

SARvision|SM: Building Better Screening Libraries

You may have a chemical collection that needs to be expanded with compounds from a new vendor catalog. At the same time, you probably want to remove undesirable chemical classes commonly found in toxic compounds, and be consistent with physicochemical properties characteristic of drugs. SARvision|SM offers an alternative to traditional cheminformatics approaches that allow you to identify not just new molecules, but chemotypes that are not present in your existing collection. Work at the scaffold level offers an alternative not present in other software for library design.

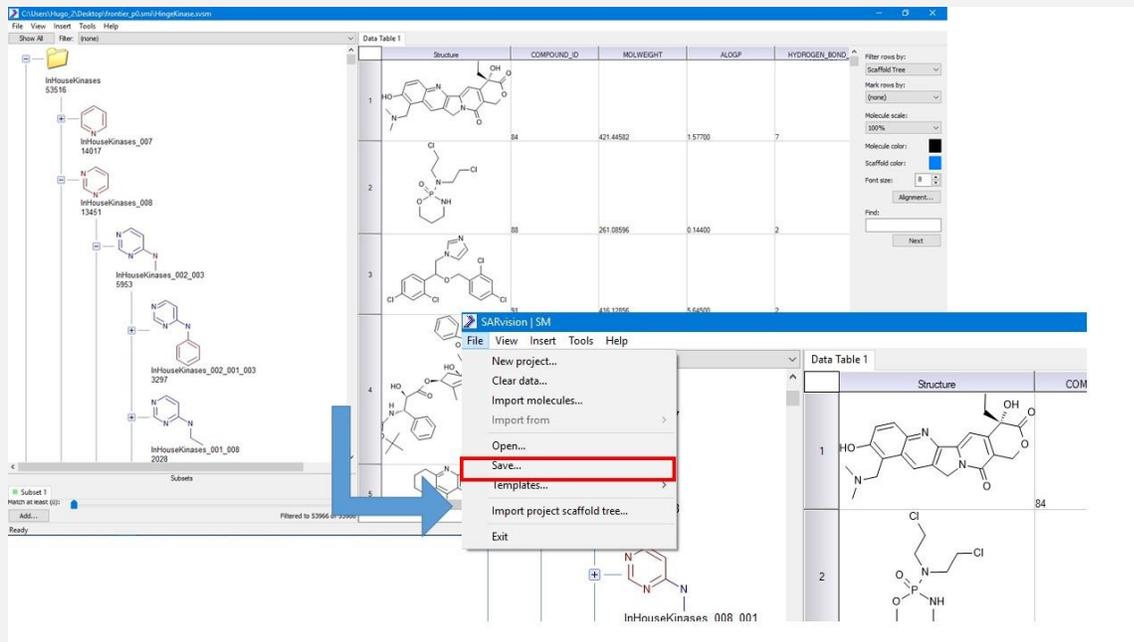
1. You probably have already a complete list of chemotypes in your in-house collection. If you do not, the first step is to generate a list of your scaffolds simply using our Automated Scaffold Perception algorithms in SARvision|SM. Just load your molecules and right click on the top left working area:

The screenshot displays the SARvision|SM software interface. A right-click context menu is open over the 'Structure' column of a 'Data Table 1', with the 'Identify scaffolds...' option selected. A blue starburst graphic with the text 'Right Click' points to this menu. The 'Identify Scaffolds' dialog box is open, showing various options for scaffold identification and display. The 'Tree building options' section includes checkboxes for 'Delete trivial scaffolds (recommended)', 'Identify chemically relevant chain scaffolds', and 'Simplify rings'. The 'Scaffold display' section has radio buttons for 'List' and 'Hierarchical tree'. The 'Scaffold pruning' section includes radio buttons for 'None', 'Number of rings', 'Number of conj. rings', 'Number of heteroatoms', 'Number of atoms', 'Population', and 'Active property match'. There are also input fields for 'Lower bound' and 'Upper bound'. A blue arrow points from the 'Identify scaffolds...' menu option to the dialog box. In the background, a data table is visible with columns for 'MPOUND_ID', 'MOLWEIGHT', 'ALOGP', and 'HYDROGEN'. The table contains three rows of data:

MPOUND_ID	MOLWEIGHT	ALOGP	HYDROGEN
421 44582		1.57700	7
261 08596		0.14400	2
416 12856		5.64500	2

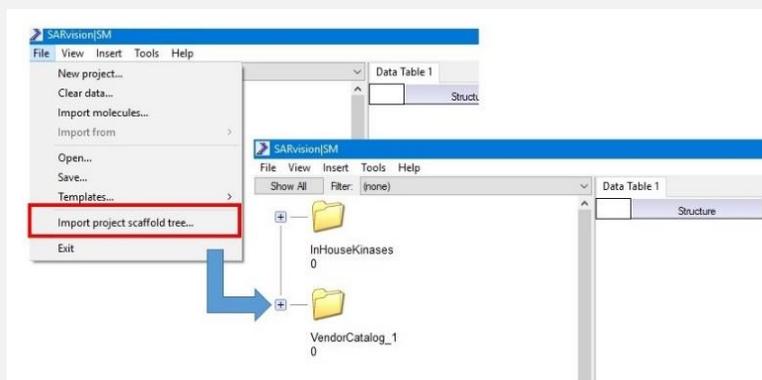
While the popup will provide you with some options, the default settings are typically sufficient. You may want to consider to what extent non cyclic scaffolds would be important in your analysis, other than that, the defaults are appropriate for this analysis. The program will automatically generate a tree of chemotypes organized in a hierarchy.

2. Save the results as a SARvision|SM file (.svsm). Now you have a SARvision|SM file that contains all your scaffolds. In this case a library of kinases has been labelled In House Kinases.

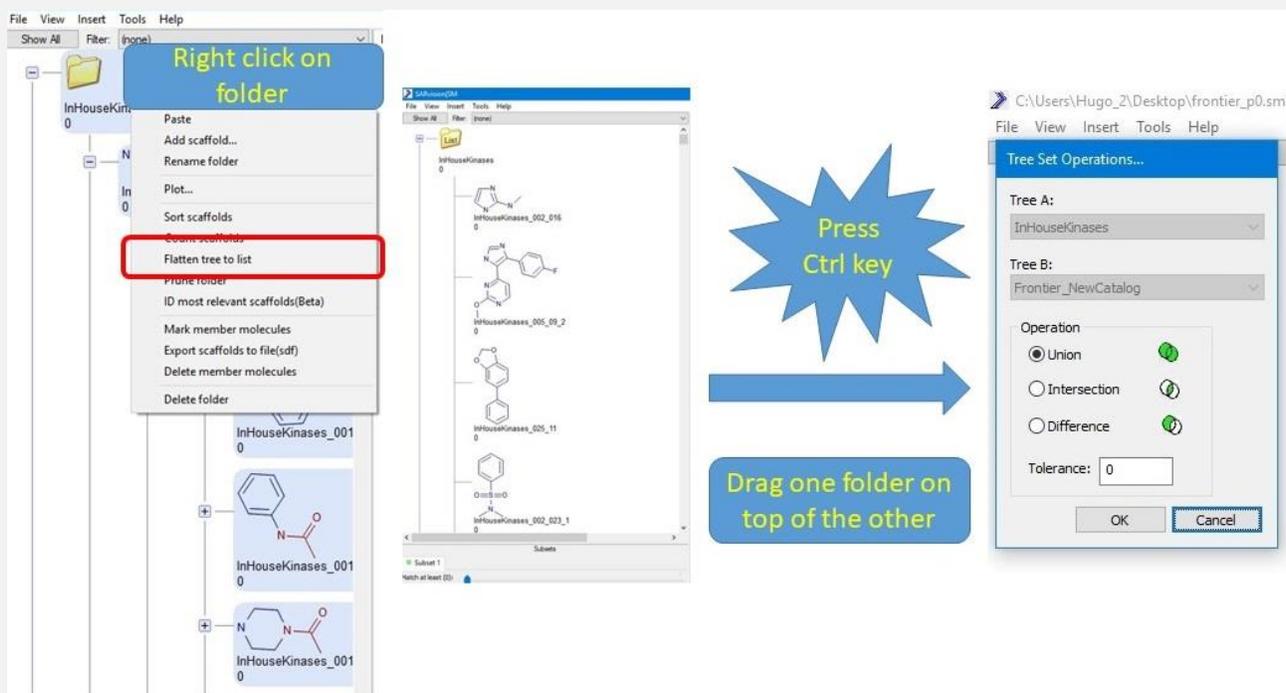


3. You want to repeat the same steps for your Vendor Catalog for example. In that way you will have two .svsm files, one includes a list of scaffolds from the vendor and the other your corporate library.

4. If you want to compare the scaffolds in each of those two libraries, you read in the scaffolds from each of those files. Open the File pull-down and select "Import Project Scaffold Tree" for each of the libraries you want to compare.



5. You will have a folder for each Scaffold Tree. In this example, the In House Kinases list of scaffolds and the list of scaffolds in a Vendor Catalog. Even though it is optional, you can first flatten the tree to a list of chemotypes, to speed up the comparisons.



Press the CTRL key and drag one of the folders on top of the other. If you do not press the CTRL key the folder you drag will be placed inside the other. When you drop one folder on top of the other while pressing control the menu shown above will popup. The menu allows you to choose if you want to create the union of the two lists, find the intersection of find the difference between them.

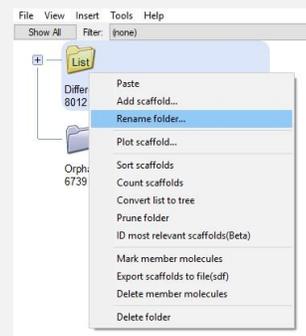
Let's select Difference. Note that in this last case which folder you drag and drop on top of the other will give you different results, the list being dragged will be subtracted from the other one. For the union or intersection the results are the same regardless of which folder is dragged.

You will end up with three folders. The two original ones and one that is labelled as the difference of one folder minus the other. In our case, the list of Vendor scaffolds minus the list of In House Kinases. You can then open the file containing the Vendor Catalog, or carry out a Scan of the vendor Catalog. The scanning and loading of molecules from the SD file selected will be carried out at the same time for all the folders.

	Structure	SMILES	ZINC ID
1		<chem>CC(C)(NH2+)[C-](H)[CO]c1c...</chem>	ZINC00000023
2		<chem>Cc1c(-O)n1c1c2cccc2N[C]ZINC00000031</chem>	ZINC00000031
3		<chem>CC(+O)Oc1cccc1C(+O)[O-]</chem>	ZINC00000053
4		<chem>C[NH+]1[C-](H)2CC[C-](H)1CZINC00000056</chem>	ZINC00000056
5		<chem>c1ccc(cc1)COC(=O)[C-](H)c2cZINC00000078</chem>	ZINC00000078
		<chem>CCOc1ccc(cc1)NC(=O)[C-](H)c2cZINC00000099</chem>	ZINC00000099

6. You can discard the two folders that are no longer needed and focus on the folder that contains the scaffolds found in the vendor catalog and not in our In House library. Of course you could have removed the two folders that related to the vendor catalog and the In House Kinase library before carrying out the scan.

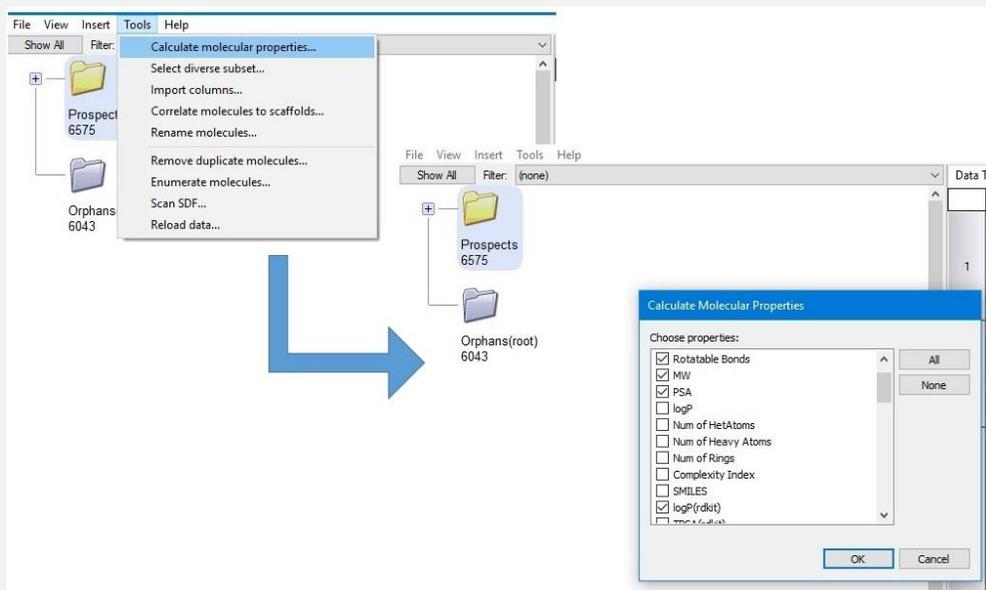
7. Rename the folder Prospects. Right click on the folders you want to delete and select Delete Folder. Right click on the one you want to rename and select Rename Folder. In the popup will enter Prospects.



8. Subtract from the list of molecules undesirable chemotypes. For example chemotypes that are frequently found in toxic compounds (toxicophores), or lists of reactive scaffolds. If you have a list of unwanted scaffolds which in this case it has been labeled BadScaffolds, you can delete all molecules that contain any of those scaffolds. For that you import the list of unwanted scaffolds. In the scaffolds area right click and select Import Scaffold.

Your unwanted scaffolds that will be added as a new folder. Right click on the folder with the unwanted chemotypes and select Delete member molecules which will eliminate all molecules that contain any of those unwanted chemotypes. Notice that the number of Prospects has been reduced.

You can take an alternative approach and subtract the toxicophores from the Prospects Scaffolds, following the approach outlined in Step 4.



9. If you want to consider drug like properties, you can now compute some simple physical properties and adopt some drug like criteria.

Under tools select calculate molecular properties and choose those that you want to use. The program includes a number of simple parameters some of which come from RDKit.

10. You can use the Subset filters to select the range of the properties you want to adopt. With the criteria chosen and the different filters adopted the number of Prospects for acquisition has been reduced to 2300 in this example.

	Structure	ZINC ID	Rotatable Bonds	MW	PSA	logP(rdkit)
733	<chem>O=C1C=CC2CN1CC2</chem>	ZINC0000023	1	217.3	40.5	1.9
734	<chem>C1CN2CCN(C1)C2</chem>	ZINC03881460	1	217.3	40.5	1.9
735	<chem>CC1(C)N(C1)C(=O)O</chem>	ZINC14808517	1	217.3	41.3	1.6
736	<chem>CC(C)N(C)C(=O)O</chem>	ZINC00057001	2	217.3	57.6	0.6
737	<chem>C1=CN=C(C1)N</chem>	ZINC02555022	5	217.3	58.6	1.9
738	<chem>Clc1ccc(cc1)NC(=O)N</chem>	ZINC03162993	1	217.7	41.6	2.7
	<chem>NC(=O)N</chem>	ZINC02006061	3	218.1	61.9	2.0

If that number of compounds is still too large there is one more capability of SARvision|SM you can use in the program to select a diverse subset based on the scaffolds present in the set.

11. You can now use the methods for Diversity Selection in SARvision|SM that are based on counting chemotypes and ensuring that a maximum number of the chemotypes identified are represented in the diverse subset. It does not resort to the use of fingerprints or other cheminformatics tools, it is purely based on the scaffolds present.

First under the Tools pulldown choose Select Diverse Subset... which will open a popup menu that will let you choose if you want your distribution to be evenly spread over all scaffolds or in a way that is proportional to the representation of the chemotype in the collection. The program will add a column that provides a ranking of the chemicals according to their scaffold based diversity. All compounds in the set are numbered.

	Structure	ZINC	Diversity Selection
1	<chem>CN1C(=O)NC2=CC=CC=C2C1=O</chem>	ZINC0000023	
2	<chem>OCC1=CC=CC=C1C(=O)N2CCNCC2</chem>	ZINC0000031	1
3	<chem>CC1(C)OC(=O)N1C2=CC=C(F)C=C2</chem>	ZINC0000056	2
4	<chem>CC(=O)S(=O)(=O)C1=CC=C(Cl)C=C1</chem>	ZINC7178827	3
5	<chem>CC1=CC=NC=C1C(=O)N2C=CC=CC2</chem>	ZINC00034175	4

If you wanted, for example, to acquire about 1000 molecules out of the remaining 2300, we have to remember that ALL compounds are numbered including those that are filtered out by the subset of properties. In that case, you have to see what is the Diverse Selection Rank for the 1000s molecule. You then add as a filter the Diversity rank.

			Diversity Column			
999		ZINC19965404	2499	3	212.3	41.6
1000		ZINC02383863	2501	1	243.3	83.9
1001		ZINC02579794	2502	2	281.2	40.9

Add as a subset criteria the Diversity Selection column and limit it from 1 to 2501 (the rank of the 1000s molecule that satisfies the other filtering criteria). You can then export the molecule set in an sd file and purchase the molecules we selected by right clicking on the tab of the table and selecting export. Among the properties that you can calculate for any molecule are SMILE strings, so if you prefer, you can calculate the SMILES strings and export those as text or in a MS Excel spreadsheet or simply the vendor ID number.

Right Click 2

			Diversity Selection	Rotatable Bonds	MW	PSA	logP
996		ZINC00023	2496	2	296.3	61.2	
997		ZINC05726372	2497	3	270.1	59.3	
998		ZINC04285937	2498	3	285.1	76.4	
999		ZINC19965404	2499	3	212.3	41.6	
1000		ZINC02383863	2501	1	243.3	83.9	

The functionality illustrated here can be combined in multiple ways depending on the criteria to be used or the processes in place. Nevertheless this ability to perform analysis at the scaffold level is chemically intuitive and unique to SARvision|SM. You can use it before or after a virtual screening exercise and as a complement to any other methods you chose to select for diversity.

Of course at any point you can remove duplicate molecules using the Remove Duplicate molecules in the Tools pulldown.

If you are interested in knowing more about this topic you can always contact us at info@altonis.com