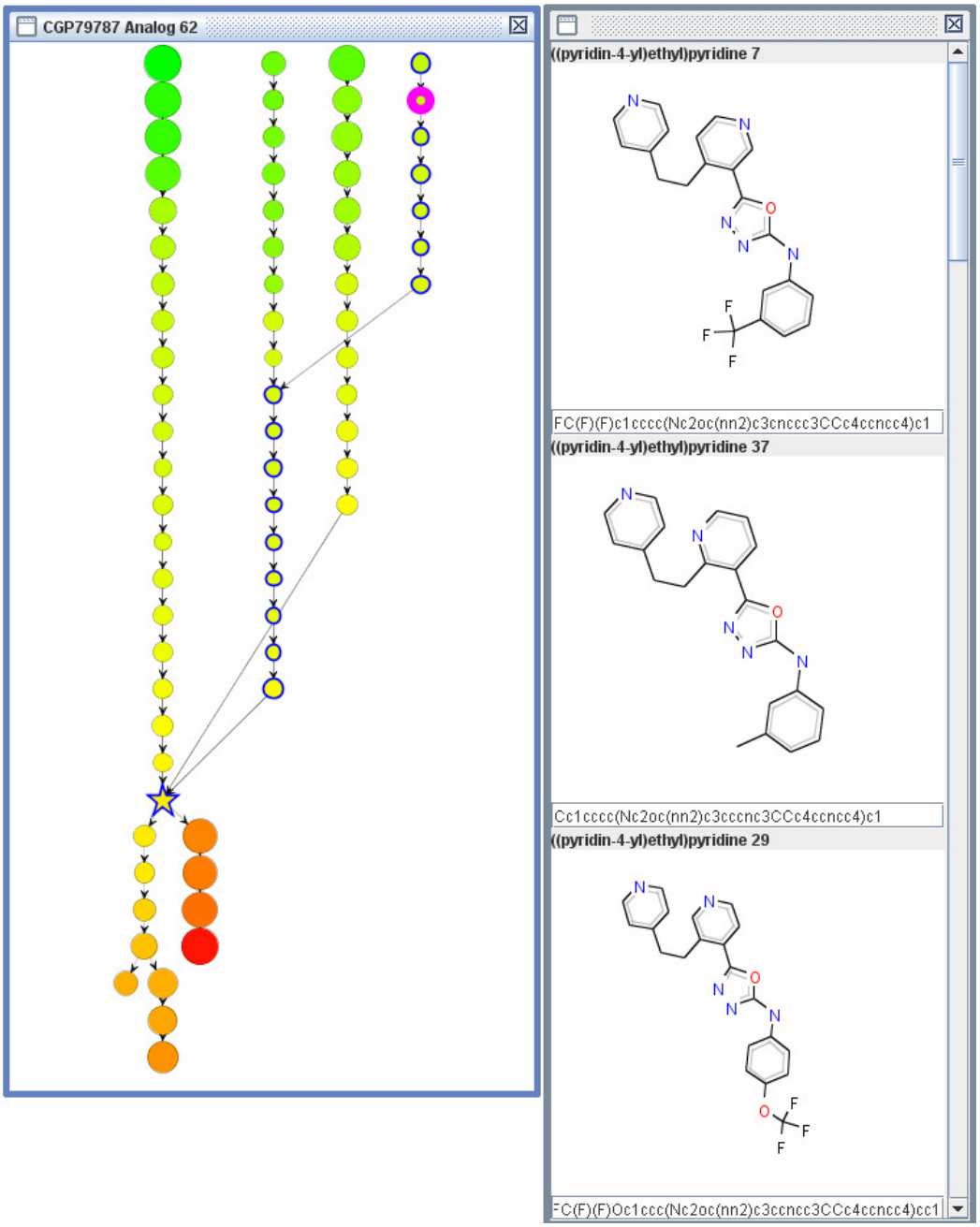
Find path

If two nodes are selected and this button is pressed, SARANEA tries to find the shortest path between the two nodes. It takes into account potency/selectivity and molecular similarity. Paths are prioritized that show a linear increase in potency / selectivity and represent gradual structural changes (high molecular similarity between every two compounds that are connected in the path). Sorting the table by potency / selectivity and showing the selected structures provides a convenient way to examine the identified path.

Show SAR Tree

This option is present in the context menu of each node. It opens a new window showing a subgraph in a tree-like layout. The star represents the compound for which the SAR tree was calculated. The SAR tree shows all pathways originating and ending at this compound. The pathways are scored and the scores are shown as tooltips of the edges. High scores represent pathways with a linear increase in potency / selectivity and gradual structural changes. Zooming and moving the graph are analogous to the NSG window.
Double-clicking on a vertex selects the maximally scored path that the vertex belongs to. Right-clicking opens a context menu that allows editing the path in addition to selecting a molecule's neighbors and showing the molecule info panel. The selection

in the SAR Tree viewer is coupled to the NSG window. Whichever compounds are selected / deselected here are also selected / deselected in the NSG window and table. The following screenshot illustrates the SAR Tree viewer and the molecule list. The selected path is highlighted.



Path Builder

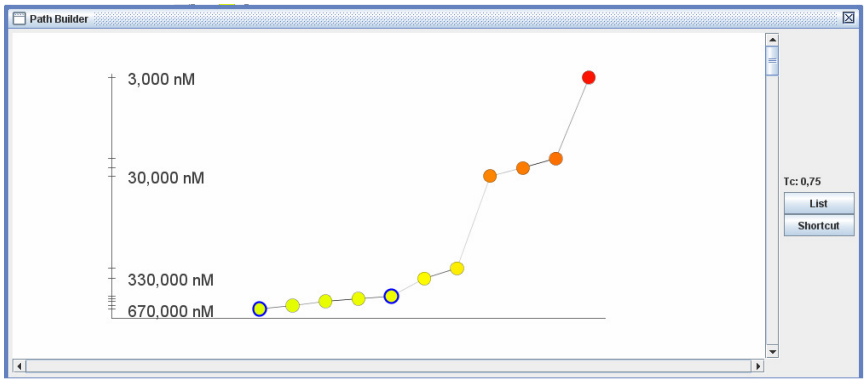
The Path Builder can be opened in three different ways: by pressing Shift+B in the NSG window graph; clicking on the button “Edit path”, or selecting the context menu entry in the SAR Tree Viewer.

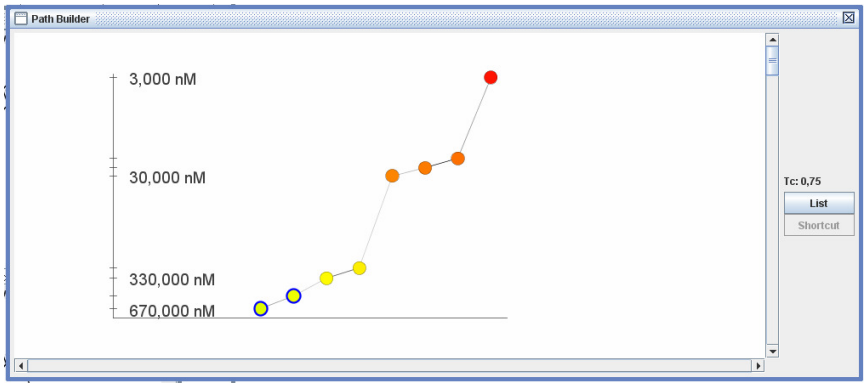
If the current selection does not contain compounds with ascending potency / selectivity that are connected, an error message will appear. The SAR Tree Viewer always produces a valid path.

The path builder shows individual nodes (equally sized) in a line graph reporting the potency / selectivity of each compound.

Shortcuts

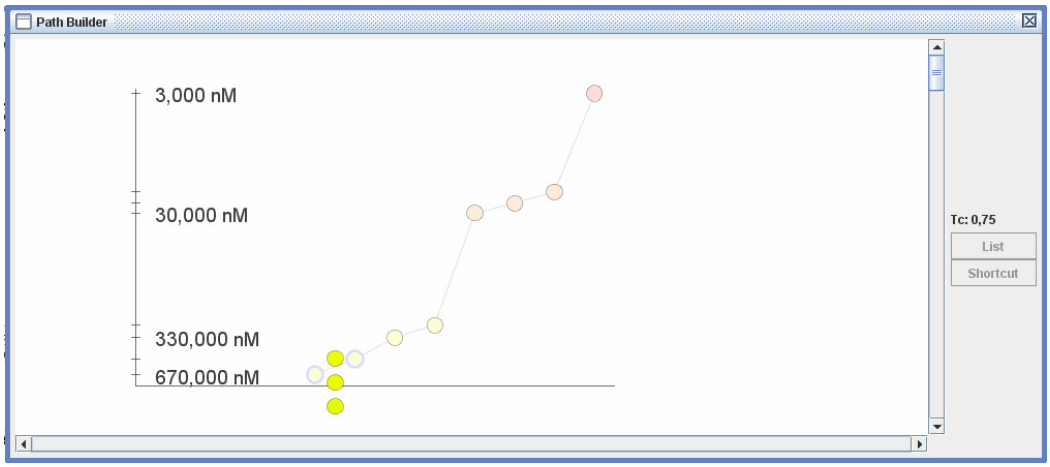
If two compounds are connected in the NSG and are selected in the Path Builder, the button “Shortcut” becomes activated. Clicking this buttons deletes all nodes between the two compounds. This feature is useful for paths with minimal change in potency over a large number of similar compounds.

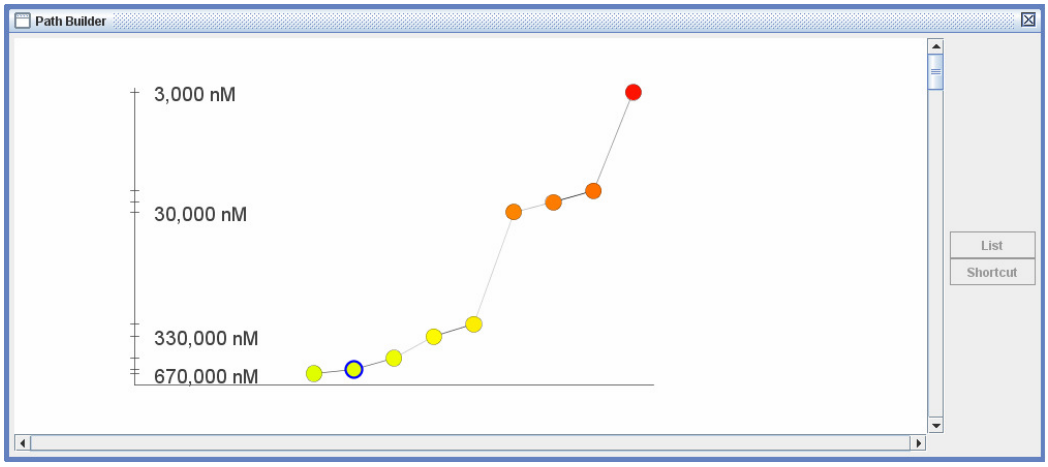




List

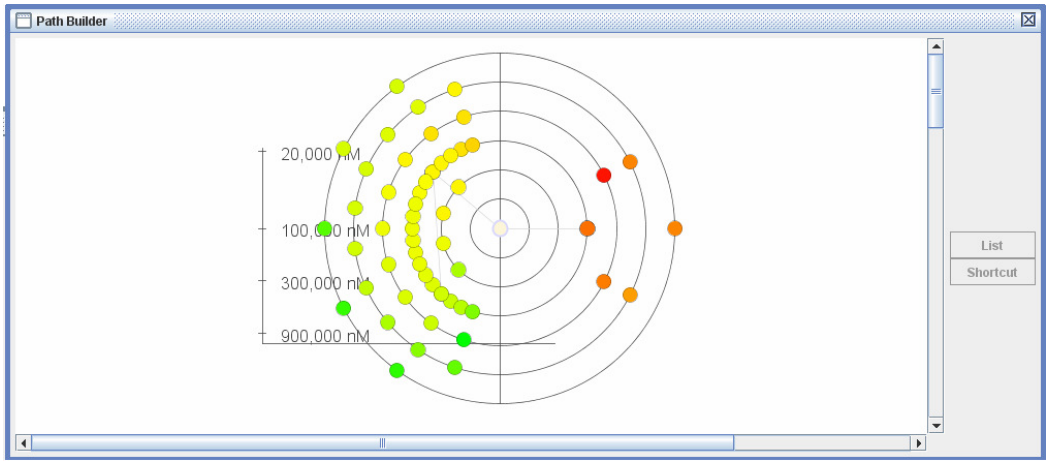
If two vertices are selected and successors of one vertex overlap with predecessors of the other vertex, the button “List” becomes activated. Clicking this button shows a list of compounds between the two vertices that can be inserted between them. If vertices are part of the path, they are shown in their proper position. The rest of the vertices is faded out. By clicking and dragging the compounds in the list can be selected (for example, to show them in the molecule list or look at their properties in the NSG window table). Left-clicking on a vertex from the list inserts it between the two selected vertices of the path. Hitting *Esc* on the keyboard or clicking somewhere else aborts the list view.





Radial neighborhood view

This view can be activated by double-clicking on a node or by selecting the respective entry in the node context menu. It shows all neighbors of the node in a radial graph. Nodes are positioned on concentric circles that represent bins of structural similarity to the selected note, which is shown at the center. Peripheral nodes are structurally dissimilar to the central node. On the left are the predecessors, on the right the successors of the central node representing less potent compounds and more potent compounds, respectively:



Clicking on a node positioned on a shell selects it as the new predecessor / successor and deletes all old predecessors / successors. Clicking on the central node or anywhere in the graph aborts the radial neighborhood view.

Normalizer

The Normalizer is a command-line tool to derive the settings needed for normalization for different fingerprints and similarity threshold values. It can be started by executing Normalizer.bat and specifying the SARANEA zip folder and the desired threshold. The normalization parameters are printed out at the standard output. In order to write them into a file, redirect the standard output to a file, e.g.:

```
Normalizer.bat MySet.zip 0.75 > .\Settings\075.csv
```