Subject: 'Search ChEMBL Database' function
Posted by kerryfowler on Mon, 27 Aug 2018 19:46:59 GMT
View Forum Message <> Reply to Message

DataWarrior is the tool I've wanted for decades, so thanks to all the developers!

In version 4.7.2 the "Search ChEMBL Database..." function works for

similar structure to (sometimes)
equal structure to
stereo isomers of
tautomers of

but not for "superstructures of" which returns the message "Error in task 'Search ChEMBL Database': Your query did not retrieve any records."

The similarity search sometimes returns an error until the similarity threshold is lowered. For example, similarity search for dopamine as query fails until similarity is decreased from 90% to %89. Substructure query fails for dopamine.

Subject: Re: 'Search ChEMBL Database' function
Posted by thomas on Tue, 28 Aug 2018 19:10:51 GMT
View Forum Message <> Reply to Message

Thank you very much for the hint. I wasn't aware of the problem, which was caused by a recent server
update that now uses 64-bit descriptors rather than 32 bit ones in order to accelerate the search. Unfortunately this wasn't consistently done and the pre-screening of candidates for the substructure search
didn't work because it expected the wrong type of descriptors.
I updated the server and the problem disappeared. There is still a more cosmetic problem. Large result sets
as dopamine as substructure (which retrieves 207'075 rows) are retrieved, but cause an error when retrieving
retrieving the respective assay descriptions (about 30'000). The error message is inconclusive. After closing
the error dialog DataWarrior opens a window with the 207075 results nonetheless.

Whether I can remove the remaining problem with another server update or whether is comes with the next DataWarrior
update I still need to find out.

Thanks again and best wishes,

Thomas

Subject: Re: 'Search ChEMBL Database' function
Posted by thomas on Tue, 28 Aug 2018 21:14:43 GMT
View Forum Message <> Reply to Message

I found the issue: When the results contain more than about 3000 distinct assay references, the request to
get the assay descriptions exceeds the maximum allowed size for GET requests on the Apache web-server.
When I use a POST request, then there is no error message. Since this is a DataWarrior client issues,
the next DataWarrior update will fix this.

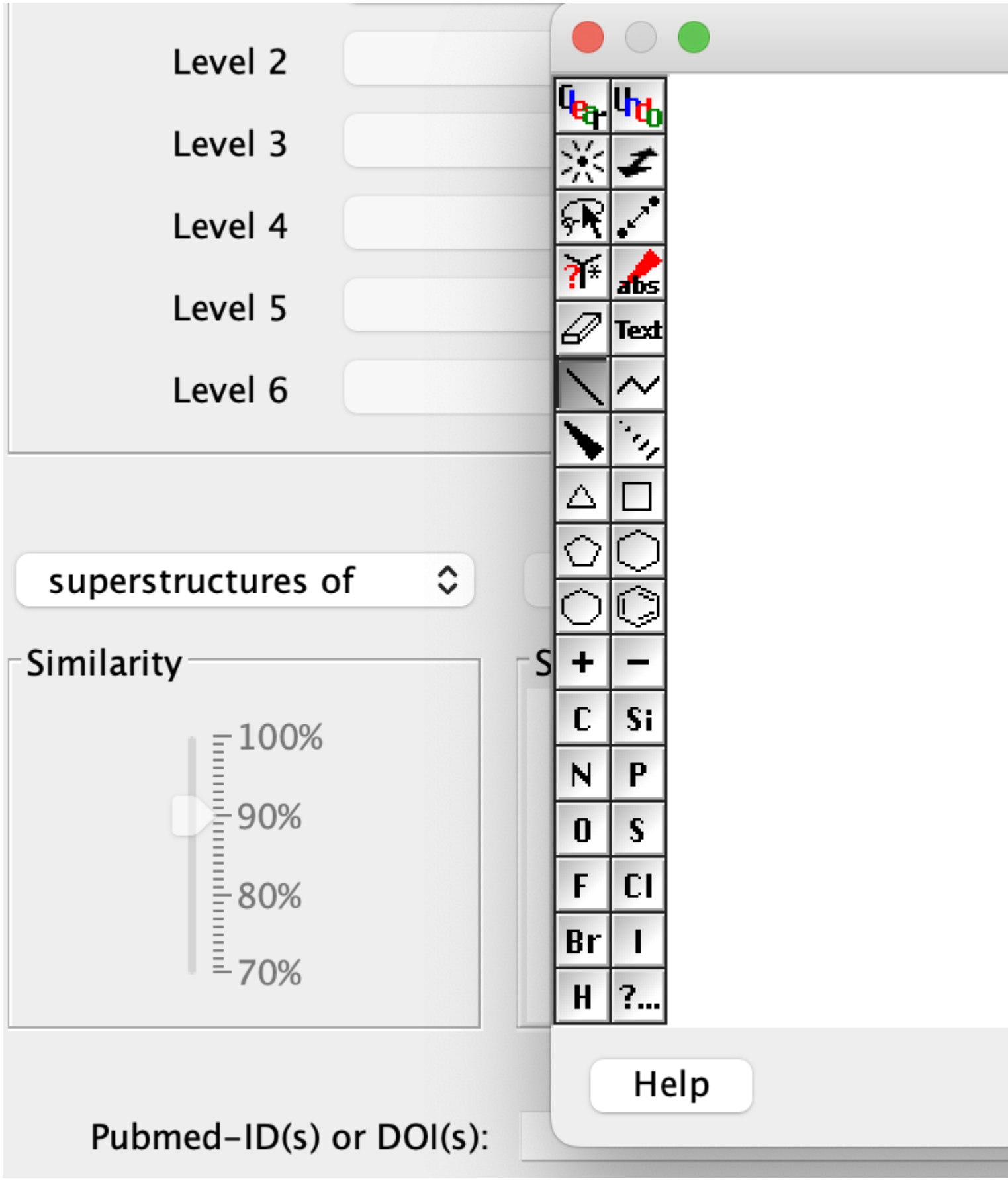
Thomas

Subject: Re: 'Search ChEMBL Database' function
Posted by rkp@23 on Mon, 01 May 2023 14:45:40 GMT
View Forum Message <> Reply to Message

I am having a similar issue,
I am trying to query organohalides on ChemBL and searching using the "superstructures of" filter.
The search did not yield any result.
I have attached a screenshot of the query I used.

Is there any issue with the ChemBL server search in the version 5.5.0?

File Attachments
1) Screen Shot 2023-05-01 at 10.46.38 AM.png, downloaded 844 times


Level 2
Level 3
Level 4
Level 5
Level 6
Clear
Undo
abs
Text
superstructures of
Similarity
100%
90%
80%
70%
C
Si
N
P
O
S
F
Cl
Br
I
H
?...
Help
Pubmed-ID(s) or DOI(s):

Subject: Re: 'Search ChEMBL Database' function
Posted by thomas on Sat, 27 May 2023 16:01:01 GMT
View Forum Message <> Reply to Message

Assuming that your '[3]' atom list contains Cl,Br,I I could confirm that no rows are returned on the current chembl server with the version 32 database. It also shows a message that your search is not specific enough and would return too many records. If you use the same query, but use iodine instead of all three halogene atoms, then 101'394 rows are returned containing 10'531 distinct structures. Currently, the server limits results to 50'000 distinct structures for sub-structure queries and to 100'000 structures for other structure searches. The idea of the limit is a historic one to make sure that a result download is not taking forever and both, client and server resources are fit to handle the result. For the next server version I will increase these limits by a factor of 2.