
Subject: Comparing Structure Files

Posted by [biolive31!!](#) on Wed, 12 Jan 2022 12:42:25 GMT

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

Good morning I'am Olivier a new datawarior user !

I tried to use the "Find similar compound in other file" tool but it doesnt work.

I know that some molecule are similar to those in the "other file " and the tool did not find it . Any molecule is find similar. I try both comparison method , even to 70% of similarity any similar molecule is find .

Could you help me ?

Thank you

Olivier

Subject: Re: Comparing Structure Files

Posted by [nbehrnd](#) on Thu, 13 Jan 2022 22:44:47 GMT

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

Dear Oliver,

your question contains little material to replicate your finding. Thus, it is difficult to identify a plausible cause for your observation.

On the other hand, I created two tiny libraries of drug-like molecules (chemistry -> generate random molecules) where the defaults were modified slightly (instead of 1000, I opted for 100 molecules each). Subsequently, while reading set_b.dwar, I selected set_a.dwar for a comparison. With a structureFp threshold lowered to 70%, DW identified 13 molecules as similar enough. Raw data, set up of the parameters, and result files in the attachment below. Perhaps the threshold set by you was too high? The criterion «structure exact» isn't a restraint, it is meant as a constraint*); any variation in the molecule to check vs. the reference molecule and the test assigns the two as dissimilar.

*) which may be weakened in regard of stereochemistry, tautomers, or application on the largest fragment only.

Norwid

File Attachments

- 1) [set_a.dwar](#), downloaded 150 times
 - 2) [set_b.dwar](#), downloaded 141 times
 - 3) [parametres.png](#), downloaded 155 times
 - 4) [Similar Compounds.dwar](#), downloaded 140 times
 - 5) [Dissimilar Compounds.dwar](#), downloaded 137 times
-

Subject: Re: Comparing Structure Files
Posted by [biolive31!!](#) on Mon, 17 Jan 2022 13:39:30 GMT
[View Forum Message](#) <> [Reply to Message](#)

Dear Norwid,
Thank you for your help.
I tired to compare the set a and seb . I found 16 similar compounds.

My first conclusion is that DW I installed work well. :)

Now I wondering if my files are, in someway, unanalysable by similarity tool in DW.
These files are sdf files with structures.

I tried an other tool to compare molecule of my files (chemdiff) and it wasn't able to identify similar structure like DW.

Do you know what is the file type required to perform a similarity analysis ?

thank you for your help.

Olivier

Subject: Re: Comparing Structure Files
Posted by [nbehnd](#) on Mon, 24 Jan 2022 12:36:48 GMT
[View Forum Message](#) <> [Reply to Message](#)

Dear Olivier,

by Monday, 2022-01-24, Thomas identified the cause for DW's problems to deploy the criterion «Structure[Exact]» in the similarity analysis of two structure files. In the development version, the code is corrected accordingly.

Norwid

Subject: Re: Comparing Structure Files
Posted by [biolive31!!](#) on Tue, 08 Mar 2022 14:37:26 GMT
[View Forum Message](#) <> [Reply to Message](#)

Dear Norwid,
Thank you to keep inform about this problem. Where the development version is downable ?
In addition in the version I used there is an other bugg in the fine similarity function when you slect Structure[FragFp] tools.
When you are using this tool to find similar compound in an other SDF file it work well only when you are not selecting colomuns to be copied into the file. When you select a columnn to copy into the file the similarity tool doesn't work.

Thank you for your help

Olivier

Subject: Re: Comparing Structure Files

Posted by [nbehrnd](#) on Tue, 08 Mar 2022 22:06:30 GMT

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

Dear Olivier,

after accepting the disclaimer on the download page[1], an additional text box is visible with links for either an installation in Linux and Mac,[2] or for the Windows operating system.[3] These archives contain only the .jar of ongoing development of DataWarrior you substitute in an already existing installation.

So far, DataWarrior's identification of similar/identical molecules in two sets of molecules worked best for me be when starting with two .dwar containing only the structures of either set without any further computed data. Then, there was no ambiguity in selecting the column in question when running an instance with one set and reaching out for the other set to compare with by Chemistry -> Find Similar Compounds In File. It is dissimilar to a copy-paste of structures of one .dwar into an other .dwar; in parlance of DataWarrior append data (without header).

Norwid

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

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

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

Subject: Re: Comparing Structure Files

Posted by [biolive31!!](#) on Wed, 09 Mar 2022 21:52:09 GMT

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

Norwid,

thank you I'am going to install the ongoing dvpt version.
I will also try to test the similar tool with the .dwar file.

I keep you in touch.

Thank you for your help.

olivier

Subject: Re: Comparing Structure Files

Posted by [biolive31!!](#) on Thu, 10 Mar 2022 09:58:46 GMT

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

Dear Norwid,

i'am sorry but the exact similar tool still not working with the .jar file I replaced in the Datawarrior windows programm file and the comparison with the simple (two column file) .dwar file for comparison.

Olivier

Subject: Re: Comparing Structure Files
Posted by [nbehrnd](#) on Thu, 10 Mar 2022 18:37:55 GMT
[View Forum Message](#) <> [Reply to Message](#)

Dear Olivier,

the comparison by «Structure[Exact]» may be more strict (think of identity) than by «Structure[FragFp]» (think of similarity), though DW allows both may be attenuated/adjusted. But set_A and set_B mentioned yield an overlap only by similarity if the threshold is less than or equal to 0.84.

Norwid

Subject: Re: Comparing Structure Files
Posted by [thomas](#) on Tue, 05 Apr 2022 14:47:59 GMT
[View Forum Message](#) <> [Reply to Message](#)

Dear Olivier and Norvid,

I am sorry that it took such a long time for me to realize the issue here. When comparing an open window against an SD-file and if columns shall be copied from the SD-file into the open window, then DataWarrior used to parse the SD-file to the end in order to create a list of existing columns. When then starting comparing molecules, it didn't find any molecules, because it was already at the end of the SD-file and, therefore, it never found matching molecules in this case. This should now be fixed in the current dev version.

Thomas
