Subject: Calculating similarity to one query compound
Posted by PaulFinn on Wed, 22 Apr 2015 19:43:42 GMT
View Forum Message <> Reply to Message

Hi,

I would like to calculate the similarity of the compounds in a table to another compound (either a compound already in the table or one in a SD file) and create a new column with these values. I cannot find a way to do this. The chemsim function maybe should do this in syntax A, but the meaning of idcode in the help text is unclear and I have not been able to get it to work.

Regards,
Paul

Subject: Re: Calculating similarity to one query compound
Posted by thomas on Sat, 25 Apr 2015 06:34:21 GMT
View Forum Message <> Reply to Message

Hi Paul,

you are right, the chemsim() function should do it. ID-Codes are DataWarrior's internal canonical structure representations. They contain all stereo features and in case of query fragments they also contain query features.

What you need to do is to copy the idcode of a compound into the clipboard (in DataWarrior: right mouse click->Copy As->ID-Code), then open the 'Add Calculated Values...' dialog, type the formula 'chemsim(,"")', move the cursor after the opening bracket, select the descriptor from the popup (e.g. FragFp_of_Structure) and click 'Add Variable', move the cursor between the double quotes and paste in the idcode. Click OK. This should do it.

ID-Codes are a little cryptic, but they are more compact than SMILES and don't have the aromaticity issues and stereo issues of SMILES and they support MDL's concept of enhanced stereo representation, which was introduced with the molfile version 3.

Kind regards,

Thomas

Subject: Re: Calculating similarity to one query compound
Posted by rkp@23 on Mon, 01 May 2023 13:48:16 GMT
View Forum Message <> Reply to Message

I have a question about ID-Code.
Are ID-code and SMARTS inter-convertible?
For example, if I want to convert SMARTS string to an ID-code or vice-versa, are there any

method within DataWarrior that can do that?

I am asking specifically for macros and editing them outside DataWarrior.
I want to provide reaction SMARTS as SMARTS string in the macro specification , as well as reactant 0 and 1 as SMILES/SMARTS, but I see that Macros are written out with ID-codes rather than SMARTS string.

My goal is to prepare a macro outside DataWarrior: provide required reaction SMARTS, reactants for combinatorial library generation , as a SMARTS/SMILES string, may be using shell variable or python. I would like to get a Macro document, with those SMARTS string converted to ID-code, which DataWaarrior can now work on.

Has anybody tried this before?
Let me know

Subject: Re: Calculating similarity to one query compound
Posted by thomas on Mon, 22 May 2023 19:55:59 GMT
View Forum Message <> Reply to Message

SMARTS and IDCodes use slightly different concepts. Therefore, a one-to-one conversion is not possible. Not all atom/bond query features, which are available in SMARTS, do exist in OpenChemLib (i.e. DataWarrior) molecule fragments and vice versa. DataWarrior does not support recursive substructures and SMARTS don't support 'exclude groups'. Most query features are translated by DataWarrior close enough, but not necessarily in an exact way. For instance, atoms that are considered aromatic in the Daylight (SMARTS) world are not necessarily considered aromatic in OpenChemLib. And OpenChemLib distinguished aromatic and delocalized atoms. The latter is not existing in SMARTS. During the recent year DataWarrior's SMARTS parser and creator were improved to closer translate features, but it will never be 100%, because the underlying concepts don't match. Another reason is that SMARTS don't support enhanced stereo recognition.

The conversion is done in DataWarrior this way: If you paste a table that contains SMARTS (or open a text file) then DataWarrior should recognize that and a create chemical structure (reaction) column from it, which is column containing IDCodes and is tagged to contain chemical information. To create a SMILES column (or SMARTS, if a structure column contains query features) call "Chemistry->From Chemical Structure->Add SMILES Code".

Your question referred to reactions rather than molecules and I realize that there is no "Chemistry->From Chemical Reaction->Add SMILES Code". You can, however, use "Copy From Reaction->Reaction As->Reaction SMILES". I could easily add the functionality "Chemistry->From Chemical Reaction->Add SMILES Code" with the limitations discussed above, if that would help?
