Subject: extension of substructure definition
Posted by nbehrnd on Thu, 24 Aug 2023 09:33:37 GMT
View Forum Message <> Reply to Message

Dear Thomas,

I would like to suggest to extend the sketcher's functionality to define motifs.

While editing a list of motifs eventually to be used as a reference set for the evolutionary library generation, I noticed the short cut `Ctrl + double left mouse click` to define one site to be e.g., either a nitrogen, or an oxygen atom currently works only if the structure in the DW's cell already carries a similar flexibility (cf. attached silent `record_1.mp4`). Instead, the whole molecule is marked. The information necessary to DW to spot potential atomic variation can be introduced by import of a .sdf with a square bracket in the corresponding line (as in the example of `compound_50.sdf`).

If not already implemented by an other command and unidentified by me, could future releases of DW provide the substructure function/atom query feature widget equally for cells about structures which still are empty (cf. second part of `record_1.mp4, and `record_2.mp4`)?

Best regards,

Norwid

File Attachments
1) compound_50.sdf, downloaded 619 times
2) nasty_functions.dwar, downloaded 588 times
3) record_1.mp4, downloaded 546 times
4) record_2.mp4, downloaded 586 times

Subject: Re: extension of substructure definition
Posted by thomas on Thu, 24 Aug 2023 14:19:33 GMT
View Forum Message <> Reply to Message

Dear Norwid,

The explanation is the following:

DataWarrior distinguishes between molecules and substructures. Molecules are considered complete structures with all open valences meant to be filled with hydrogen atoms. Molecules cannot carry query features on their atoms and bonds. Substructures (also caled fragments), however, have open valences and may have broadening or restricting query features (e.g. atom must be aromatic, or bond may be singe or double). They also may contain exclude groups.

If you open a molecule (even an empty one) in the DataWarrior editor, you cannot add query features, because they are not allowed. When editing a substructure, you may add and change query features. Molecules and substructures are logically different species, need different

handling, and are differently visualized.

An empty DataWarrior structure column contains molecules as default even if a cell is empty. If an SD-file is read, then the individual structure cells contains molecules except for those entries, where the molfile read contained a query feature and, thus, implicitly defined the structure to be a substructure. This way a structure column may contain a mix of molecules and substructures. Any new row will contains molecules, unless the column is marked to contains substructures, which usually is not the case.

You may mark a structure column manually to contain substructures by adding the column property isFragment=true. This can be done as macro task or directly in a text editor. The relevant part would then look like this:

<columnName="Structure">
<columnProperty="specialType idcode">
<columnProperty="isFragment true">

Note that the white space between 'isFragment' and 'true' must be a TAB.

Best wishes,

Thomas

Subject: Re: extension of substructure definition
Posted by nbehrnd on Wed, 30 Aug 2023 16:08:29 GMT
View Forum Message <> Reply to Message

Dear Thomas,

thanks to your answer, I understand the reasoning to retain entries about an «exact structure» and «variable (query) structure» better.

Tentatively, I attribute my failure to write a dedicated .dwam macro because I didn't identify an entry in DW's macro editor which would provide an additional column about idcoordinates2D. To me (perhaps an erroneous perception), it appears to be a requirement to render all (including future) entries in a DW structure column extendable as (similar as in Markush formulae) a query motif, because a mere look-up of a line <columnName="Structure"> (e.g. by generation of a library of random molecules) followed by the addition of the two lines did not yield the anticipated result.

As a result of testing, a minimal working example .dwar was identified. Only populated with 2,6-lutidine, the .dwar file attached below can serve future users of DW to build their set of reference of (dis)similar motifs for the generation of evolutionary libraries with Markush formula like flexibility in apparently any future row of this very column. The presence of lutidine serves as an example only.

Best regards,

Norwid

File Attachments

1) almost_empty.dwar, downloaded 625 times

Subject: Re: extension of substructure definition
Posted by thomas on Thu, 14 Sep 2023 12:52:23 GMT
View Forum Message <> Reply to Message

Dear Norwid,

in the current dev version published yesterday, when adding columns for a new or existing file, one can choose type 'Substructure', which then creates a structure column with '<columnProperty="isFragment true">'. Thus, freom now on it is possible to generate substructure columns from the user interface without hacking it in a text editor.

Best wishes,

Thomas

Subject: Re: extension of substructure definition
Posted by nbehrnd on Thu, 14 Sep 2023 20:31:39 GMT
View Forum Message <> Reply to Message

Dear Thomas,

I'm sorry to report the additional criterion to initiate the array with «substructure» instead of «structure» as first column represents an obstacle.

To venture out the new functionality, departing from a pristine installation of DW for Linux 5.5.0 (fetched 2022-05-18) in Linux Debian 13/trixie, I fetched the new update dw_wl_linux.zip. Under the assumption renaming file /opt/datawarrior/datawarrior would move it out of the way for the new files (of the update) to come, the content of the archive was moved into the folder. Lines 3--6 of the new shell script were understood as optional, and thus I retained the suggested defaults.

Case 1): Normal launch of DW, creating the array with «structure» as first column, no problem. Nor the generation of a library of random molecules and subsequent assignment of computed properties.

Case 2): Normal launch of DW, creating the array with «substructure» as first column. Now, the command shell became talkative. Shortly after, the message of «uncaught exception 3» on a widget appears which is above an other without a note (except the «cancel» button»). If I click on «cancel», this second widget claims to initiate some space cleaning. At this point, the task manager of the operational system does not show noticeable RAM allocated to DW.

However leaving DW from its GUI is blocked, an exit possible the hard way by killing the corresponding PID.

Because of the noteworthy reorganization towards DW 5.6.0 and «it is a beta version, if something goes wrong, restore the old /opt/datawarrior/datawarrior file»), I equally gave this spin; in my case however it is on expense of new column type «substructure».

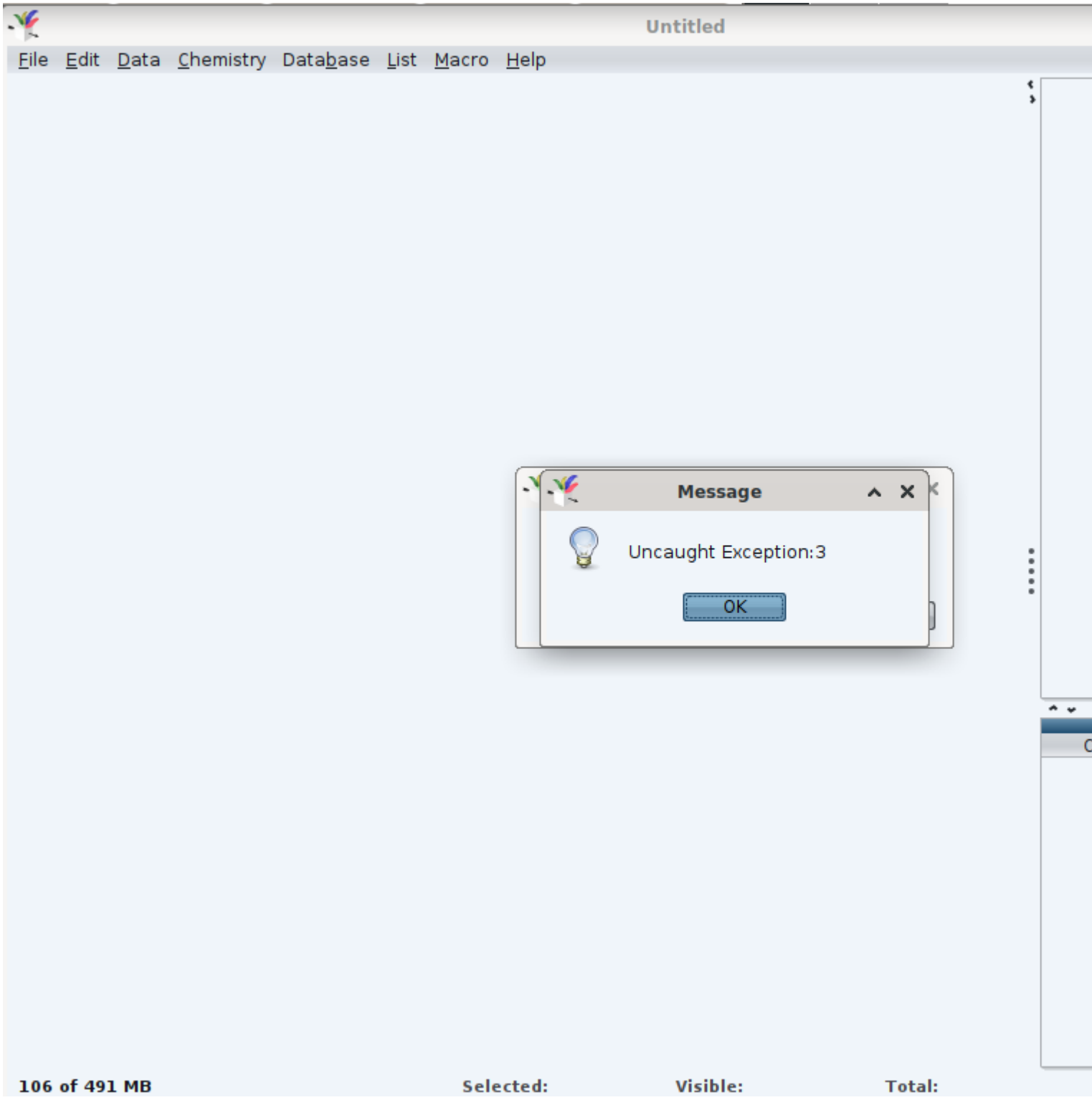
Attached below, screen photos recorded and what was logged via the shell (in this case, bash).

Best regards,

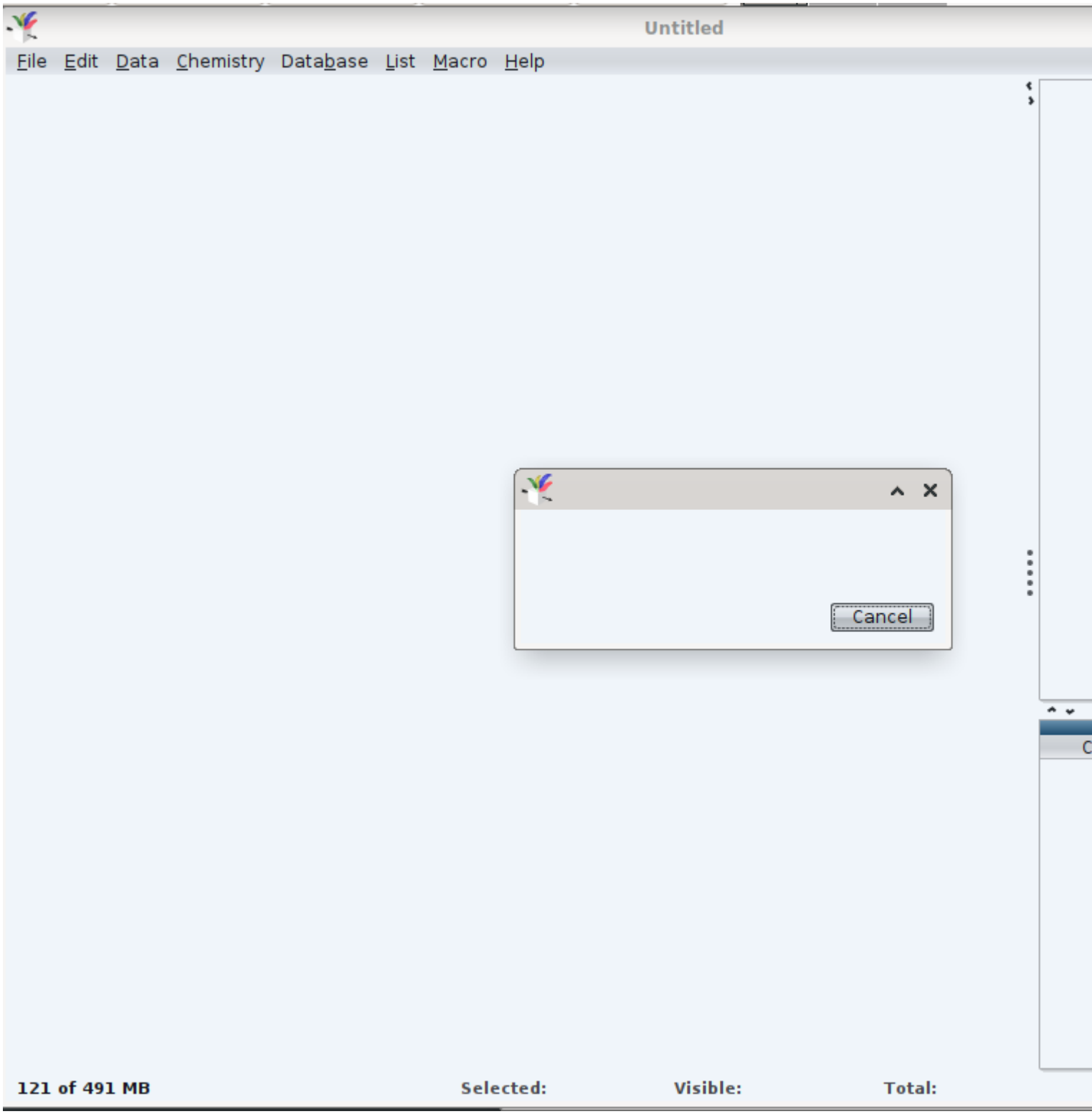
Norwid

File Attachments

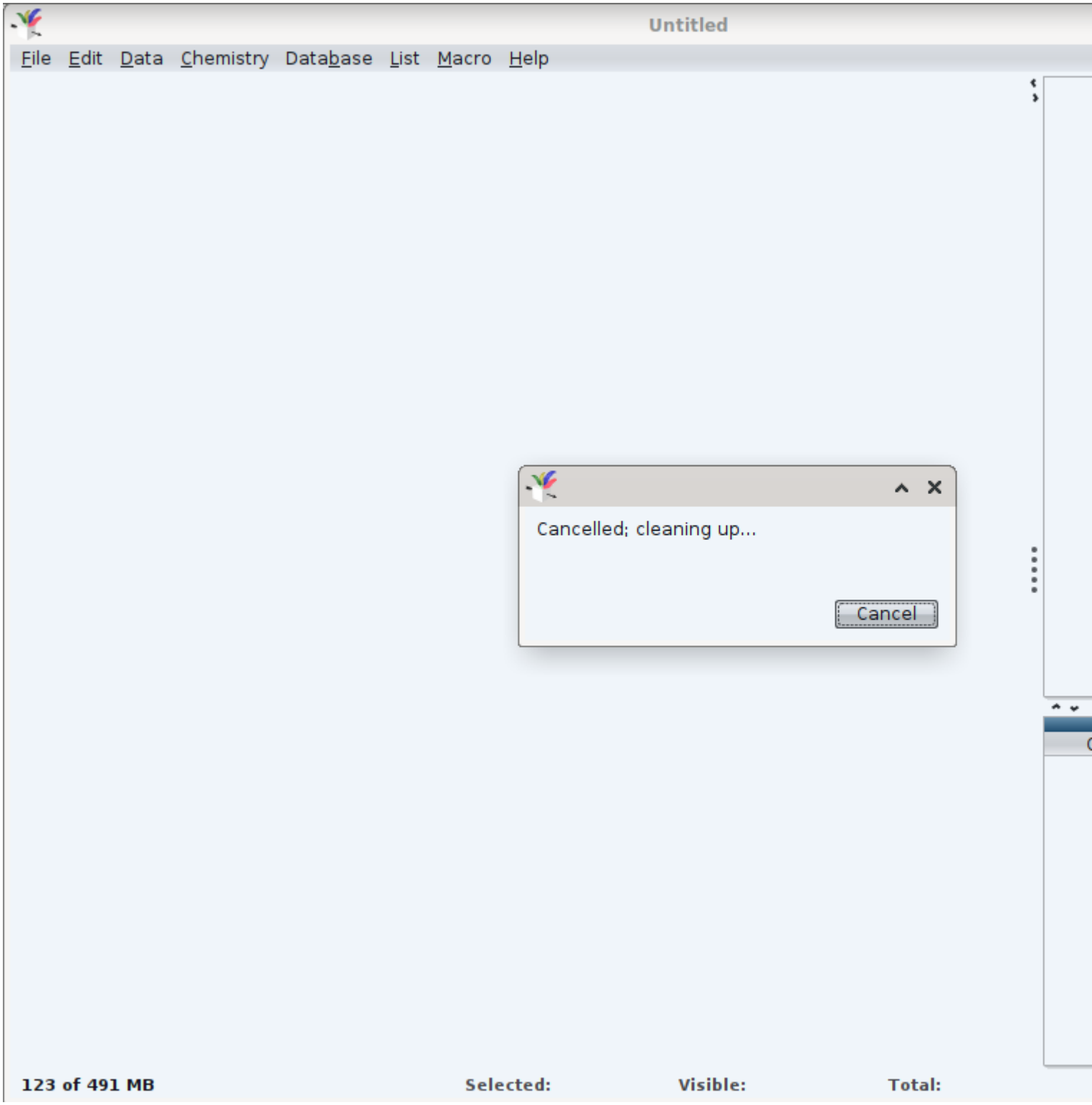
1) problem_01.png, downloaded 889 times



2) problem_02.png, downloaded 889 times



3) problem_03.png, downloaded 756 times



4) 2023-09-14_relevant_log -- dw.txt, downloaded 525 times

Subject: Re: extension of substructure definition
Posted by thomas on Sun, 17 Sep 2023 19:32:28 GMT
View Forum Message <> Reply to Message

Dear Norwid,

I am very sorry, but evidently there slipped a stupid bug into the new functionality. I have fixed it, but it requires to update your files with the newest dw_wl_linux.zip.

Sorry for the inconvenience any many thanks for your support...

Thomas
