
Subject: SMILES-based scaffold search

Posted by [nbehrnd](#) on Tue, 25 Jun 2019 11:36:01 GMT

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

It seems worth to add to the manual pages [1] that once DataWarrior identified scaffolds, like the Murcko scaffolds, DataWarrior may determine their SMILES and add this as a new column to the array. Advantageously, it is possible to subsequently search entries in the array by setting up a new text based filter, too.

As a word of caution, results of such a text-based search depend on how the pull-down menu in front of the string to be deployed is set:

+ «equal», followed for example by C1CCCCC1 (capital characters) retrieves both entries with the scaffold of cyclohexane, as well as entries with the scaffold of benzene (for which, given the aromatization, other programs deploy the string of c1cccc1).

+ «matches regex» will retrieve scaffolds described by a string exactly as described in the string entered here. Since it is case sensitive, c1cccc1 will retrieve only entries with a scaffold of benzene; conversely, C1CCCCC1 will yield only those about the of cyclohexane. It requires backslashes should your SMILES string contain parentheses, or brackets.

DataWarrior's SMILES occasionally differ from the output provided by other programs, e.g. openbabel.

Norwid

<http://www.openmolecules.org/help/chemistry.html#ScaffoldAnalysis>

File Attachments

1) [murcko_equal.png](#), downloaded 762 times

File Edit Data Chemistry Database List Macro Help

Table	Structure 2	Murcko_scaffold	Smiles_Murcko
32			C1CCCCC1
33			C1CCCCC1
34			C1CCCCC1
35			c1ccccc1
36			c1ccccc1

Murcko scaffold_Ring Systems

Structure of structure
Murcko_scaffold_Ring Systems

Smiles Murcko

equals C1CCCCC1

Data
Structure 2

Murcko_scaffold

Selected: 1 Visible: 638 Total: 689

2) [murcko_regex_a.png](#), downloaded 754 times

File Edit Data Chemistry Database List Macro Help

Table	Structure 2	Murcko_scaffold	Smiles_Murcko
7			C1CCCCC1
8			C1CCCCC1
9			C1CCCCC1
10			C1CCCCC1
11			C1CCCCC1

Murcko scaffold_Ring Systems

Structure of structure
Murcko_scaffold_Ring Systems

Smiles_Murcko

matches regex C1CCCCC1

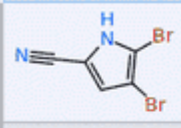

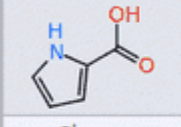

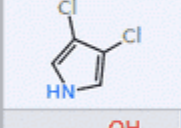

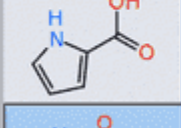

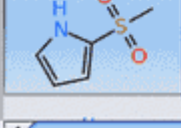

Data
Structure 2

Murcko_scaffold

Selected: 1 Visible: 34 Total: 689

3) [murcko_regex_b.png](#), downloaded 835 times

File Edit Data Chemistry Database List Macro Help

Table	Structure 2	Murcko_scaffold	Smiles_Murcko
1			<chem>c1c[nH]cc1</chem>
2			<chem>c1c[nH]cc1</chem>
3			<chem>c1c[nH]cc1</chem>
4			<chem>c1c[nH]cc1</chem>
5			<chem>c1c[nH]cc1</chem>

Selected: 1 Visible: 15 Total: 689

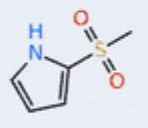
Murcko_scaffold_Ring Systems

- C(CC1)CC2C1CCCC2
- C(CC1)Cc2c1cccc2
- C1CCCC1
- c1c[nH]cc1
- c1cccc1

Smiles_Murcko

matches regex c1c[nH]cc1

Data Structure 2



Murcko_scaffold

