
Subject: How to perform general R group searches in DW

Posted by [Jo W](#) on Tue, 28 Dec 2021 14:49:52 GMT

[View Forum Message](#) <> [Reply to Message](#)

In the Chemical Editor Concepts section of the online manual, the instructions state to use "?" when using specify an unusual/custom atom label, a specific isotope, an abnormal valence and/or a radical state. when drawing a structure in the structure editor.

However, when I type "R" to specify any group / atom and then do a substructure search in the dataset or via chembl, no structures are returned. So this is obviously the wrong command/label. I have searched the manual and the forums but can't find the answer.

So for example, if you wanted to search for a series of alkane alcohols in a dataset. How do you create:

R-CH₂OH as opposed to tediously searching for CH₃CH₂OH, CH₃CH₂CH₂OH

Also how do you search using two different R groups at the same time within the molecule?

e.g., R₁-CH₂O-R₂

Many thanks

Subject: Re: How to perform general R group searches in DW

Posted by [nbehrnd](#) on Tue, 11 Jan 2022 06:06:19 GMT

[View Forum Message](#) <> [Reply to Message](#)

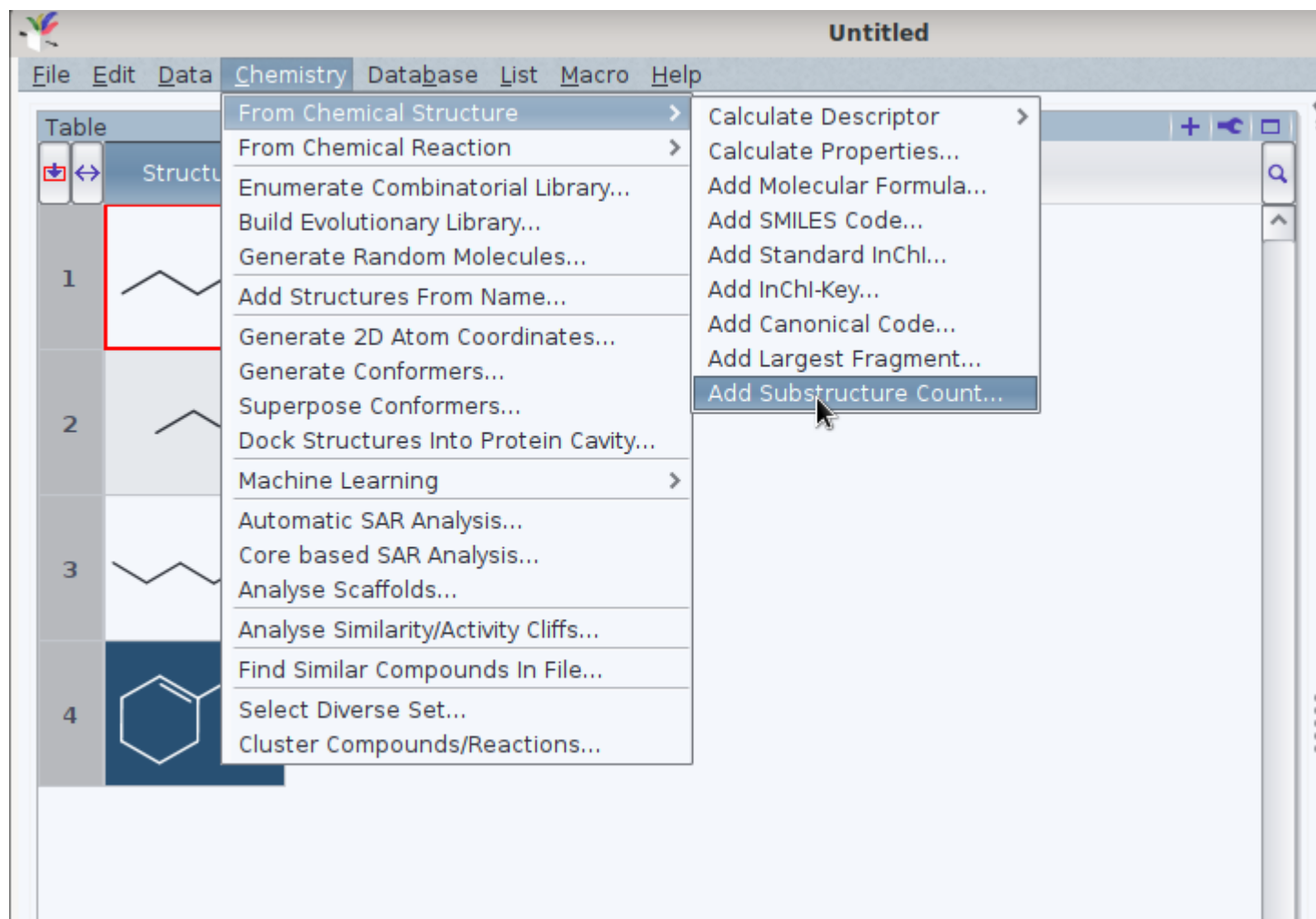
Dear JonW,

at least the nightly-builds version of DW contains a function to define a substructure query. Follow Chemistry -> From Chemical Structure -> Add Substructure Count to open a new instance of the sketcher, and decide in favour/against the criterion of overlapping substructure matches. In the matrix view, you obtain a new column with the number of occurrences of the pattern; at the right hand side, there will be a new slide ruler to filter the display. By substructure, one may define queries permitting more than one atom type on one site, too:

Norwid

File Attachments

1) [01.png](#), downloaded 650 times




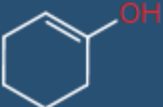


2) [02.png](#), downloaded 715 times

Untitled

File Edit Data Chemistry Database List Macro Help


Table

	Structure
1	
2	
3	
4	

Add Substructure Count

Structure column:

Column Name:

Structure: 

Include overlapping substructure matches


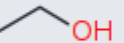

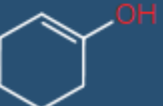
Help Cancel OK

3) [03.png](#), downloaded 691 times

Untitled

File Edit Data Chemistry Database List Macro Help

Table

	Structure	Substructure Count
1		1
2		0
3		1
4		1

Substructure Count

0 1

4) [04.png](#), downloaded 1415 times

The screenshot shows a software window titled "Untitled" with a menu bar (File, Edit, Data, Chemistry, Database, List, Macro, Help). Below the menu is a "Table" with four rows of chemical structures. The first row is highlighted with a red border. A dialog box titled "Add Substructure Count" is open over the table. The dialog has a "Structure column:" dropdown set to "Structure" and a "Column Name" text box containing "Substructure Count". Below this is a "Structure" text box containing a chemical structure of a C-O bond with the label "[C,N]". There is an unchecked checkbox labeled "Include overlapping substructure matches" and three buttons at the bottom: "Help", "Cancel", and "OK".

Subject: Re: How to perform general R group searches in DW

Posted by [thomas](#) on Tue, 11 Jan 2022 15:41:44 GMT

[View Forum Message](#) <> [Reply to Message](#)

DataWarrior uses an atom table derived from MDL's original when defining the Molfile format. This table contains R, R1 till R16, A, the amino acids, e.g. 'Ala' and a few others as pseudo atoms. Therefore, you can use these labels, but there is no logic behind them. If a ChEMBL molecule would contain an R as pseudo atom, you could find it with a substructure search.

You could use the following substructure definition to exclusively find alkyl alcohols:

This is a C-O with the carbon defined to have no pi electrons and the oxygen set to have no further non-H neighbour. Then there are to exclude groups:

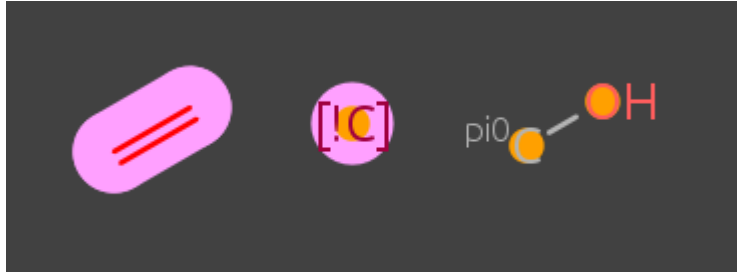
1) The '!C' is an ANY atom except carbon. This prevents any matches to molecules with non-carbon atoms in addition to the required OH.

2) A C=C double bond to prevent double bonds anywhere. Actually this is not quite correct. Better would a carbon atom with 'at least one pi-electron' as exclude group.

Two R-groups at an otherwise specified molecules could be done analogously.

File Attachments

1) [t.png](#), downloaded 1276 times



Subject: Re: How to perform general R group searches in DW

Posted by [Jo W](#) on Thu, 17 Mar 2022 21:17:07 GMT

[View Forum Message](#) <> [Reply to Message](#)

Thanks Thomas and Norwid

Many thanks for your comments and suggestions.

I didnt respond straight away because, to be honest, I didn't fully understand your answers!

Having now come back to this from time to time over the past few months its been interesting to try out those approaches, but I believe they still don't quite address what I want to do. If you or anyone else has further suggestions, please do post them here or PM me directly. When is the next version of DW going to be available and is there any new instruction videos coming out?

Subject: Re: How to perform general R group searches in DW

Posted by [nbehrnd](#) on Thu, 17 Mar 2022 21:42:19 GMT

[View Forum Message](#) <> [Reply to Message](#)

Dear Jon,

as soon as you have a running installation of DataWarrior,[1] a click on «I have read and understood the disclaimer» opens a new menu to provide access to the development versions with the .jar to substitute for improved functionality. Depending on the operating system available to you, this either is the archive for Mac/Linux,[2] or for Windows.[3] At present, the current development version was prepared Pi day, 2022-03-14.

Norwid

[1] <https://openmolecules.org//datawarrior/download.html>

[2] <https://openmolecules.org/datawarrior/dw550x.zip>

[3] <https://openmolecules.org/datawarrior/dw550win.zip>
