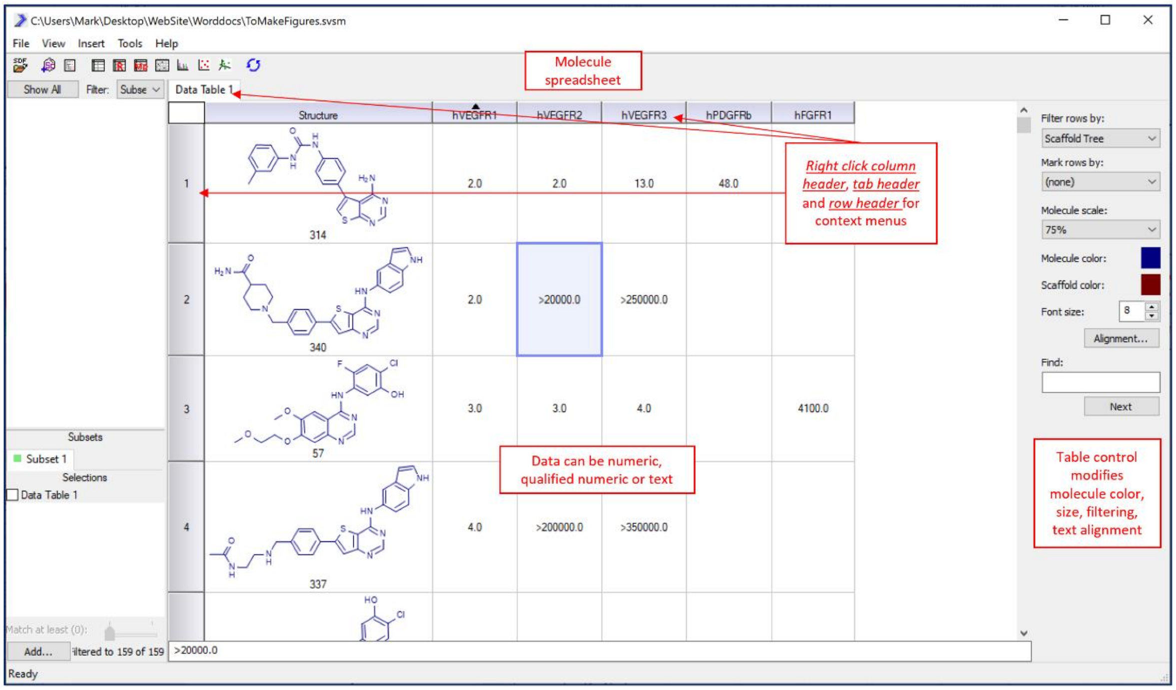
# 1. Create a Molecule Spreadsheet Data Table to Explore Structure-Activity Relationships

**SARVision$^{TM}$** allows you to load a list of molecules as SDF or CSV file (with SMILES code) and create a database of compounds and associated data including chemical descriptors such as molecular weight, polar surface area, and lipophilicity (clogP).

In the main menu FILE-IMPORT (sdf.file or csv with SMILES codes) opens a browser that allows one to navigate and import a molecule file. You can import multiple sdf. or csv files into a SARVision project file and large files of compounds are possible.

Molecules will appear in the left most column under the heading Structure. Any other associated data included in the file will appear as columns located to the right of the structure column. Your data sheets used by SARVision must have the SMILES code in the 1st column and labeled as Structure for presentation as structures in column 1.



Molecule spreadsheet

*Right click column header, tab header* and *row header* for context menus

Data can be numeric, qualified numeric or text

Table control modifies molecule color, size, filtering, text alignment

| | Structure | hVEGFR1 | hVEGFR2 | hVEGFR3 | hPDGFRb | hFGFR1 |
|---|---|---|---|---|---|---|
| 1 | 314 | 2.0 | 2.0 | 13.0 | 48.0 | |
| 2 | 340 | 2.0 | >20000.0 | >250000.0 | | |
| 3 | 57 | 3.0 | 3.0 | 4.0 | | 4100.0 |
| 4 | 337 | 4.0 | >200000.0 | >350000.0 | | |

>20000.0

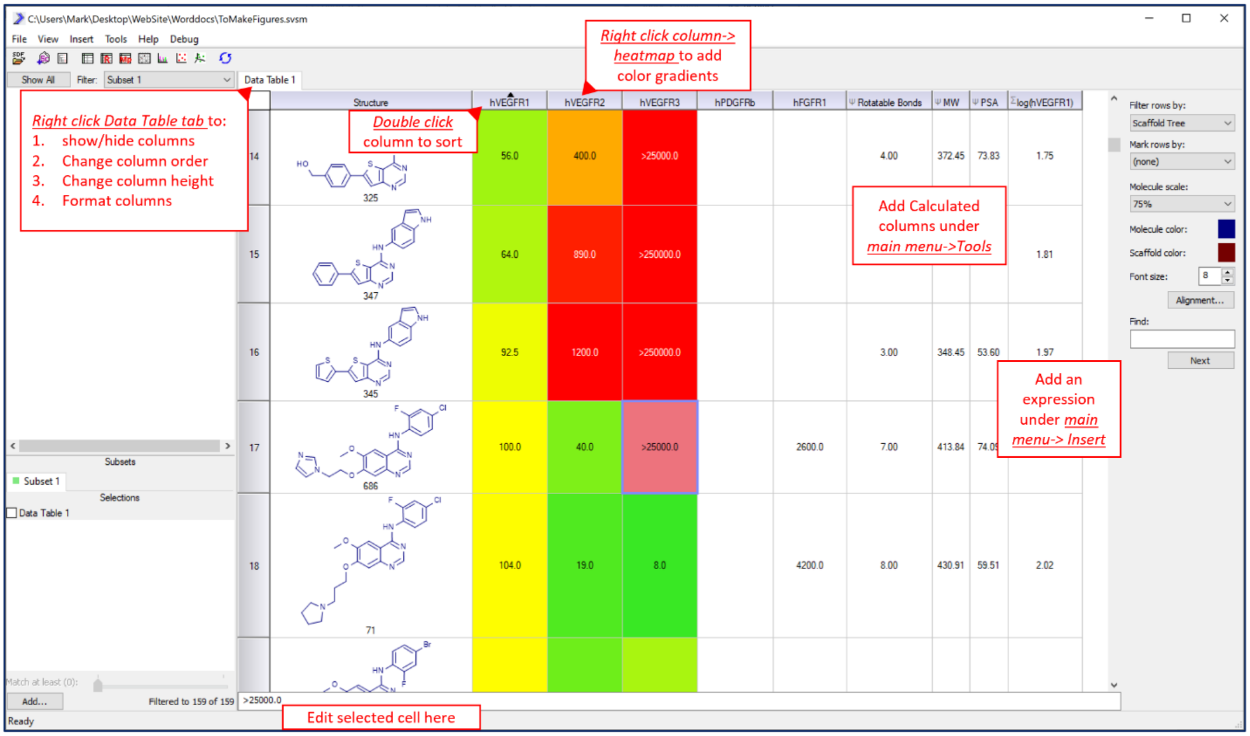
Filter rows by: Scaffold Tree
Mark rows by: (none)
Molecule scale: 75%
Molecule color: 🟦
Scaffold color: 🟥
Font size: 8

- **PANEL 1. Data Table.** Right click on row headers, spreadsheet header tab, column header and cells to get context menus to modify objects. Click on any data cell to modify or add data.

The data table is a fully functional Excel spreadsheet where columns can be changed by:

- ascending of descending data values (double click the column header to sort)
- columns can be hidden and reordered

- columns can be heat-mapped to highlight activity data.
- Additional data can be added or calculated using the *main menu->insert* functionality
- Adding a *Column* adds an empty column on the end of the spreadsheet that the user can type in data, comments, or other information.
- Adding an *Expression* allows the user to mathematically transform another column; for example the **log**() function can be applied to another column to create a new one.



Filter: Subset 1

*Right click Data Table tab* to:
1. show/hide columns
2. Change column order
3. Change column height
4. Format columns

*Double click* column to sort

*Right click column-> heatmap* to add color gradients

Add Calculated columns under *main menu->Tools*

Add an expression under *main menu-> Insert*

| | Structure | hVEGFR1 | hVEGFR2 | hVEGFR3 | hPDGFRb | hFGFR1 | Rotatable Bonds | MW | PSA | log(hVEGFR1) |
|---|---|---|---|---|---|---|---|---|---|---|
| 14 | 325 | 56.0 | 400.0 | >25000.0 | | | 4.00 | 372.45 | 73.83 | 1.75 |
| 15 | 347 | 64.0 | 890.0 | >250000.0 | | | | | | 1.81 |
| 16 | 345 | 92.5 | 1200.0 | >250000.0 | | | 3.00 | 348.45 | 53.60 | 1.97 |
| 17 | 686 | 100.0 | 40.0 | >25000.0 | | 2600.0 | 7.00 | 413.84 | 74.0 | |
| 18 | 71 | 104.0 | 19.0 | 8.0 | | 4200.0 | 8.00 | 430.91 | 59.51 | 2.02 |

☐ Data Table 1

Subset 1

Match at least (0):

Filtered to 159 of 159

>25000.0

Edit selected cell here

Ready

Filter rows by: Scaffold Tree

Mark rows by: (none)

Molecule scale: 75%

Molecule color:

Scaffold color:

Font size: 8

Find:

Molecular Data Tables can be sorted, edited, heat-mapped and cells formatted to help identify trends in biological assay data.

Under Data Table tab (*right click*) you can export to Excel and/or Word for further data input or presentations of structure-activity data.

**Common Structural Cores: The Basis of Biological and Chemical Properties**

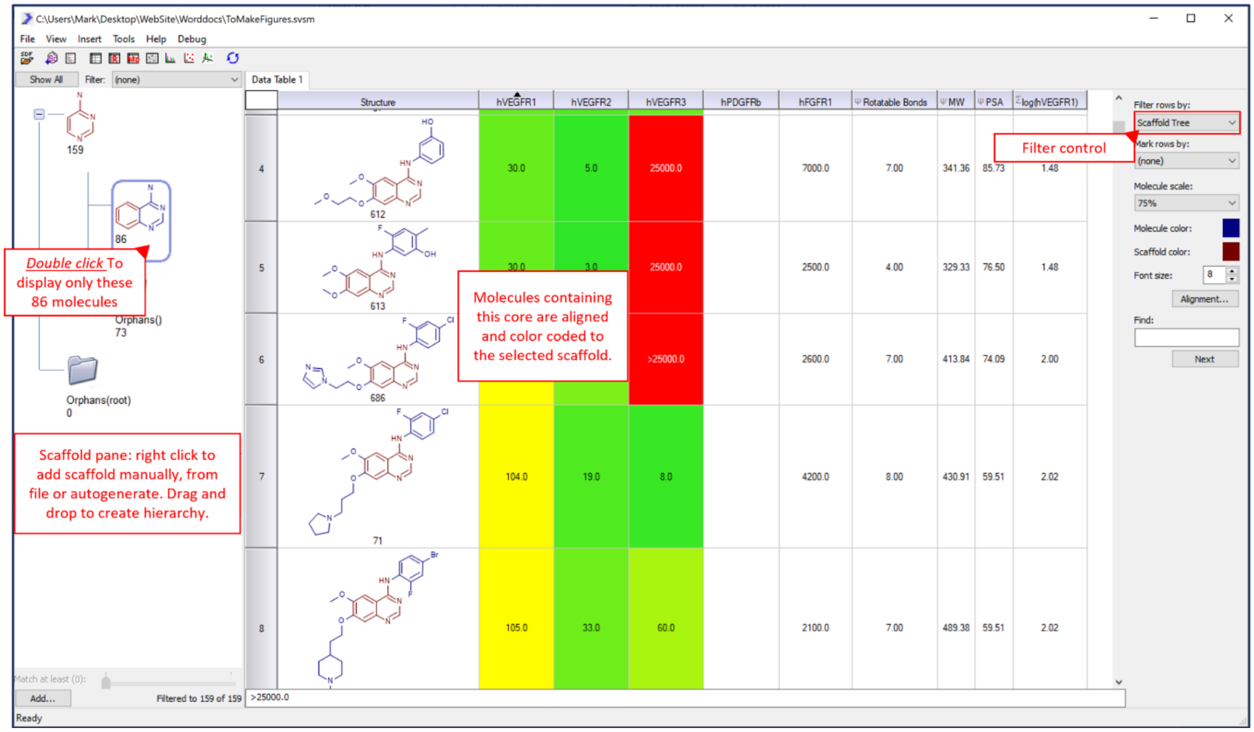
In a Data Table, molecules are grouped by a ***common structural core*** termed a ***scaffold*** to create families of related molecules.

In **SARVision** any number of scaffolds can be added for analysis; or scaffolds specific to this data set can be drawn by the user.

By right clicking in the scaffold pane (panel to the left) options are given:

- **right click->Identify scaffolds**- automatically finds chemical scaffolds and lists them
- **right click->Add scaffold**- draw your desired compound or scaffold
- **right click->Import scaffold**-import a list of compounds to search a Data Table with

Molecules in **SARvision** can be aligned and color coded relative to any scaffold by double clicking on the scaffold of interest. Additionally, clicking on scaffolds can generate other views such as **R-Group tables** and **scaffold centered molecular pair tables.**



Filter: (none)

Filter rows by: Scaffold Tree

Mark rows by: (none)

Molecule scale: 75%

Font size: 8

Filtered to 159 of 159

>25000.0

| | Structure | hVEGFR1 | hVEGFR2 | hVEGFR3 | hPDGFRb | hFGFR1 | Rotatable Bonds | MW | PSA | log(hVEGFR1) |
|---|---|---|---|---|---|---|---|---|---|---|
| 4 | 612 | 30.0 | 5.0 | 25000.0 | | 7000.0 | 7.00 | 341.36 | 85.73 | 1.48 |
| 5 | 613 | 30.0 | 3.0 | 25000.0 | | 2500.0 | 4.00 | 329.33 | 76.50 | 1.48 |
| 6 | 686 | | | >25000.0 | | 2600.0 | 7.00 | 413.84 | 74.09 | 2.00 |
| 7 | 71 | 104.0 | 19.0 | 8.0 | | 4200.0 | 8.00 | 430.91 | 59.51 | 2.02 |
| 8 | | 105.0 | 33.0 | 60.0 | | 2100.0 | 7.00 | 489.38 | 59.51 | 2.02 |

Filter control

*Double click* To display only these 86 molecules

Molecules containing this core are aligned and color coded to the selected scaffold.

Scaffold pane: right click to add scaffold manually, from file or autogenerate. Drag and drop to create hierarchy.

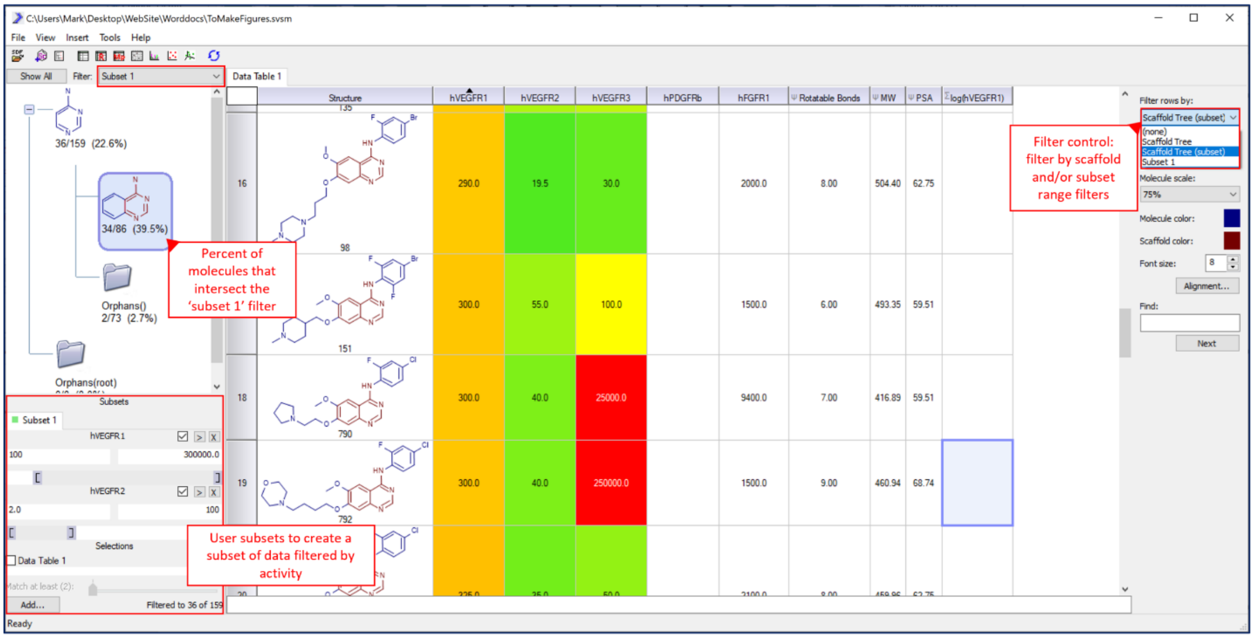
Molecules in a related series are clustered on scaffolds and their substructures. Additionally, you can Add a scaffold and double click on it to filter, color code and align molecules to the selected scaffold for easy analysis.

Under **main menu->tools** are options to:

- Calculate molecular properties including molecular weight, polar surface area, and lipophilicity constants. These values are added as columns to the right side of the Data Table
- Select a Diverse subset for screening in bioassays
- Add Columns to the Data Table
- Rearrange or Hide Columns

## Filtering Data to Create Structure-Data Subsets

A fully functional filter panel to create *subsets* of data based on textual or numeric data is located on the lower left, can be used to **subset** (or *highlight*) columns in the Data Table that a user defines as a range of properties. In the same way that scaffold subset molecules based on substructure, these range filters can subset molecules based on bioactivity or chemical properties. Note that scaffolds and filters can operate independently or additively on the spreadsheet.



hVEGFR1: 100 – 300000.0

hVEGFR2: 2.0 – 100

Percent of molecules that intersect the 'subset 1' filter

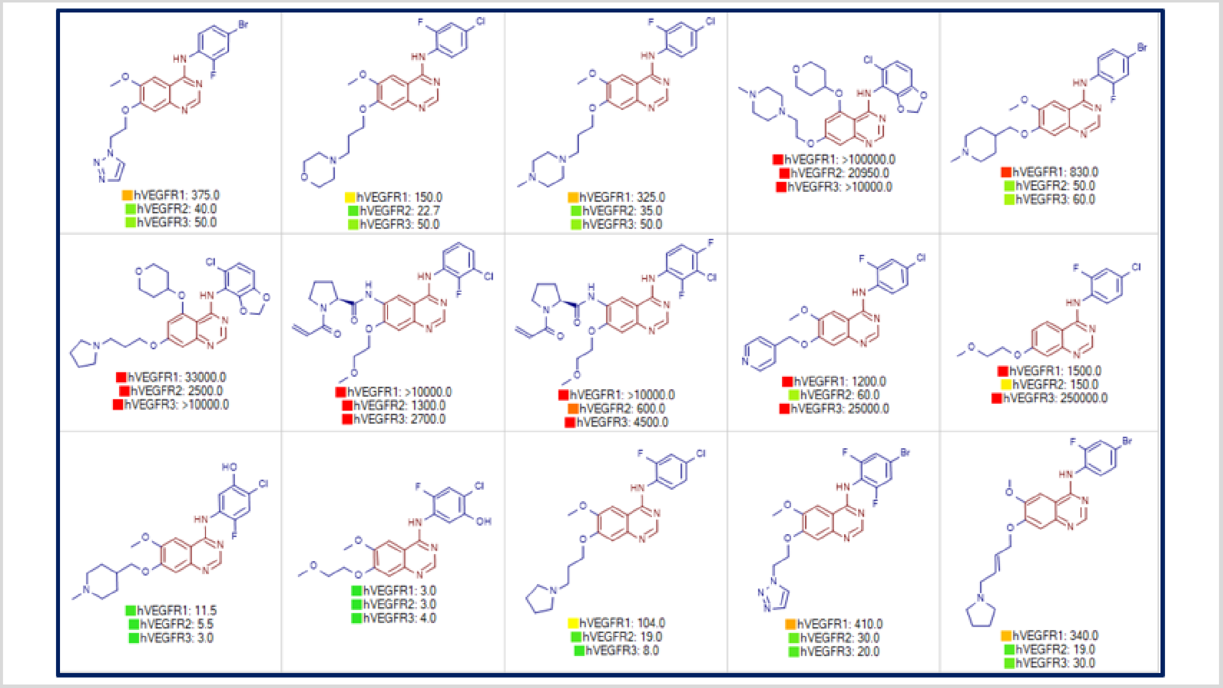
Filter control: filter by scaffold and/or subset range filters

User subsets to create a subset of data filtered by activity

| | Structure | hVEGFR1 | hVEGFR2 | hVEGFR3 | hPDGFRb | hFGFR1 | Rotatable Bonds | MW | PSA | log(hVEGFR1) |
|---|---|---|---|---|---|---|---|---|---|---|
| 16 | 98 | 290.0 | 19.5 | 30.0 | | 2000.0 | 8.00 | 504.40 | 62.75 | |
| | 151 | 300.0 | 55.0 | 100.0 | | 1500.0 | 6.00 | 493.35 | 59.51 | |
| 18 | 790 | 300.0 | 40.0 | 25000.0 | | 9400.0 | 7.00 | 416.89 | 59.51 | |
| 19 | 792 | 300.0 | 40.0 | 250000.0 | | 1500.0 | 9.00 | 460.94 | 68.74 | |

Filter data by scaffold substructure and/or data ranges to see only relevant molecules in the spreadsheet.

# II. Creating Grids of Molecules with Data



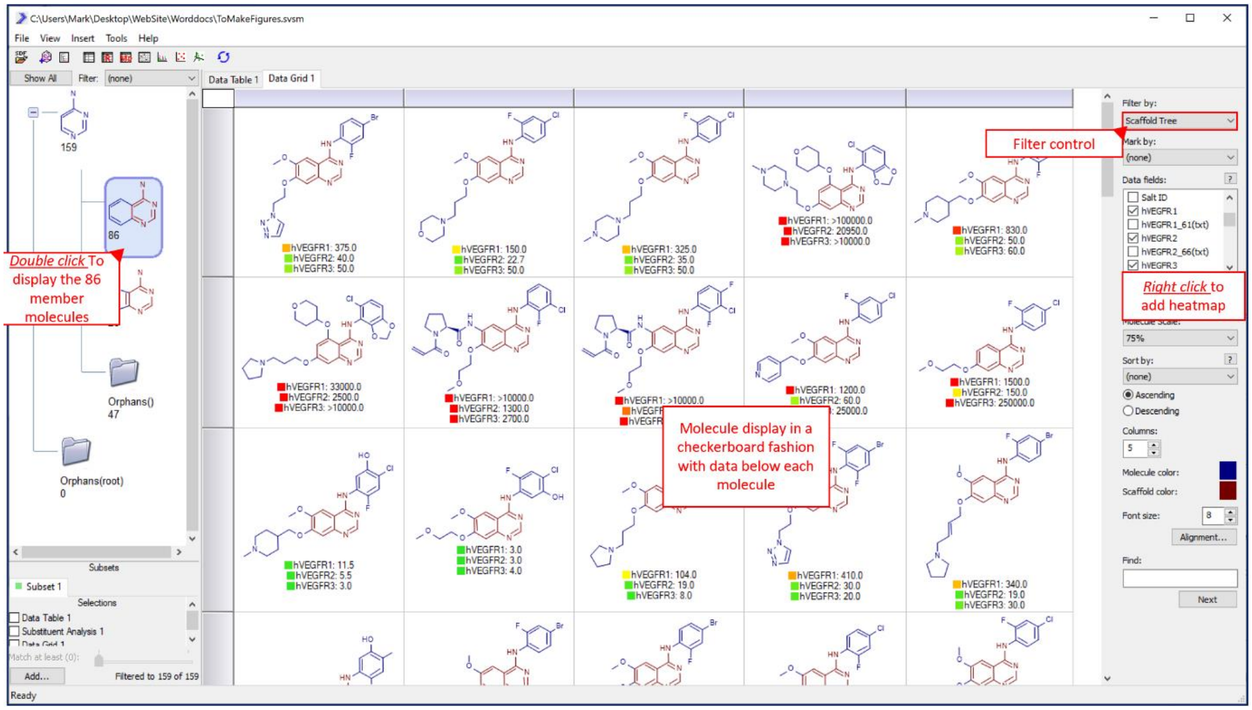
| | | | | |
|---|---|---|---|---|
| 🟧hVEGFR1: 375.0<br>🟩hVEGFR2: 40.0<br>🟩hVEGFR3: 50.0 | 🟨hVEGFR1: 150.0<br>🟩hVEGFR2: 22.7<br>🟩hVEGFR3: 50.0 | 🟧hVEGFR1: 325.0<br>🟩hVEGFR2: 35.0<br>🟩hVEGFR3: 50.0 | 🟥hVEGFR1: >100000.0<br>🟥hVEGFR2: 20950.0<br>🟥hVEGFR3: >10000.0 | 🟥hVEGFR1: 830.0<br>🟩hVEGFR2: 50.0<br>🟩hVEGFR3: 60.0 |
| 🟥hVEGFR1: 33000.0<br>🟥hVEGFR2: 2500.0<br>🟥hVEGFR3: >10000.0 | 🟥hVEGFR1: >10000.0<br>🟥hVEGFR2: 1300.0<br>🟥hVEGFR3: 2700.0 | 🟥hVEGFR1: >10000.0<br>🟧hVEGFR2: 600.0<br>🟥hVEGFR3: 4500.0 | 🟥hVEGFR1: 1200.0<br>🟩hVEGFR2: 60.0<br>🟥hVEGFR3: 25000.0 | 🟥hVEGFR1: 1500.0<br>🟨hVEGFR2: 150.0<br>🟥hVEGFR3: 250000.0 |
| 🟩hVEGFR1: 11.5<br>🟩hVEGFR2: 5.5<br>🟩hVEGFR3: 3.0 | 🟩hVEGFR1: 3.0<br>🟩hVEGFR2: 3.0<br>🟩hVEGFR3: 4.0 | 🟨hVEGFR1: 104.0<br>🟩hVEGFR2: 19.0<br>🟩hVEGFR3: 8.0 | 🟧hVEGFR1: 410.0<br>🟩hVEGFR2: 30.0<br>🟩hVEGFR3: 20.0 | 🟧hVEGFR1: 340.0<br>🟩hVEGFR2: 19.0<br>🟩hVEGFR3: 30.0 |

A molecule data grid is an efficient and compact way to visualize structure-activity data.

Molecules can be arranged in a checkerboard fashion in sequential cells from left to right and top to bottom with relevant data in columns describing each molecule.

The Data grid can be subset by a scaffold by double clicking on a scaffold, and displaying and checking the data fields desired (right panel).

The **Data grid** can be exported to Excel and Word to rapidly create reports (*right click->Data Grid tab* options).



*Double click* To display the 86 member molecules

Filter control

Molecule display in a checkerboard fashion with data below each molecule

*Right click* to add heatmap

Filter by: Scaffold Tree

Mark by: (none)

Data fields:

☐ Salt ID
☑ hVEGFR1
☐ hVEGFR1_61(txt)
☑ hVEGFR2
☐ hVEGFR2_66(txt)
☑ hVEGFR3

Molecule Scale: 75%

Sort by: (none)

● Ascending
○ Descending

Columns: 5

Molecule color: 🟦

Scaffold color: 🟥

Font size: 8

A molecular Data Grid displays molecules and data in a compact form.

In **SARvision**, select *main menu->Insert->Data grid*. To filter the molecule data grid table, double click on any scaffold substructure to apply a substructure filter. This will reorient the molecules that belong to this scaffold and color code them based on the scaffold structure. The data displayed under each molecule can be selected by a check box in the table control on the right side. *Right click* on any column in the check box control to create a heat-map icon next to the values in the table. To export this view, *right click* on the Data Grid tab.

# III. Build R-Group Tables to Analyze Structure-Activity Relationships

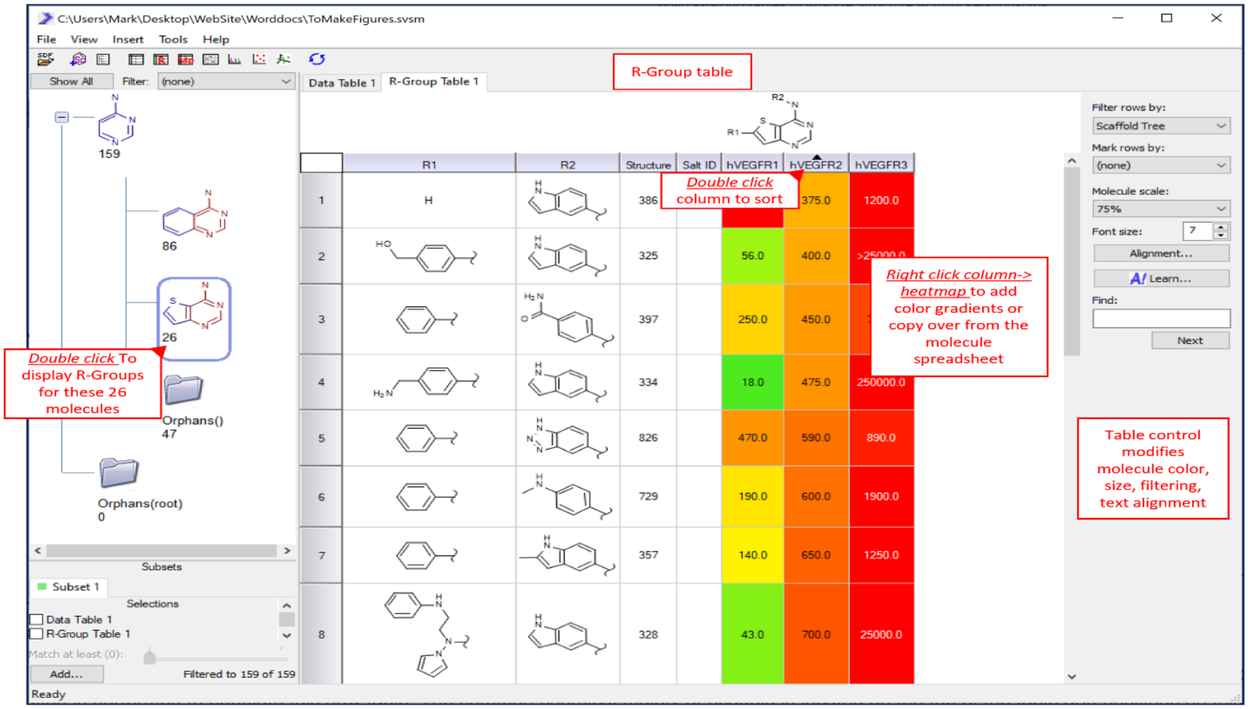
R-Group tables are built around a scaffold or chemical core common to a group of molecules under study. This type of table displays molecules broken down by substitution patterns around the common chemical moiety.

This allows for in depth analysis of Structure Activity Relationships (SAR) between chemical substituent patterns at each position with respect to different activities displayed by the molecules. Data can be derived from any experimental technique or even derived from *in silico* methods such as molecular docking.

In **SARvision**, once a set of molecules has been loaded into a molecule spreadsheet, an R-Group table can be added under the *main menu->Insert->R-Group table*. By double-clicking on a scaffold in the scaffold pane, R-Group analysis and deconvolution is performed to populate this new table.

The common chemical moiety or scaffold is located on the top of the table. Positions on this scaffold where the molecules in the set vary are sequentially numbered ***R1***, ***R2***….***Rn***. Under this scaffold are the columns labeled ***R1***, ***R2***,…***Rn*** where the R-Groups or chemical fragments associated with each molecule are shown as rows in the table.

R-Group tables are built on demand by double clicking on any scaffold.



R-Group table

Double click column to sort

*Right click column-> heatmap* to add color gradients or copy over from the molecule spreadsheet

*Double click* To display R-Groups for these 26 molecules

Table control modifies molecule color, size, filtering, text alignment

| | R1 | R2 | Structure | Salt ID | hVEGFR1 | hVEGFR2 | hVEGFR3 |
|---|---|---|---|---|---|---|---|
| 1 | H | | 386 | | | 375.0 | 1200.0 |
| 2 | | | 325 | | 56.0 | 400.0 | >25000.0 |
| 3 | | | 397 | | 250.0 | 450.0 | |
| 4 | | | 334 | | 18.0 | 475.0 | 250000.0 |
| 5 | | | 826 | | 470.0 | 590.0 | 890.0 |
| 6 | | | 729 | | 190.0 | 600.0 | 1900.0 |
| 7 | | | 357 | | 140.0 | 650.0 | 1250.0 |
| 8 | | | 328 | | 43.0 | 700.0 | 25000.0 |

To optimize the table, columns can be hidden, heat-mapped and sorted. To sort a column **double click** the header to sort the rows from ascending or descending values. **Double clicking** a column of R-Groups sorts the table by complexity of the R-Group.

To heat-map a column, **right click** on the column header and create coloring scheme in the popup user interface. The applied coloring scheme can be stoplight and/or a gradient in nature. Finally, to show/hide/change the order of the columns, **right click** on the R-Group table tab and select **Columns** option.

Predicted physicochemical properties can be calculated for the R-Group columns to compare activity to properties. Under main **Menu->Tools->Calculate molecular properties**, individual properties can be selected for calculation. These appear in the table and function as any other column in the R-Group table.

There are a number of customizable features to improve the presentation of the R-Group table.

- Any R-Group structure in the R-Group table can be substituted with a name (on R-Group: **right click->rename**).

- The user can rename (and reorder) R-Groups by simply re-drawing the scaffold **(right click->edit)** and adding R-Groups with numbers on them to the desired position on the scaffold in the Scaffold panel.

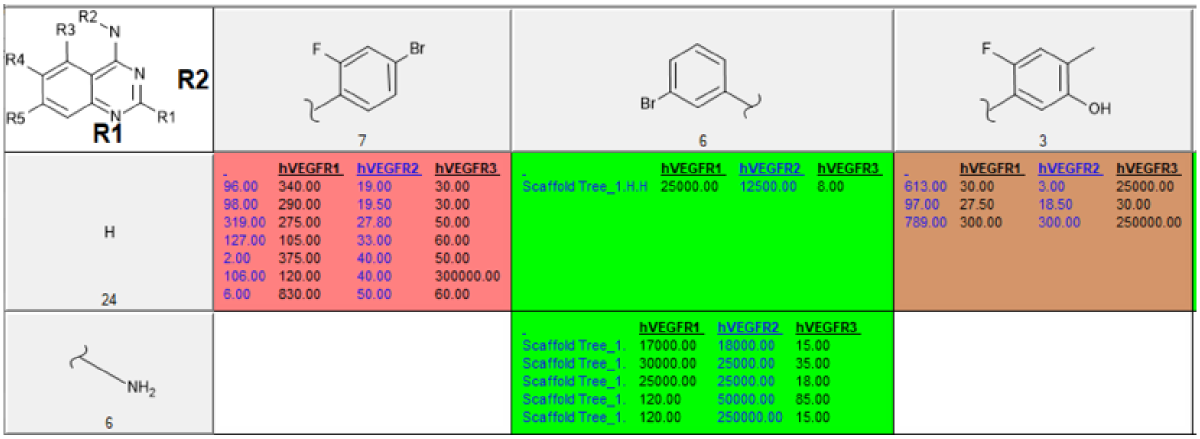
Similar to the Data Table, R-group tables can be exported by **right clicking** on the R-Group table tab and selecting **Export**, creating an Excel spreadsheet or Word file for further presentations.

# IV. Build two-way R-Group tables to analyze Structure Activity Relationships

Two-way R-Group tables are a great way to analyze and visualize R-Group substitution patterns. R-Group combinations that have not been tried are easily identified while patterns of R-Groups that lead to bioactivity can be highlighted.

Two-way tables are slices through an N-dimensional chemical space where each dimension corresponds to one R-Group position in the R-Group table (R1, R2….Rn). In the two-way R-Group tables, R-Group structures from two positions on the scaffold (Rx and Ry) are displayed in the column headers and the row header respectively. Inside each of the table's cells are a list of the molecules with data that contain these two substation patterns.

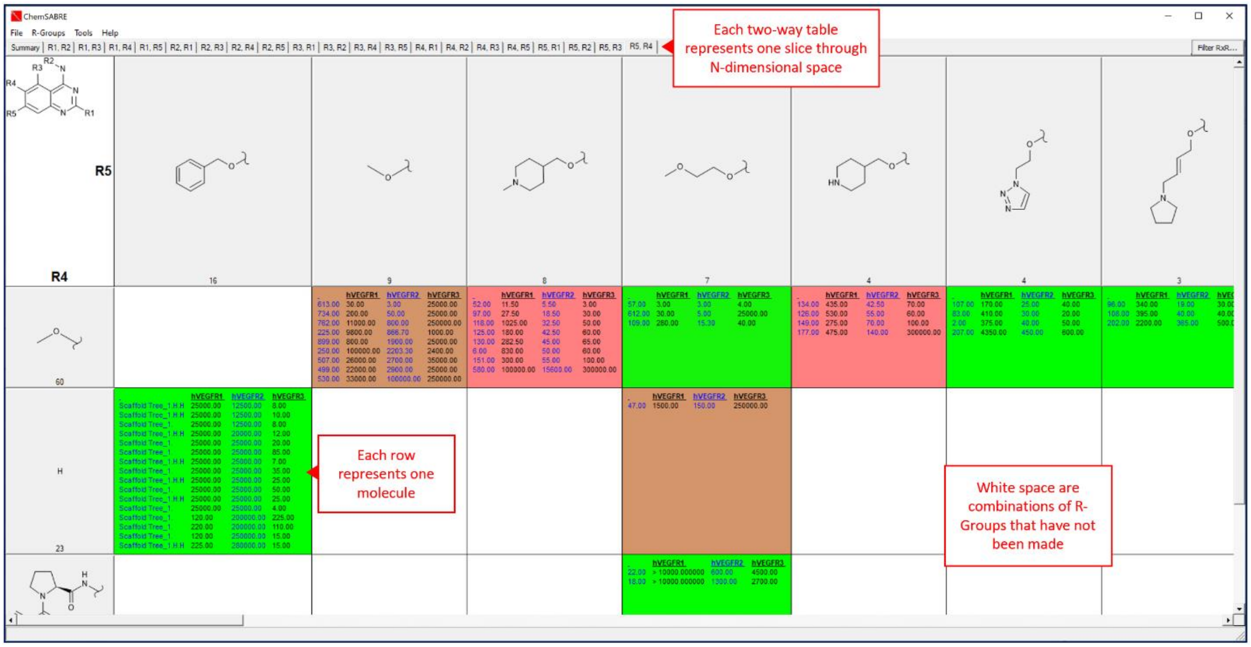
Ideally, the table is heat-mapped to make bioactive groups of molecules stand out. Using a mouse fly over, individual molecule structures on each row in the cell can be viewed. Two-way R-Group tables can be created for each pair of R-Group positions in the R-Group table. An example is shown below.



| R2<br>R1 | 7 | 6 | 3 |
|---|---|---|---|
| H<br>24 | - hVEGFR1 hVEGFR2 hVEGFR3<br>96.00 340.00 19.00 30.00<br>98.00 290.00 19.50 30.00<br>319.00 275.00 27.80 50.00<br>127.00 105.00 33.00 60.00<br>2.00 375.00 40.00 50.00<br>106.00 120.00 40.00 300000.00<br>6.00 830.00 50.00 60.00 | - hVEGFR1 hVEGFR2 hVEGFR3<br>Scaffold Tree_1.H.H 25000.00 12500.00 8.00 | - hVEGFR1 hVEGFR2 hVEGFR3<br>613.00 30.00 3.00 25000.00<br>97.00 27.50 18.50 30.00<br>789.00 300.00 300.00 250000.00 |
| 6 | | - hVEGFR1 hVEGFR2 hVEGFR3<br>Scaffold Tree_1. 17000.00 18000.00 15.00<br>Scaffold Tree_1. 30000.00 25000.00 35.00<br>Scaffold Tree_1. 25000.00 25000.00 18.00<br>Scaffold Tree_1. 120.00 50000.00 85.00<br>Scaffold Tree_1. 120.00 250000.00 15.00 | |

A two-way R-Group table highlights key features in R-Group space.

There are multiple two-way tables that can be built for a molecule set. Each represents a slice through N-dimensional R-Group space defined by a scaffold core.



Each two-way table represents one slice through N-dimensional space

| - | hVEGFR1 | hVEGFR2 | hVEGFR3 |
|---|---|---|---|
| 613.00 | 30.00 | 3.00 | 25000.00 |
| 734.00 | 200.00 | 50.00 | 25000.00 |
| 762.00 | 11000.00 | 800.00 | 250000.00 |
| 225.00 | 9800.00 | 866.70 | 1000.00 |
| 899.00 | 800.00 | 1900.00 | 25000.00 |
| 250.00 | 100000.00 | 2203.30 | 2400.00 |
| 507.00 | 26000.00 | 2700.00 | 35000.00 |
| 499.00 | 22000.00 | 2900.00 | 25000.00 |
| 530.00 | 33000.00 | 100000.00 | 250000.00 |

| - | hVEGFR1 | hVEGFR2 | hVEGFR3 |
|---|---|---|---|
| 52.00 | 11.50 | 5.50 | 3.00 |
| 97.00 | 27.50 | 18.50 | 30.00 |
| 118.00 | 1025.00 | 32.50 | 50.00 |
| 125.00 | 180.00 | 42.50 | 60.00 |
| 130.00 | 282.50 | 45.00 | 65.00 |
| 6.00 | 830.00 | 50.00 | 60.00 |
| 151.00 | 300.00 | 55.00 | 100.00 |
| 580.00 | 100000.00 | 15600.00 | 300000.00 |

| - | hVEGFR1 | hVEGFR2 | hVEGFR3 |
|---|---|---|---|
| 57.00 | 3.00 | 3.00 | 4.00 |
| 612.00 | 30.00 | 5.00 | 25000.00 |
| 109.00 | 280.00 | 15.30 | 40.00 |

| - | hVEGFR1 | hVEGFR2 | hVEGFR3 |
|---|---|---|---|
| 134.00 | 435.00 | 42.50 | 70.00 |
| 126.00 | 530.00 | 55.00 | 60.00 |
| 149.00 | 275.00 | 70.00 | 100.00 |
| 177.00 | 475.00 | 140.00 | 300000.00 |

| - | hVEGFR1 | hVEGFR2 | hVEGFR3 |
|---|---|---|---|
| 107.00 | 170.00 | 25.00 | 40.00 |
| 83.00 | 410.00 | 30.00 | 20.00 |
| 2.00 | 375.00 | 40.00 | 50.00 |
| 207.00 | 4350.00 | 450.00 | 600.00 |

| - | hVEGFR1 | hVEGFR2 | hVE |
|---|---|---|---|
| 96.00 | 340.00 | 19.00 | 30.0 |
| 108.00 | 395.00 | 40.00 | 40.0 |
| 202.00 | 2200.00 | 385.00 | 500.0 |

| - | hVEGFR1 | hVEGFR2 | hVEGFR3 |
|---|---|---|---|
| Scaffold Tree_1.H.H | 25000.00 | 12500.00 | 8.00 |
| Scaffold Tree_1.H.H | 25000.00 | 12500.00 | 10.00 |
| Scaffold Tree_1 | 25000.00 | 12500.00 | 8.00 |
| Scaffold Tree_1.H.H | 25000.00 | 20000.00 | 12.00 |
| Scaffold Tree_1 | 25000.00 | 25000.00 | 20.00 |
| Scaffold Tree_1 | 25000.00 | 25000.00 | 85.00 |
| Scaffold Tree_1.H.H | 25000.00 | 25000.00 | 7.00 |
| Scaffold Tree_1 | 25000.00 | 25000.00 | 35.00 |
| Scaffold Tree_1.H.H | 25000.00 | 25000.00 | 25.00 |
| Scaffold Tree_1 | 25000.00 | 25000.00 | 50.00 |
| Scaffold Tree_1.H.H | 25000.00 | 25000.00 | 25.00 |
| Scaffold Tree_1 | 25000.00 | 25000.00 | 4.00 |
| Scaffold Tree_1 | 120.00 | 200000.00 | 225.00 |
| Scaffold Tree_1 | 220.00 | 200000.00 | 110.00 |
| Scaffold Tree_1 | 120.00 | 250000.00 | 15.00 |
| Scaffold Tree_1.H.H | 225.00 | 280000.00 | 15.00 |

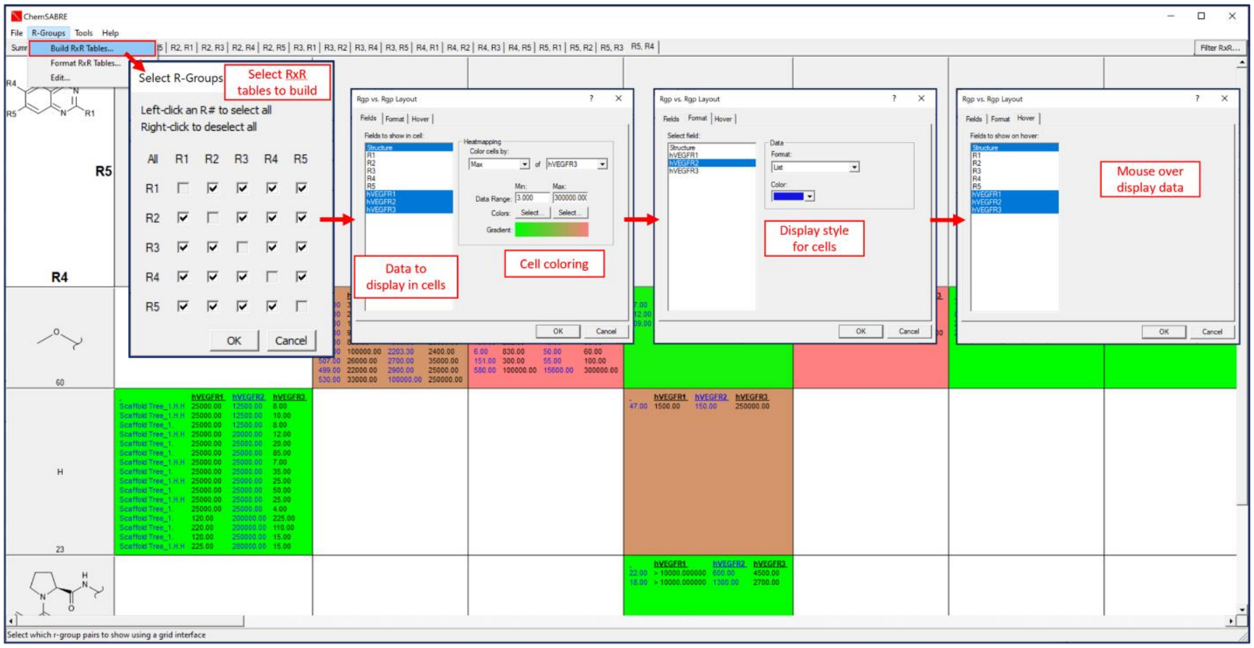
Each row represents one molecule

| - | hVEGFR1 | hVEGFR2 | hVEGFR3 |
|---|---|---|---|
| 47.00 | 1500.00 | 150.00 | 250000.00 |

White space are combinations of R-Groups that have not been made

| - | hVEGFR1 | hVEGFR2 | hVEGFR3 |
|---|---|---|---|
| 22.00 | > 10000.000000 | 600.00 | 4500.00 |
| 18.00 | > 10000.000000 | 1300.00 | 2700.00 |

To construct a set of two-way R-Group tables: **right click** on the R-Group tab in the R-Group table in **SARvision** and select **Exports->SABRE**.  Inside **SABRE**, under **main menu->R-Groups->Build RxR tables**, begin construction of all desired pairwise tables. In the first user interface, select the two-way tables to be built by selecting the check-boxes that correspond to the desired R-Group positions. In the second user interface, select 1) data to appear inside the table cells, 2) the desired heat-map coloring for the cells, 3) the format for the data and 4) the data that should appear in the mouse fly overs. Note that the two-way R-Group tables can be modified and rebuilt using this **Build RxR tables** menu option as necessary.



Select RxR tables to build

Select R-Groups

Left-click an R# to select all
Right-click to deselect all

| All | R1 | R2 | R3 | R4 | R5 |
|---|---|---|---|---|---|
| R1 | ☐ | ☑ | ☑ | ☑ | ☑ |
| R2 | ☑ | ☐ | ☑ | ☑ | ☑ |
| R3 | ☑ | ☑ | ☐ | ☑ | ☑ |
| R4 | ☑ | ☑ | ☑ | ☐ | ☑ |
| R5 | ☑ | ☑ | ☑ | ☑ | ☐ |

Data to display in cells

Cell coloring

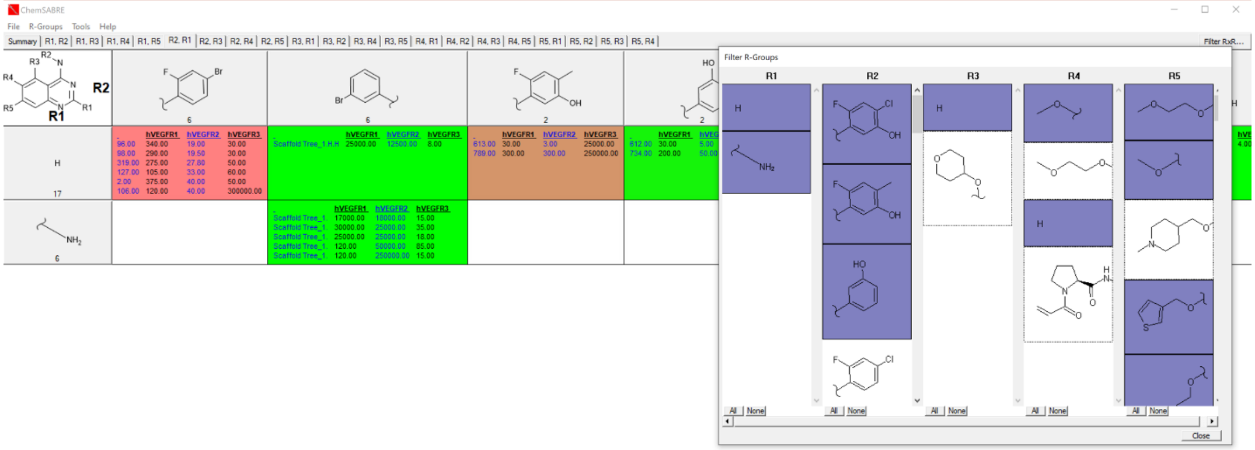
Display style for cells

Mouse over display data

Select which r-group pairs to show using a grid interface

Under the main menu are a series of dialogues that will build custom formatted two-way tables for SAR analysis.

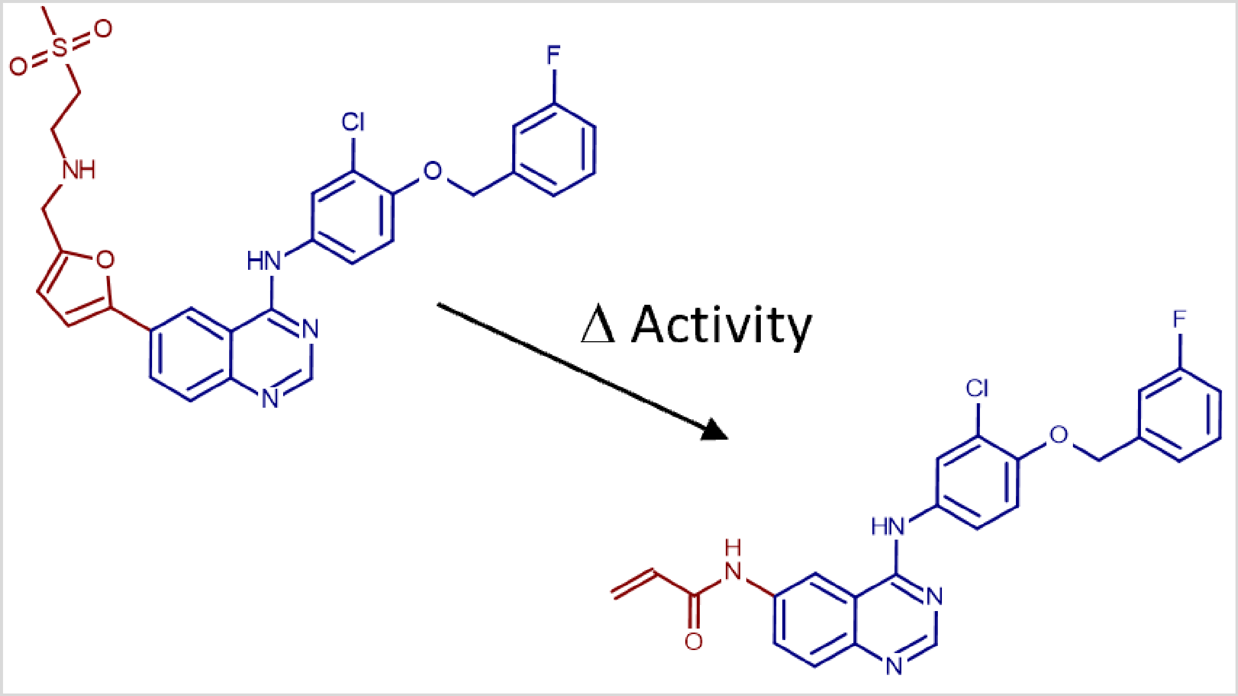
For each two-way R-Group table, there is a filter option on the right of the display. Using this filter, the molecules that populate the table can be filtered in or out of the display to help focus results onto relevant molecules. For example, in the R1 x R2 table below, the filter can be used to select for only molecules to populate the table that have desired R-Group substitutions in ***R1***, ***R2***, ***R3…Rn*** positions. Irrelevant or uninteresting substitutions can be removed from analysis using this filter.



The population of molecules in the two-way table can be filtered at any R-position to study only those with R-Groups that are relevant to the current analysis.

Note that under exports these two-way tables can be exported into Excel or Word.

# V. Explore Scaffold Centered Molecular Pairs to Understand Structure Activity Relationships

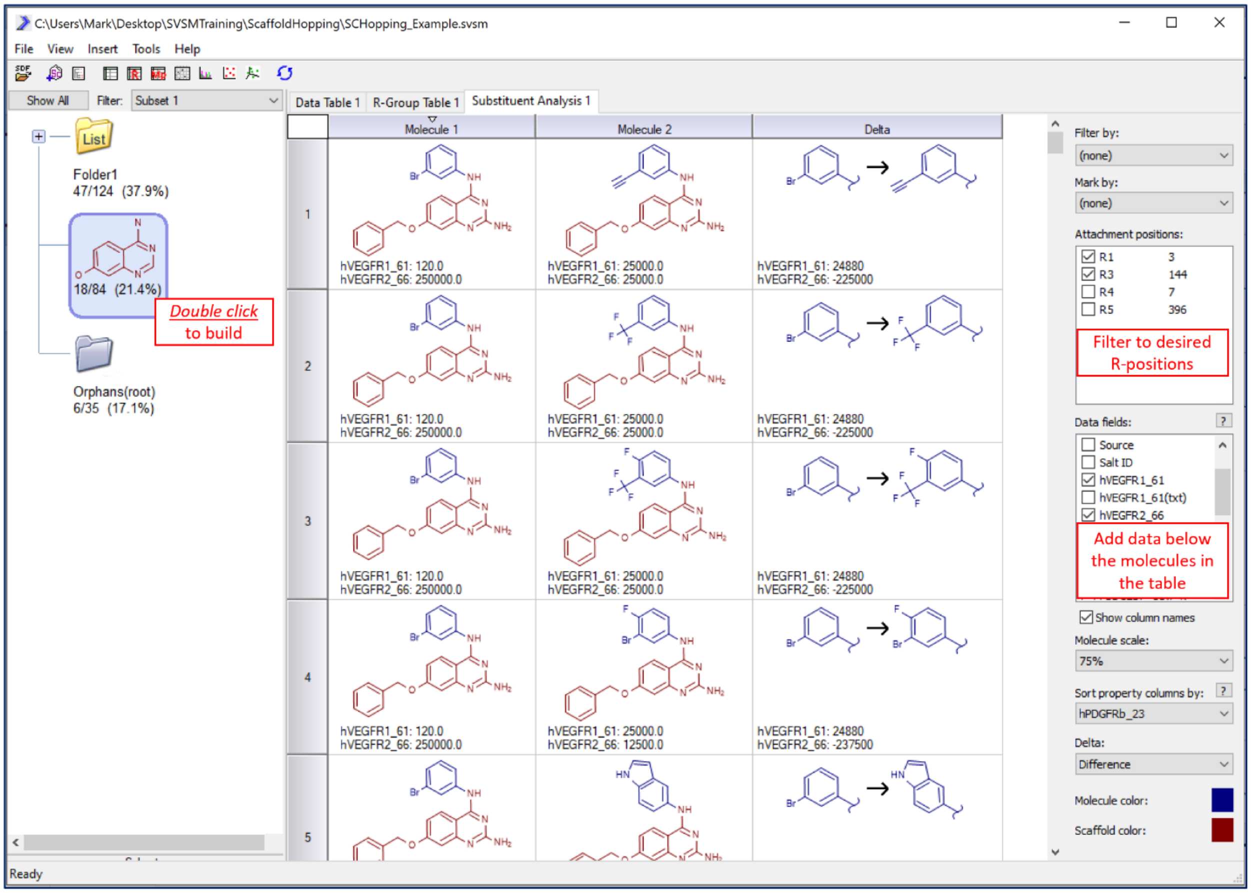


Molecular pairs are molecules that differ at only a single position. Isolating structural changes in this way useful to study Structure Activity Relationships.

A molecular pair table is built by analyzing a group of molecules to identify all pairs of molecules that are otherwise identical except for a change at a single position that can change activity data. The result is to remove all possible confounding structural variables to unequivocally connect a single structural modification to a change in the observable bioactivity data.

In **SARvision**, the molecular pair relevance problem is solved by building molecule pairs that are anchored to a scaffold of interest selected by the user. The user can build as many pairs for as many scaffolds as desired simply by adding a molecule pair table (**Main menu->Insert->Substituent analysis**) and selecting the relevant scaffolds one at a time (**double clicking**). The pair finding algorithm begins with a subset of molecules that belong to this scaffold and identifies molecular pairs that include this scaffold core.  The two pairs are shown side by side with structural differences denoted by different colors. Below each structure is observable data measured for each molecule.

The third column focuses on the change in structure (top) and data(bottom) for this molecular pair. Note that the change in data can be depicted as difference or a ratio (table control: right bottom: drop down).


C:\Users\Mark\Desktop\SVSMTraining\ScaffoldHopping\SCHopping_Example.svsm

| | Molecule 1 | Molecule 2 | Delta |
|---|---|---|---|
| 1 | hVEGFR1_61: 120.0<br>hVEGFR2_66: 250000.0 | hVEGFR1_61: 25000.0<br>hVEGFR2_66: 25000.0 | hVEGFR1_61: 24880<br>hVEGFR2_66: -225000 |
| 2 | hVEGFR1_61: 120.0<br>hVEGFR2_66: 250000.0 | hVEGFR1_61: 25000.0<br>hVEGFR2_66: 25000.0 | hVEGFR1_61: 24880<br>hVEGFR2_66: -225000 |
| 3 | hVEGFR1_61: 120.0<br>hVEGFR2_66: 250000.0 | hVEGFR1_61: 25000.0<br>hVEGFR2_66: 25000.0 | hVEGFR1_61: 24880<br>hVEGFR2_66: -225000 |
| 4 | hVEGFR1_61: 120.0<br>hVEGFR2_66: 250000.0 | hVEGFR1_61: 25000.0<br>hVEGFR2_66: 12500.0 | hVEGFR1_61: 24880<br>hVEGFR2_66: -237500 |
| 5 | | | |

Filter by: (none)

Mark by: (none)

Attachment positions:

| | |
|---|---|
| ☑ R1 | 3 |
| ☑ R3 | 144 |
| ☐ R4 | 7 |
| ☐ R5 | 396 |

Filter to desired R-positions

Data fields:

☐ Source
☐ Salt ID
☑ hVEGFR1_61
☐ hVEGFR1_61(txt)
☑ hVEGFR2_66

Add data below the molecules in the table

☑ Show column names

Molecule scale: 75%

Sort property columns by: hPDGFRb_23

Delta: Difference

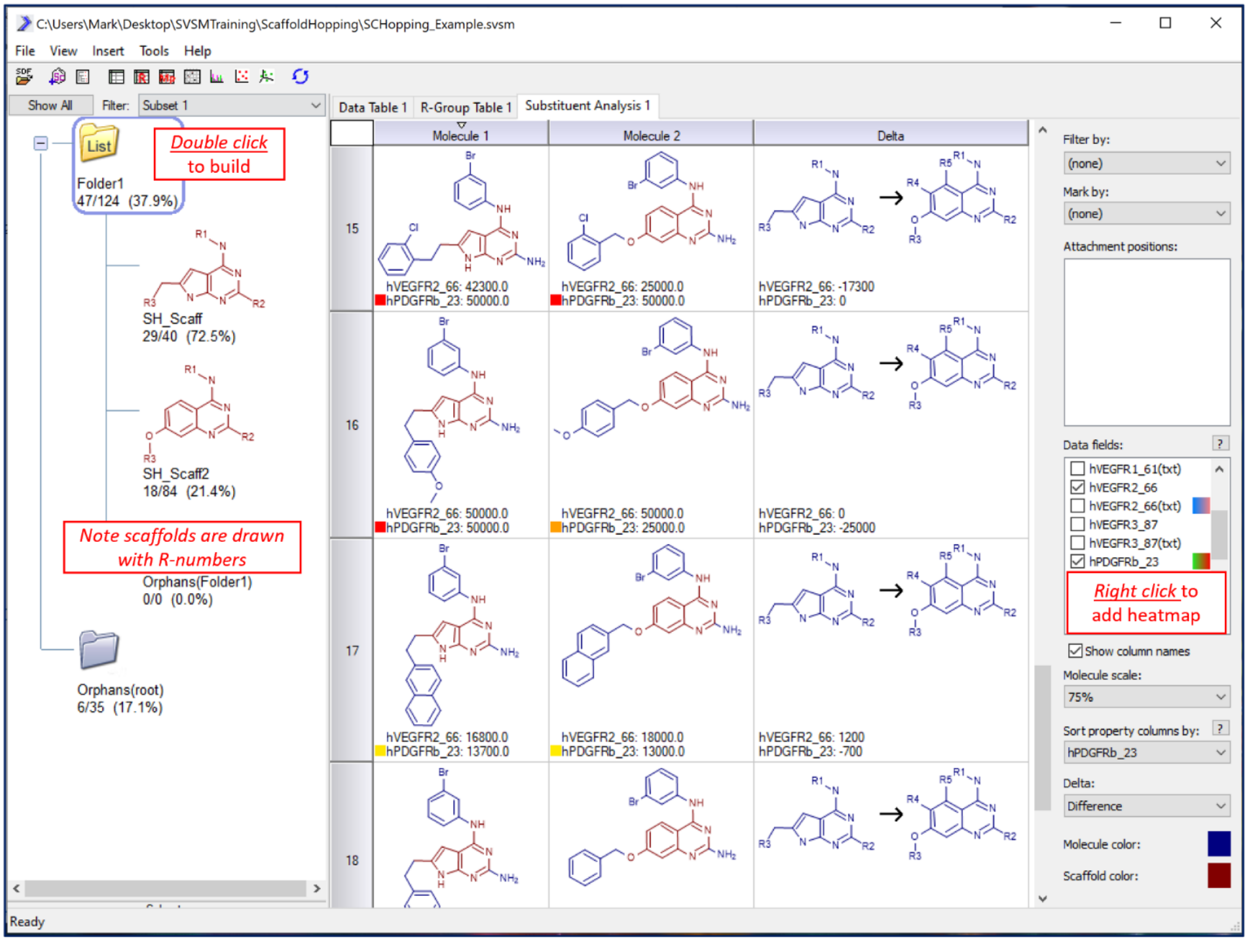
Molecule color:

Scaffold color:

**SARvision** builds molecular pair tables centered on specific scaffolds. Pairs can be easily filtered by R-position to facilitate analysis.

Often too many molecular pairs are derived for closely related sets of molecules and it is desirable to subset further by pairs that change at only specific positions. In the control panel (right) is a check box for each R-Group position (the R-Groups are the same as described in the R-Group table). Unchecking an R-Group temporarily removes it from the display allowing browsing of the data one scaffold R-position at a time. Note that molecular pairs are now filtered by a scaffold core ***and*** by position to make analyzing molecular pair changes easy to visualize.

Sometimes a series may have two closely related cores and the user may want to identify pairs of molecules that have the same R-Groups but change at the core. Placing these cores into a folder in the scaffold tree (**scaffold pane:right click->Add folder, folder:right click->Add scaffold**) and selecting the folder (**folder: double click**) creates a new molecular pair table. This is built listing only pairs where the cores listed in this folders are swapped. The table is otherwise identical to those created with a single scaffold. This can be useful for scaffold or core hopping exercises.



*Double click to build*

*Note scaffolds are drawn with R-numbers*

*Right click to add heatmap*

Filter by: (none)

Mark by: (none)

Data fields:

☐ hVEGFR1_61(txt)
☑ hVEGFR2_66
☐ hVEGFR2_66(txt)
☐ hVEGFR3_87
☐ hVEGFR3_87(txt)
☑ hPDGFRb_23

☑ Show column names

Molecule scale: 75%

Sort property columns by: hPDGFRb_23

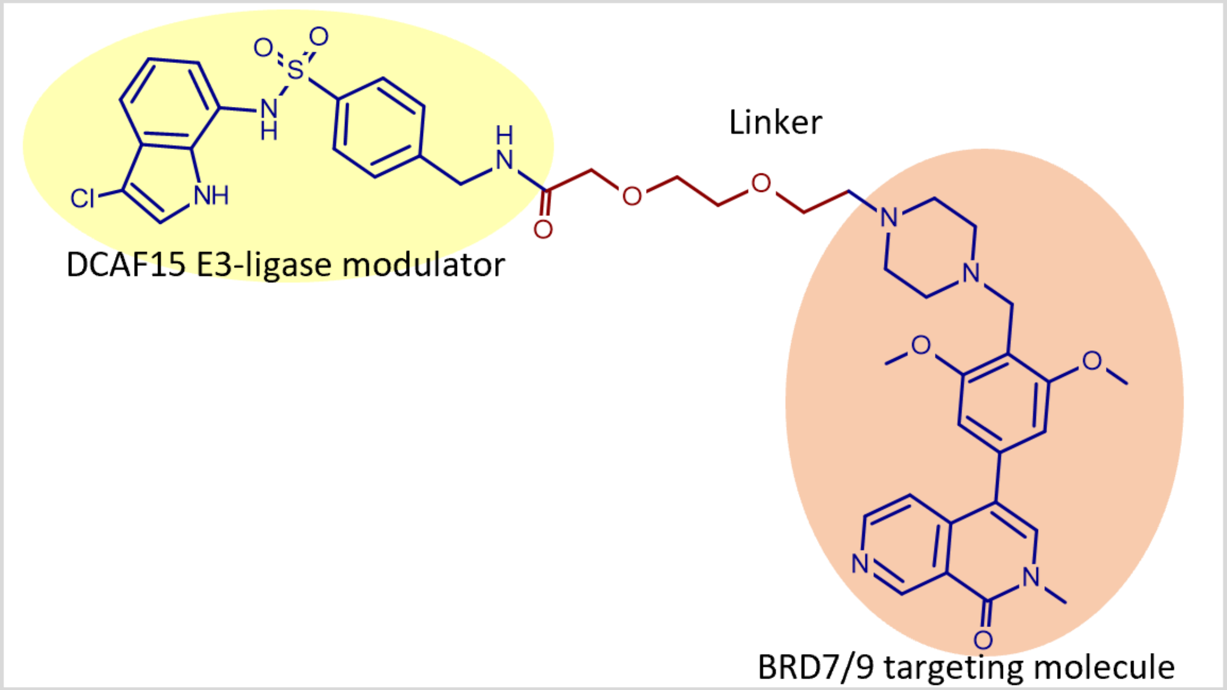
Delta: Difference

Comparing scaffold cores or core-hopping is easily performed by using folders in the scaffold tree.

# VI. Analyzing PROTAC Structure-Activity Relationships

Performing Structure-Activity analysis on bi-ligands such as PROTACS requires software that can identify the linker and break the molecule into its constituent parts-the ligase modulator-Linker-and Targeting molecule segments.

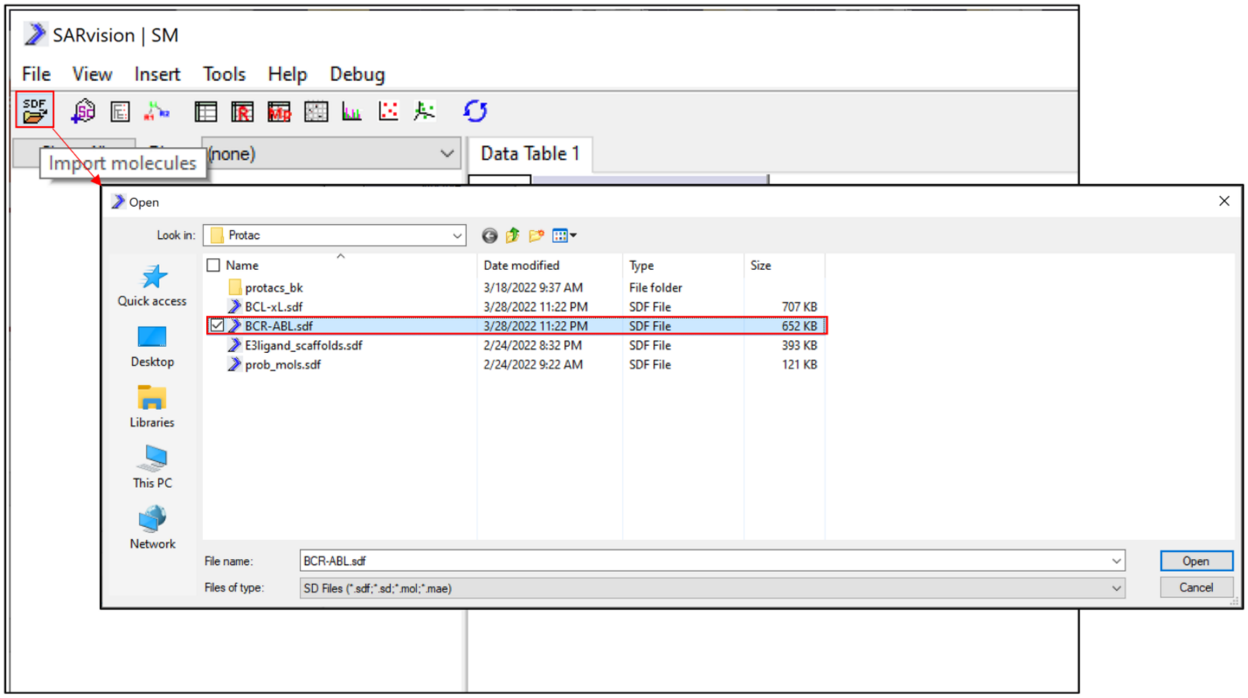
Medicinal chemists link the modulator to the targeting molecule together through a linker to achieve new and novel biological activities. PROTAC molecules best exemplify this approach where an E3 ligase modulator is linked to a targeting warhead ligand that recruits a protein target for degradation. **SARVision** allows the user to retro synthetically deconvolute these molecules into their constituent components to perform structure-activity analysis while providing insights into novel and bioactive chemical structures on a routine basis.



A PROTAC linker analysis module has been added to **SARvision** to present chemically relevant linkers and to make SAR studies of bi-ligands routine and efficient to perform. The PROTAC functionality is designed to study not only PROTACs, but also is applicable to work on any set of molecules that possess two or more ligands joined by chemical linkers.

The algorithm identifies and deconstructs the molecules into a linker and the two ligands such that they can be readily analyzed and visualized using SAR tables created by **SARvision*.***

For PROTAC molecules, the E3 ligase modulator is identified and positioned on the left side. The linker is extended horizontally between the two ligands. The remaining moiety, the targeting warhead, is oriented to the right.



Look in: Protac

| Name | Date modified | Type | Size |
|---|---|---|---|
| protacs_bk | 3/18/2022 9:37 AM | File folder | |
| BCL-xL.sdf | 3/28/2022 11:22 PM | SDF File | 707 KB |
| ☑ BCR-ABL.sdf | 3/28/2022 11:22 PM | SDF File | 652 KB |
| E3ligand_scaffolds.sdf | 2/24/2022 8:32 PM | SDF File | 393 KB |
| prob_mols.sdf | 2/24/2022 9:22 AM | SDF File | 121 KB |

File name: BCR-ABL.sdf

Files of type: SD Files (*.sdf;*.sd;*.mol;*.mae)

Molecules in SDF or SMILES format can be easily imported into **SARvision** (**file->import molecules :** ***load SDF, Smiles***). A set of molecules designed to degrade BCR-ABL can be loaded for study (download here). We are using a set obtained from PROTAC DB: Nucleic Acids Research, 2020. Doi: 10.1093/nar/gkaa807 for this example.

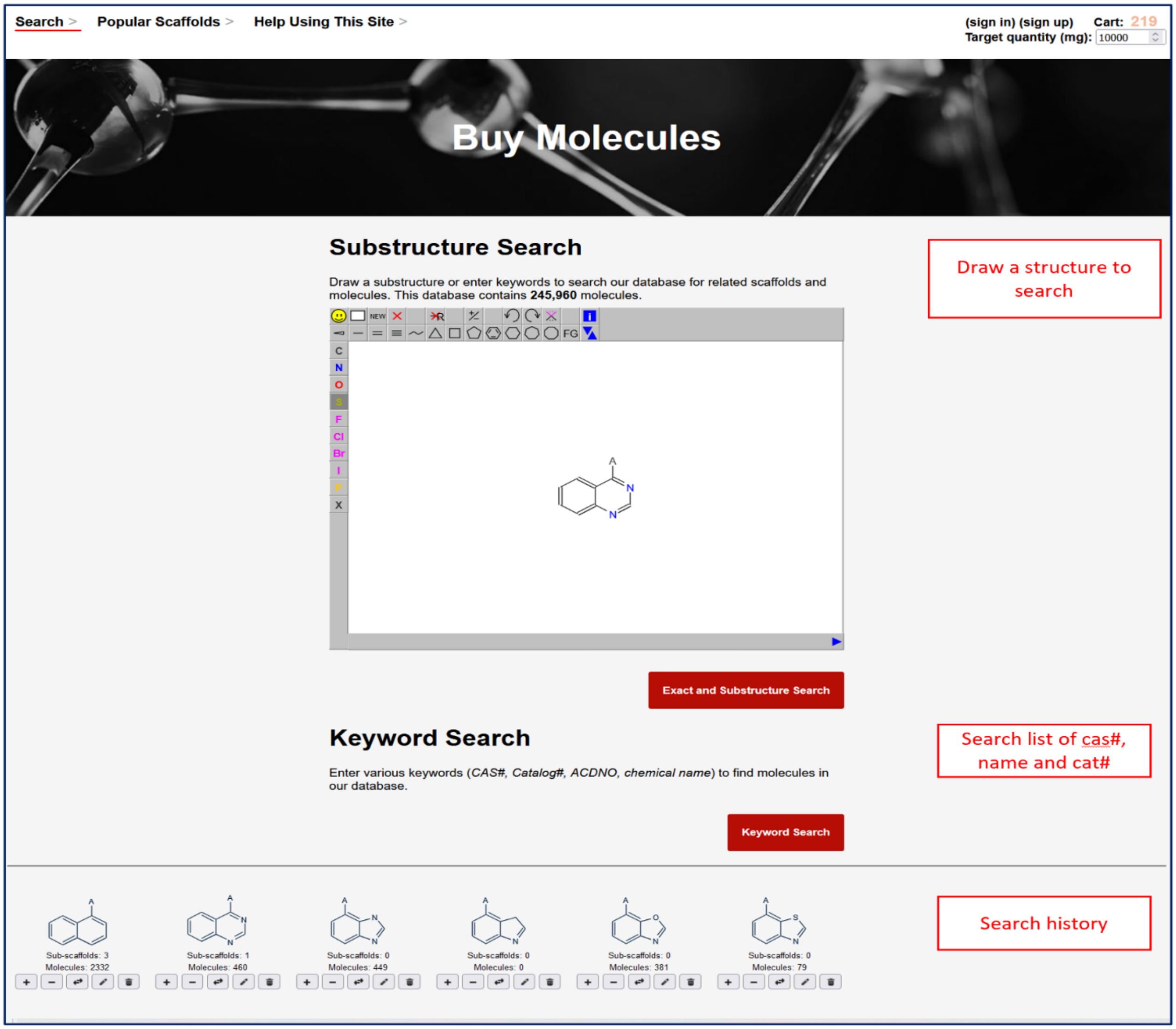
Molecules and any associated data will appear in the molecule Data Table. Once the molecules have been loaded, click the Linker analysis button to perform the analysis. The algorithm processes the molecules and populates any open tables: **Data Table**, **R-Group Table**, **Data Grid** and **Substituent Analysis** table are the most relevant. This step identifies the linker highlighting it in red (default color), reorients the molecules such that the E3 ligase ligand is on the left and the target warhead is on the right. Rows can be sorted by any data column which in turn can be heat mapped by value as shown for **DC50 (nM)** below.

# VII. Finding Chemical Reagents and Molecules from Commercial Vendors

Purchasing chemicals for organic synthesis, creating chemical series for assay, or generating chemical libraries for high throughput screening (HTS) can be a time and labor intensive process, as many manufacturers and vendors for chemicals exist, and different vendor databases can be incomplete, have compounds listed but not available, and finding the proper compounds for research can be frustrating and unproductive.

CHEMAPPS has addressed this need, and has several different proprietary chemical databases that contain vetted and qualified manufacturers and vendors of fine chemicals, pharmaceutical and agricultural chemicals, and biochemical and unique chemicals used routinely in the lab.

The 245,960 compounds found at www.buymolecules.com are actual compounds from over 14 vendors, vetted and qualified, and are ready to ship to customers worldwide.



**Substructure Search**

Draw a substructure or enter keywords to search our database for related scaffolds and molecules. This database contains **245,960** molecules.

Draw a structure to search

**Keyword Search**

Enter various keywords (*CAS#, Catalog#, ACDNO, chemical name*) to find molecules in our database.

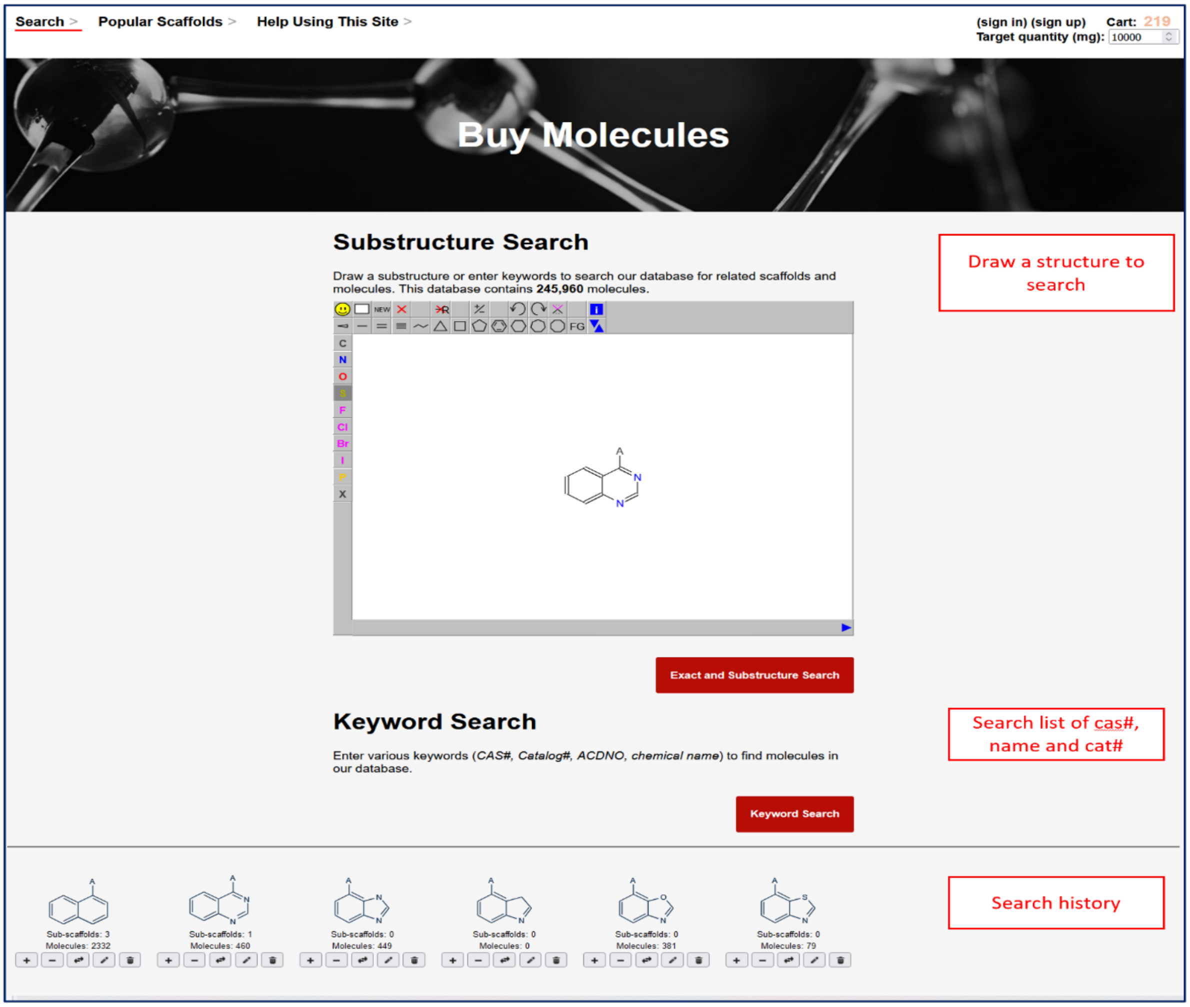
Search list of cas#, name and cat#

Search history

The CHEMAPPS search page shows results for substructure searches of vendor and other online molecules.

The **Molecule Finder** web tool organizes libraries of compounds in intuitive ways for chemists to browse and select molecules individually or in groups. The user can begin either with a **Search** using the molecule drawing tool or by browsing the **Popular Scaffolds**.

On the **Search** page, the user can draw a scaffold to perform an exact/substructure search or enter a list of names or compound names to text search.  At the bottom of the page is the users' history of previous searches that can be re-run by simply clicking on the scaffold.



**Substructure Search**

Draw a substructure or enter keywords to search our database for related scaffolds and molecules. This database contains **245,960** molecules.

**Keyword Search**

Enter various keywords (*CAS#, Catalog#, ACDNO, chemical name*) to find molecules in our database.

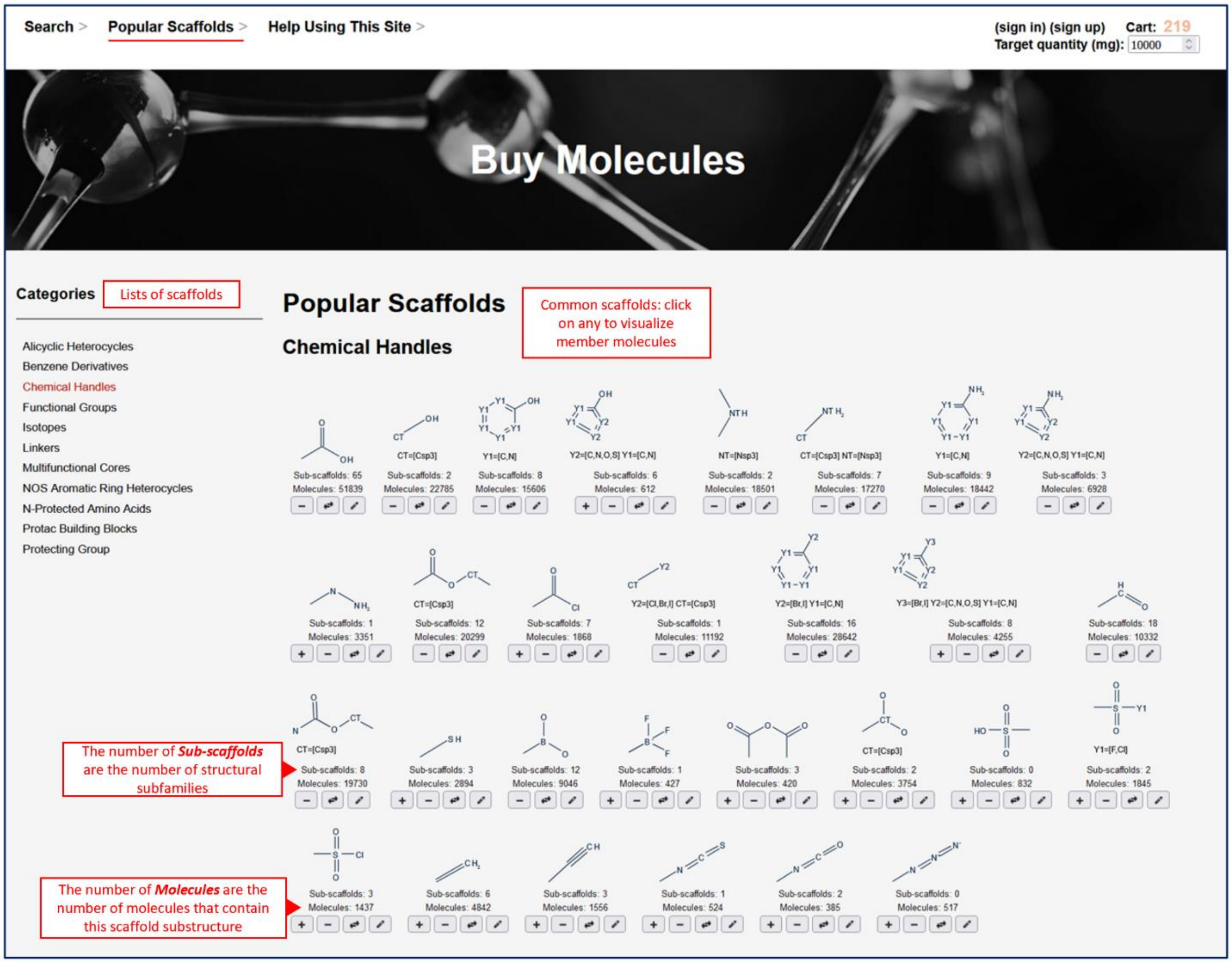
Draw a structure to search

Search list of cas#, name and cat#

Search history

Each scaffold is grouped by chemical type and can be clicked to perform a search against a library of molecules.

The **Results** page is presented in three sections: 1) The scaffold currently being searched (top). 2) Sub-scaffolds that are children (contain substructure) of the current search scaffold (middle). 3) Molecules that belong to the current scaffold family (bottom).



Common scaffolds: click on any to visualize member molecules

The number of ***Sub-scaffolds*** are the number of structural subfamilies

The number of ***Molecules*** are the number of molecules that contain this scaffold substructure

Search page showing the selected scaffold (top), child scaffold families (middle) and molecule hits (bottom). Exact substructure matches are highlighted with blue squares. Below each molecule descriptive data is display such as the cas# in this case.

Through out the program are buttons under molecules that preform operations on the cart.

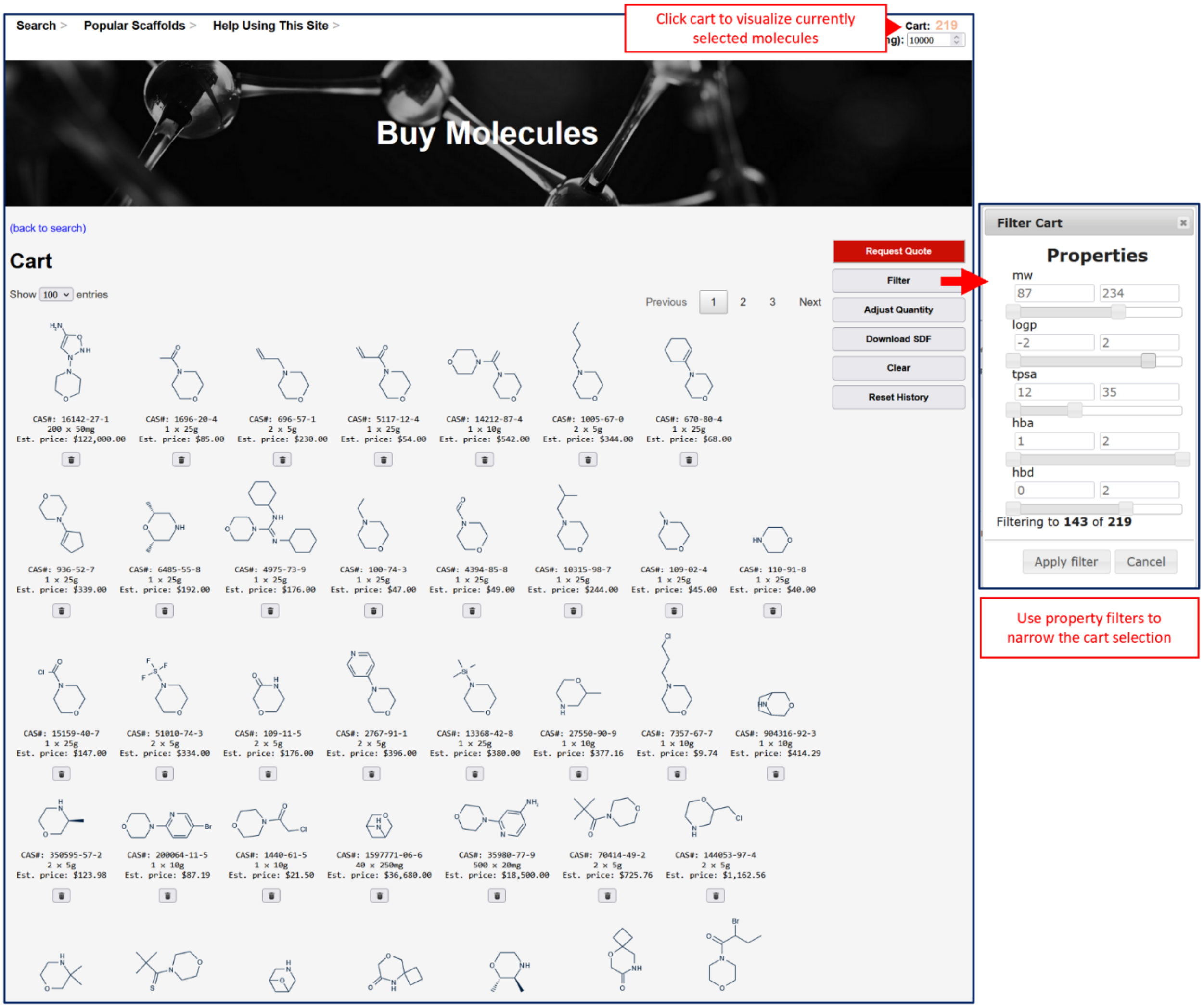
- Adds the molecules or in the case of a scaffold, the scaffold-family of molecules to the cart [+].
- Removes any molecules in the cart the belong to this scaffold-family (i.e. remove all molecules that ***have*** a carboxylic acids) [-].
- Performs a intersection with the cart with the current scaffold family. Cart will keep only molecules that are already in the cart and belong to the scaffold-family (i.e. remove all molecules that ***do not have*** a carboxylic acid) [opposing arrows].
- Clones the current search molecule and opens it in the editor to refine the search by modifying the scaffold structure [pencil].

- Removes the current molecule from the cart or remove scaffold search from history [trash].

On the **Cart** page the user can display and edit a cart of molecules. There are four main functions:

1) Request a price quote for the current set of molecules in the cart. Adding your username and email will send a quote directly to you.

2) Filter the cart by physicochemical properties. 3) Adjust the quantity of desired material (in mg, default = 10g or 10000mg). 4) Clear the Cart of all molecules.



Click cart to visualize currently selected molecules

Cart: 219
ng): 10000

**Cart**

Show 100 entries

CAS#: 16142-27-1
200 x 50mg
Est. price: $122,000.00

CAS#: 1696-20-4
1 x 25g
Est. price: $85.00

CAS#: 696-57-1
2 x 5g
Est. price: $230.00

CAS#: 5117-12-4
1 x 25g
Est. price: $54.00

CAS#: 14212-87-4
1 x 10g
Est. price: $542.00

CAS#: 1005-67-0
2 x 5g
Est. price: $344.00

CAS#: 670-80-4
1 x 25g
Est. price: $68.00

CAS#: 936-52-7
1 x 25g
Est. price: $339.00

CAS#: 6485-55-8
1 x 25g
Est. price: $192.00

CAS#: 4975-73-9
1 x 25g
Est. price: $176.00

CAS#: 100-74-3
1 x 25g
Est. price: $47.00

CAS#: 4394-85-8
1 x 25g
Est. price: $49.00

CAS#: 10315-98-7
1 x 25g
Est. price: $244.00

CAS#: 109-02-4
1 x 25g
Est. price: $45.00

CAS#: 110-91-8
1 x 25g
Est. price: $40.00

CAS#: 15159-40-7
1 x 25g
Est. price: $147.00

CAS#: 51010-74-3
2 x 5g
Est. price: $334.00

CAS#: 109-11-5
2 x 5g
Est. price: $176.00

CAS#: 2767-91-1
2 x 5g
Est. price: $396.00

CAS#: 13368-42-8
1 x 25g
Est. price: $380.00

CAS#: 27550-90-9
1 x 10g
Est. price: $377.16

CAS#: 7357-67-7
1 x 10g
Est. price: $9.74

CAS#: 904316-92-3
1 x 10g
Est. price: $414.29

CAS#: 350595-57-2
2 x 5g
Est. price: $123.98

CAS#: 200064-11-5
1 x 10g
Est. price: $87.19

CAS#: 1440-61-5
1 x 10g
Est. price: $21.50

CAS#: 1597771-06-6
40 x 250mg
Est. price: $36,680.00

CAS#: 35980-77-9
500 x 20mg
Est. price: $18,500.00

CAS#: 70414-49-2
2 x 5g
Est. price: $725.76

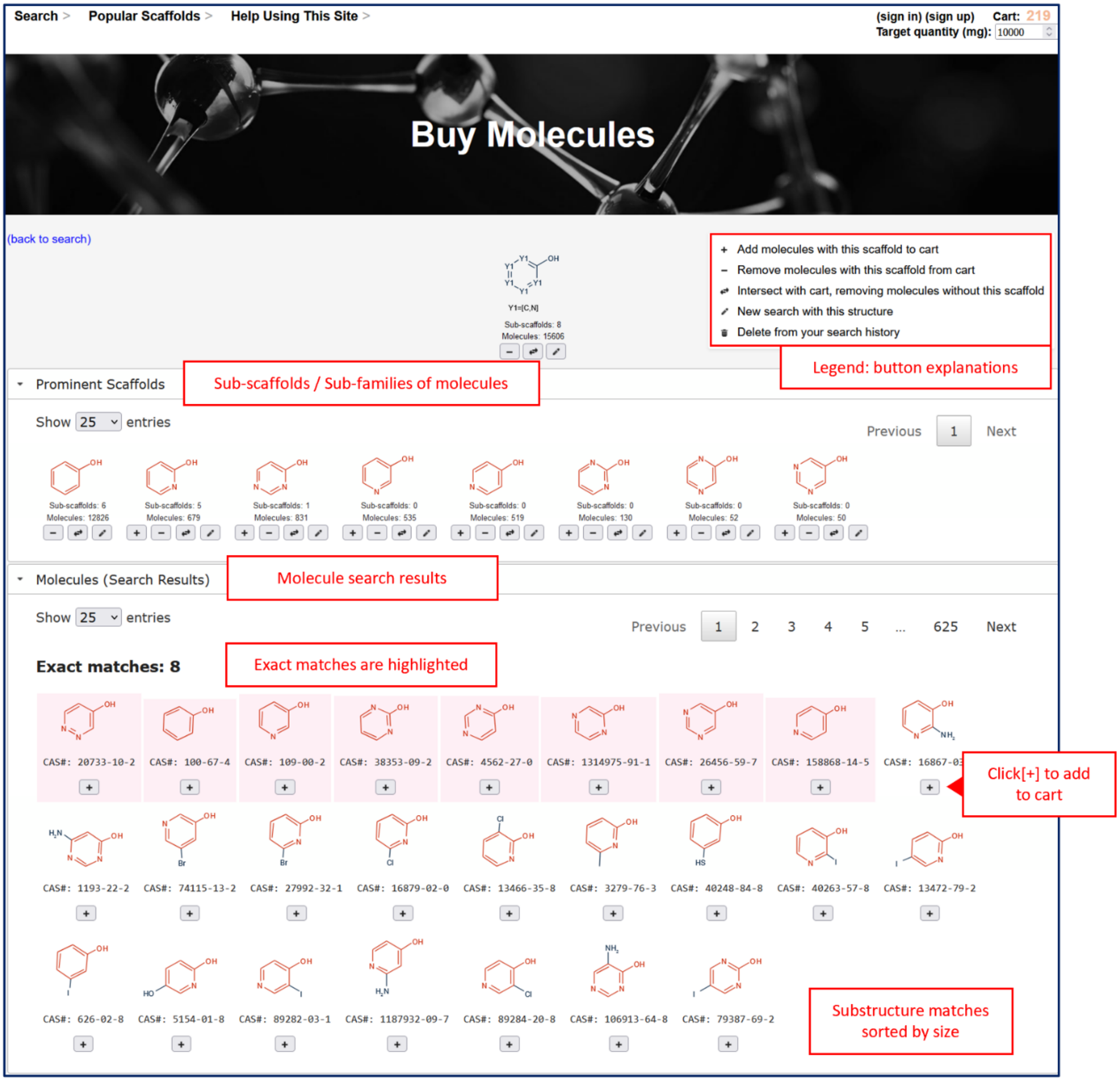
CAS#: 144053-97-4
2 x 5g
Est. price: $1,162.56

**Properties**

mw: 87 – 234
logp: -2 – 2
tpsa: 12 – 35
hba: 1 – 2
hbd: 0 – 2

Filtering to **143** of **219**

Use property filters to narrow the cart selection



- \+ Add molecules with this scaffold to cart
- − Remove molecules with this scaffold from cart
- ⇄ Intersect with cart, removing molecules without this scaffold
- ✎ New search with this structure
- 🗑 Delete from your search history

Legend: button explanations

Sub-scaffolds / Sub-families of molecules

Molecule search results

Exact matches are highlighted

Click[+] to add to cart

Substructure matches sorted by size

The cart contains collections of molecules created by the user. These display some relevant chemical information.