StarDrop™

Version 8

StarDrop™ Worked Example: Reaction-based Library Enumeration

optibrium™

Contents

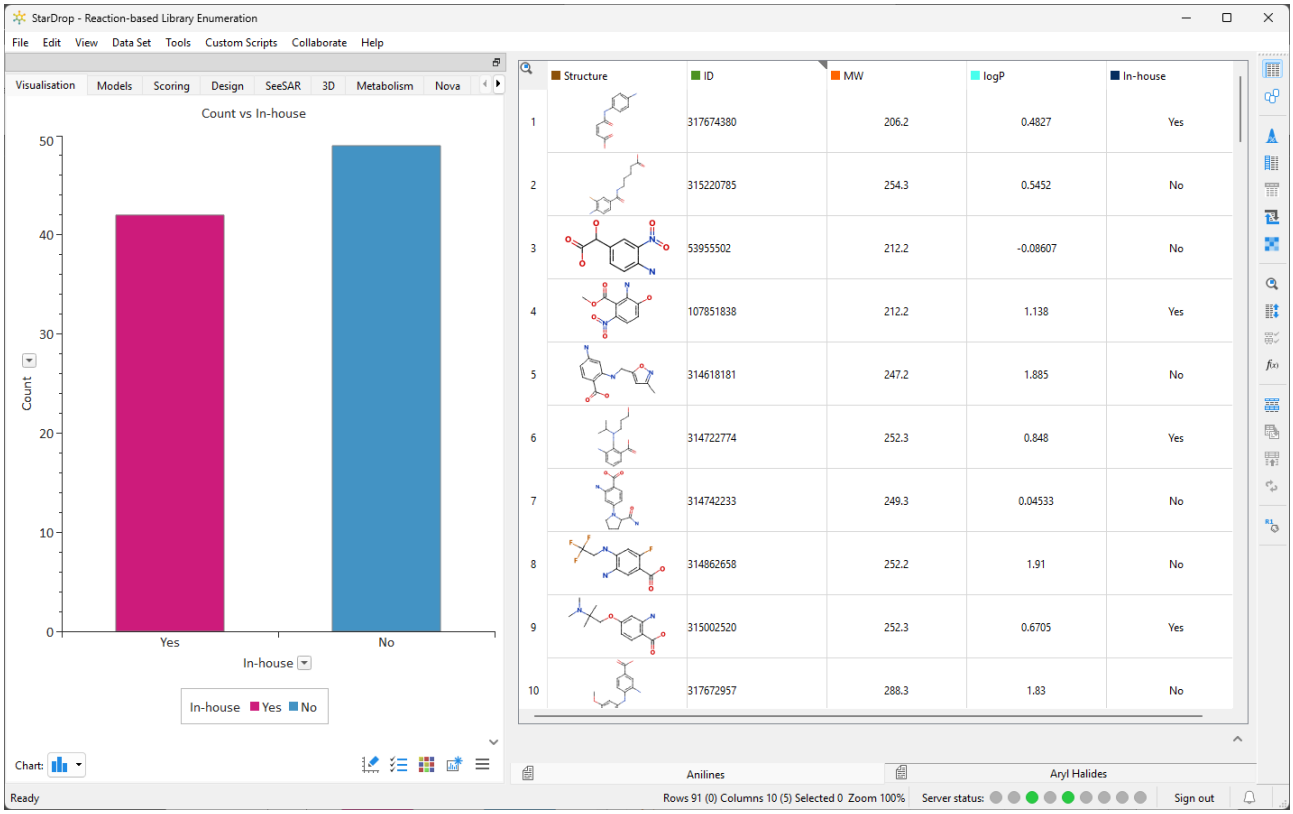
Overview

In this example, we will use the reaction-based library enumeration feature in StarDrop's Nova™ module to generate a library of virtual compounds. This will be based on predefined sets of reagents used to generate products via well-known reactions.

Creating a virtual library of compounds with Nova's Reaction-based Library Enumeration

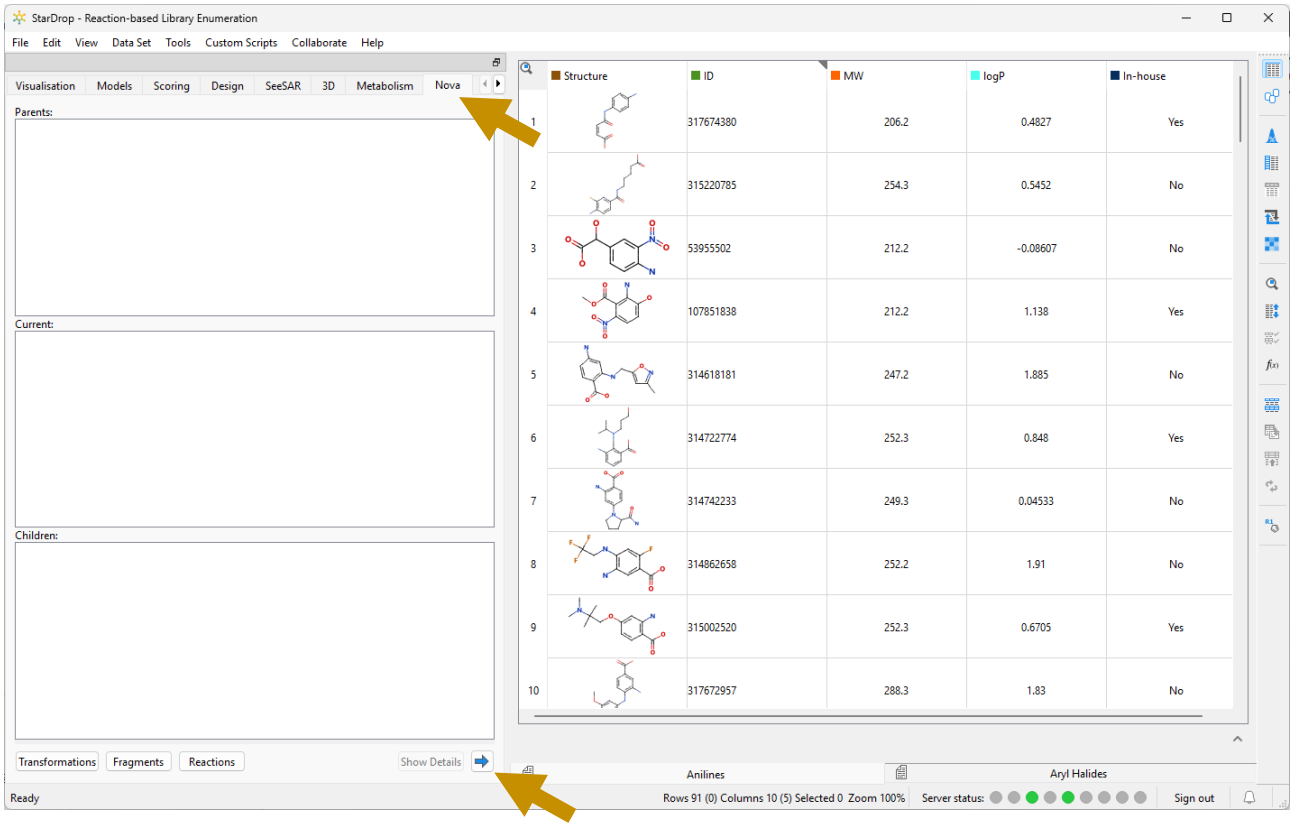
- In StarDrop, select **Open** from the **File** menu, find the workspace **Reaction-based Library Enumeration.sdproj** and open it by selecting it.

The project contains two reagent data sets. The current data set contains a set of anilines, and these have been categorised, indicating whether they are available "in-house".

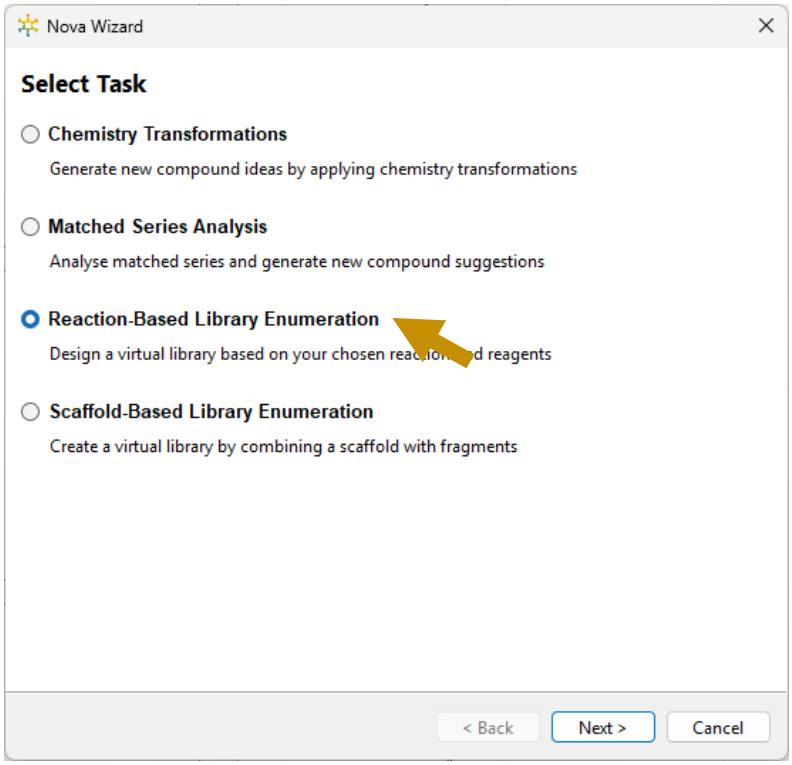


We'll enumerate a library, limiting the choice of anilines to those available in-house.

- Click on the **Nova** tab.



- Click on the  button to start the Nova dialog.
- Choose **Reaction-Based Library Enumeration** and click the **Next** button.



The first step is to choose the reaction that you would like to use.

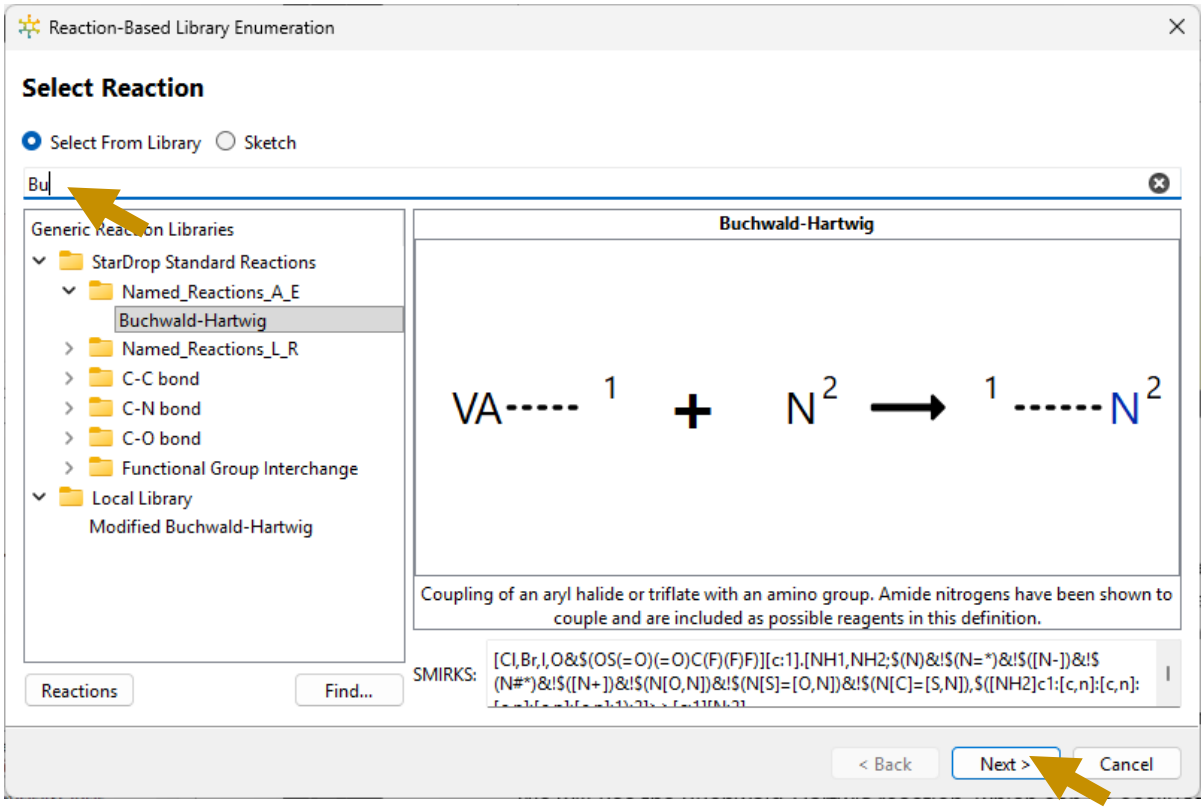
There are over 120 reactions available, and you can browse these either by their name or their general type. You will also see any reactions that you have created, along with any shared by your organisation, in the list of **Generic**

Reaction Libraries (for more information on how to create and share libraries of reactions, please get in touch with support@optibrium.com.)

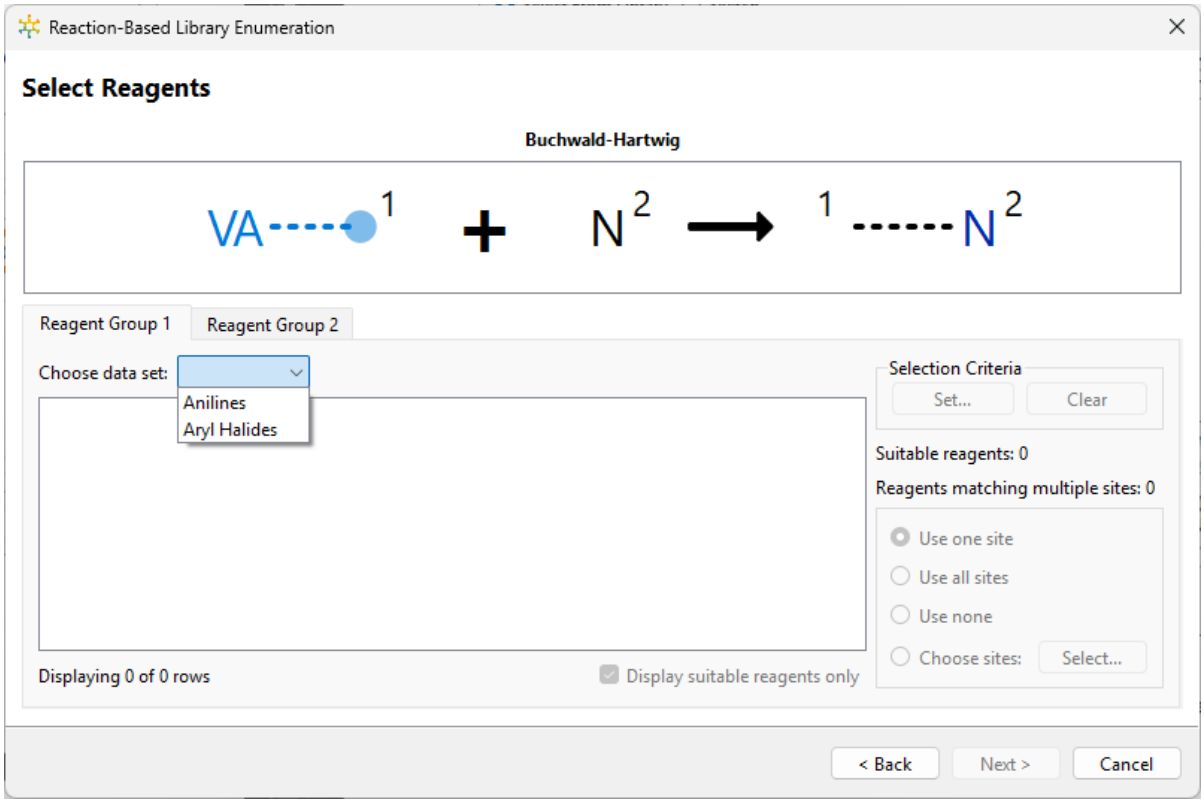
If you wish to use a reaction that is not in the list, then you can choose the **Sketch** option to display a reaction editor.

We will use the Buchwald-Hartwig reaction, which can be easily found by typing its name into the search bar at the top. As you type, the list will contract to display only matching reactions.

- Type the letters “bu” into the search bar, and Buchwald-Hartwig will appear at the top and be selected.
- Click the **Next** button.

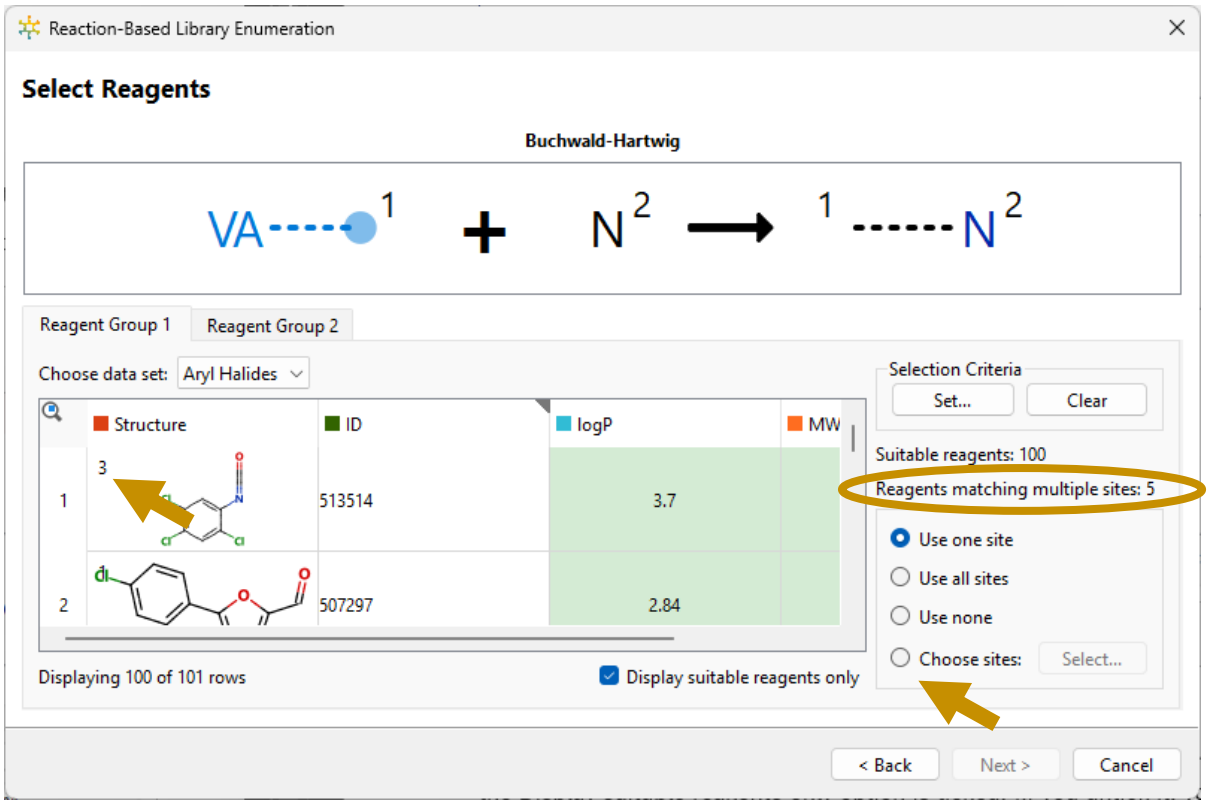


We can now define the reagents that we wish to use in the reaction. On the next page of the dialog, you can see the chosen reaction displayed, and below it, a series of tabs, one for each reagent in the reaction.



- For **Reagent Group 1**, select the **Aryl Halides** data set.

This data set contains 101 reagents, but only 100 of them are suitable and meet the reaction definition. By default, the **Display suitable reagents only** option is ticked. (If you untick it, you'll see row three appear. This has been excluded because the reaction definition includes only aryl Cl, Br, I or triflate groups, not F).

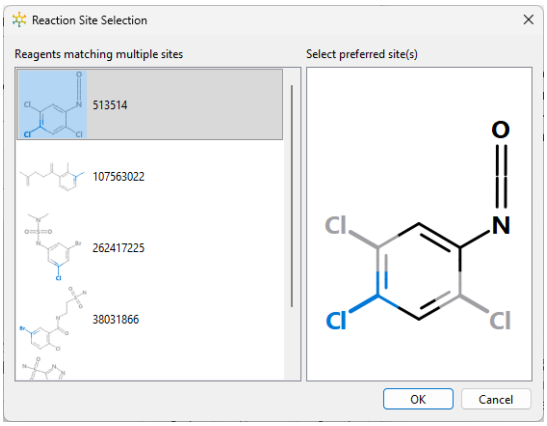


There is also an indication if there is more than one reactive site on any of the reagents. To the top left of each structure is an indicator showing the number of sites on that reagent that match the reaction. If you are only interested in general properties, then the first site option, **Use one site**, may be sufficient, but in more complex cases, you may prefer to **Use none**. The option **Use all sites** will carry out an exhaustive enumeration. In this case, we will **Choose sites**, which enable us to specify the regioselectivity.

- Tick the **Choose sites** option and then click the **Select** button.

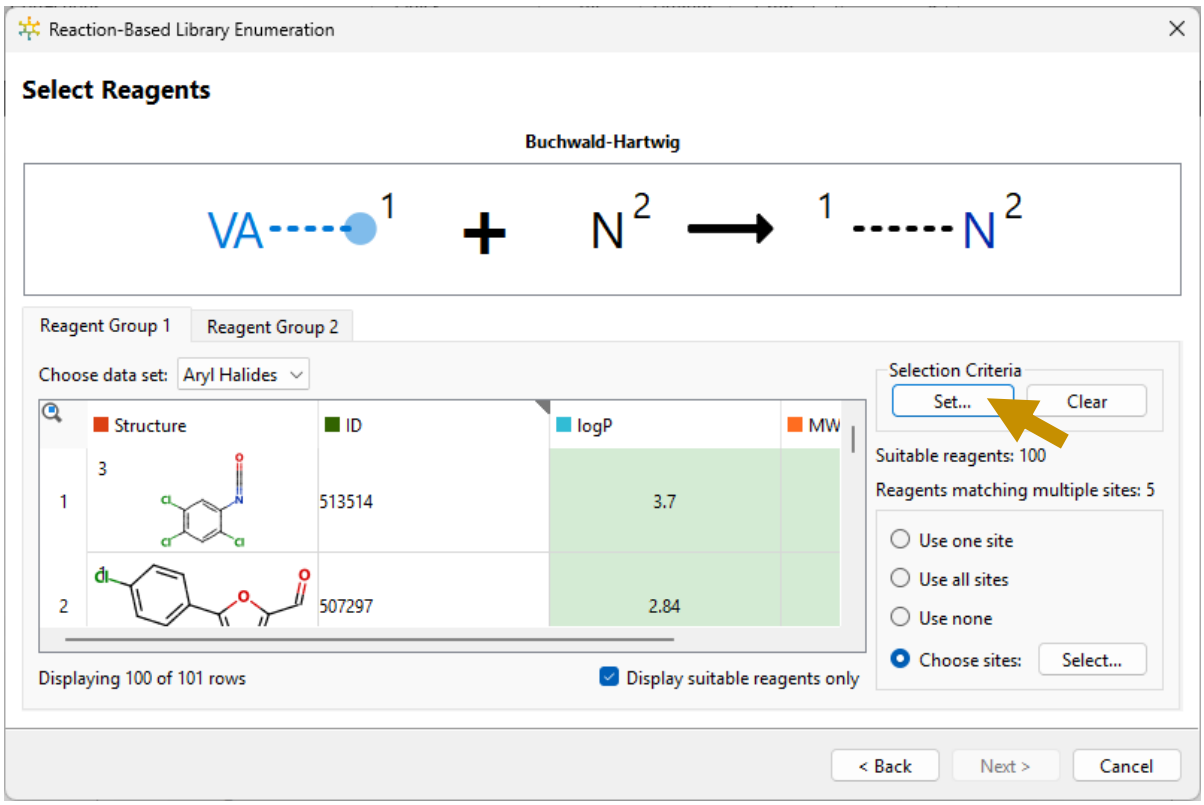
Each reagent with more than one site that matches the reaction is shown on the left. Selecting any of these will display it to the right, enabling you to choose which site(s) you wish to participate in the reaction. Sites that have been selected will be blue, and unselected sites grey. Clicking on a site will toggle it between being selected and unselected.

- Click on each reagent in turn on the left-hand side:
  - Where the sites are all **Cl**, ensure that only the para-Cl site is selected.
  - Where there is a **Br**, make sure that this is selected preferentially over **Cl**, because it is more reactive.
- Having done this for each of the five reagents, click the **OK** button.



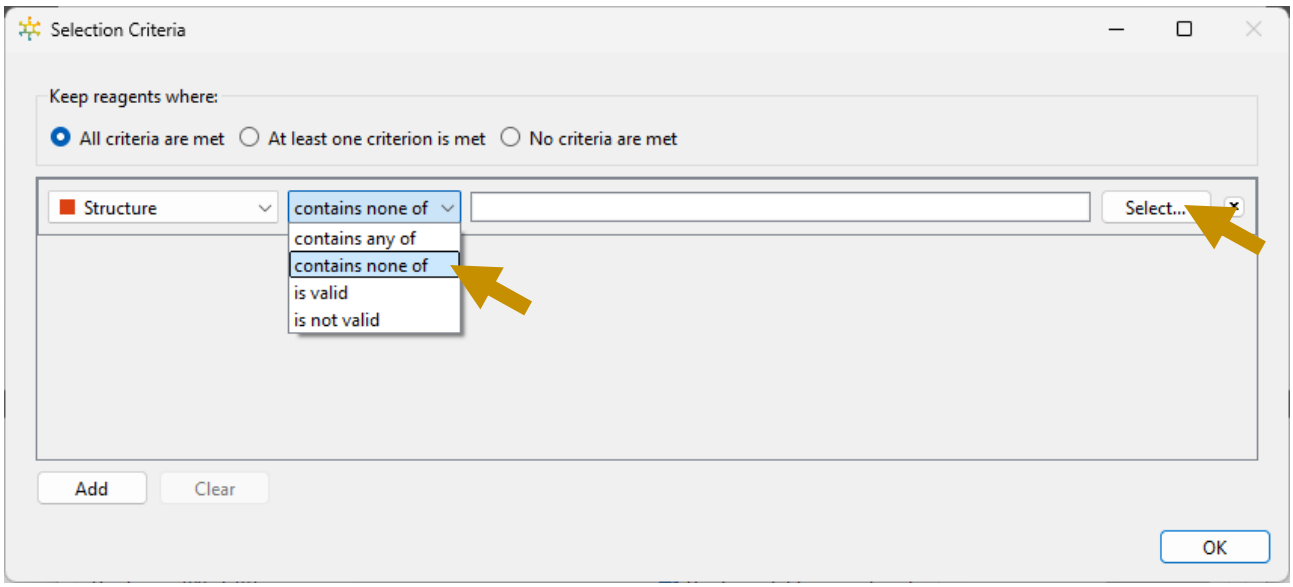
In this set of reagents, while they are all aryl halides, there are also some substituents that may be undesirable. The first reagent contains an isocyanate which is reactive to nucleophiles and therefore unlikely to survive the reaction. In addition, some of the other reagents contain aldehydes which may be undesirable. We will therefore filter this reagent set before proceeding.

- In the **Selection Criteria** section on the right, click the **Set** button.



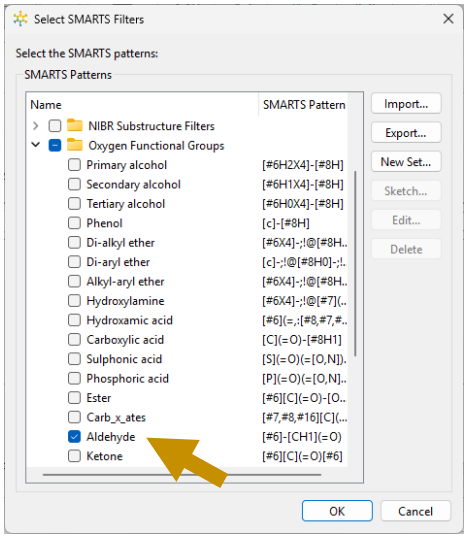
Here we can define characteristics to look for in reagents we wish to keep; in this case, reagents that do not contain any of the functional groups we wish to avoid.

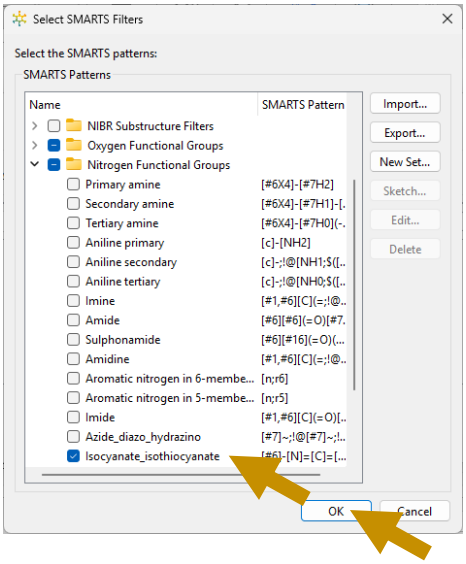
- From the drop-down list next to **Structure**, **choose contains none of**.
- Click Select to choose the functional groups to avoid.



The dialog provides many pre-defined functional groups (represented by SMARTS patterns), and you can add your own to this list by importing them or sketching.

- In the **Oxygen Functional Groups**, tick the box next to **Aldehyde**.
- In the **Nitrogen Functional Groups**, tick the box next to **Isocyanate_isothiocyanate**.
- Click the **OK** button.



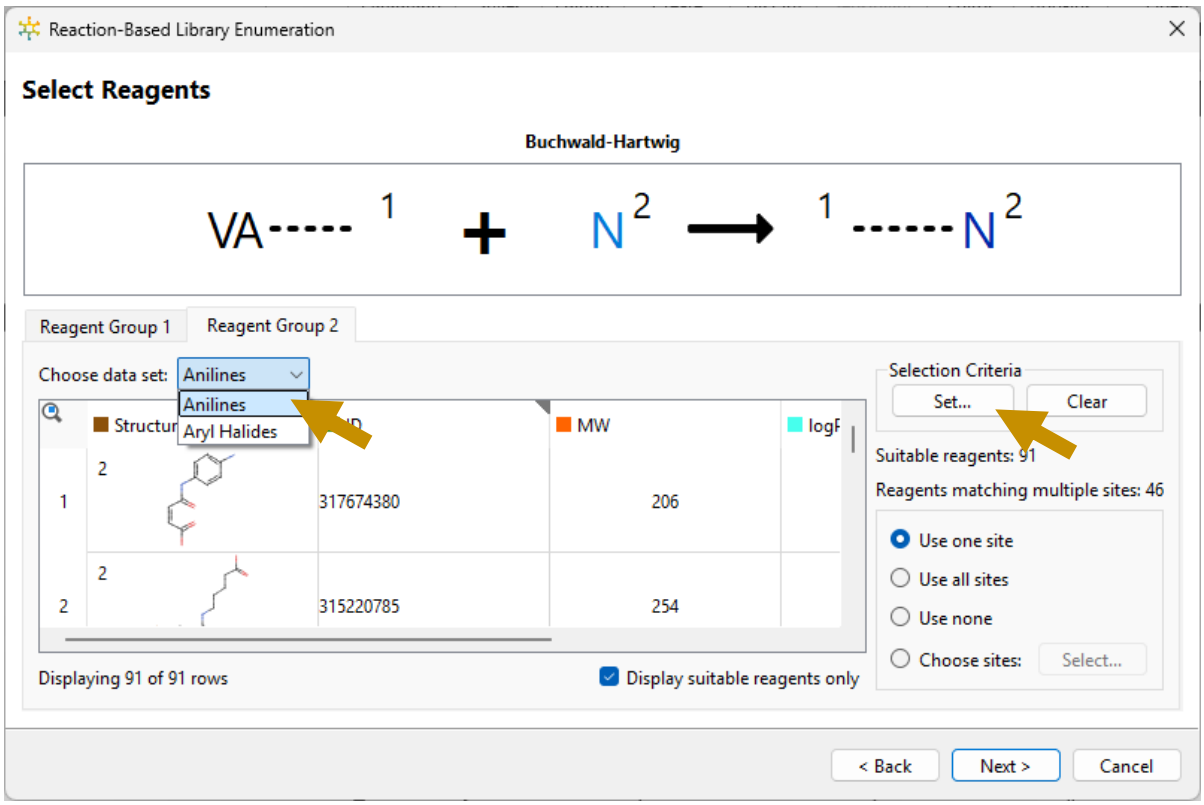


The **Selection Criteria** dialog will now show these groups to be avoided. Note that you can define multiple selection criteria based on any combination of the structure and properties of the reagents.

- Click the **OK** button to use the selection criteria we have defined.

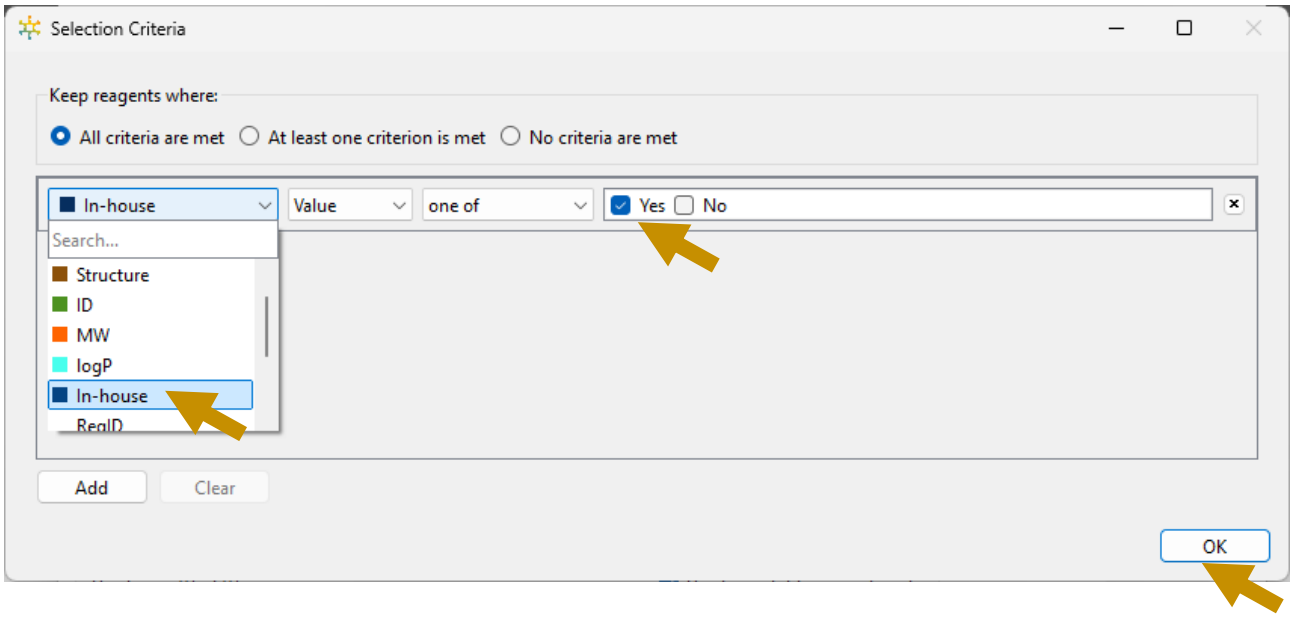
We now have 95 suitable reagents remaining in the set and can move on to **Reagent Group 2**.

- Click on the **Reagent Group 2** tab and then, from the data set list, choose **Anilines**.



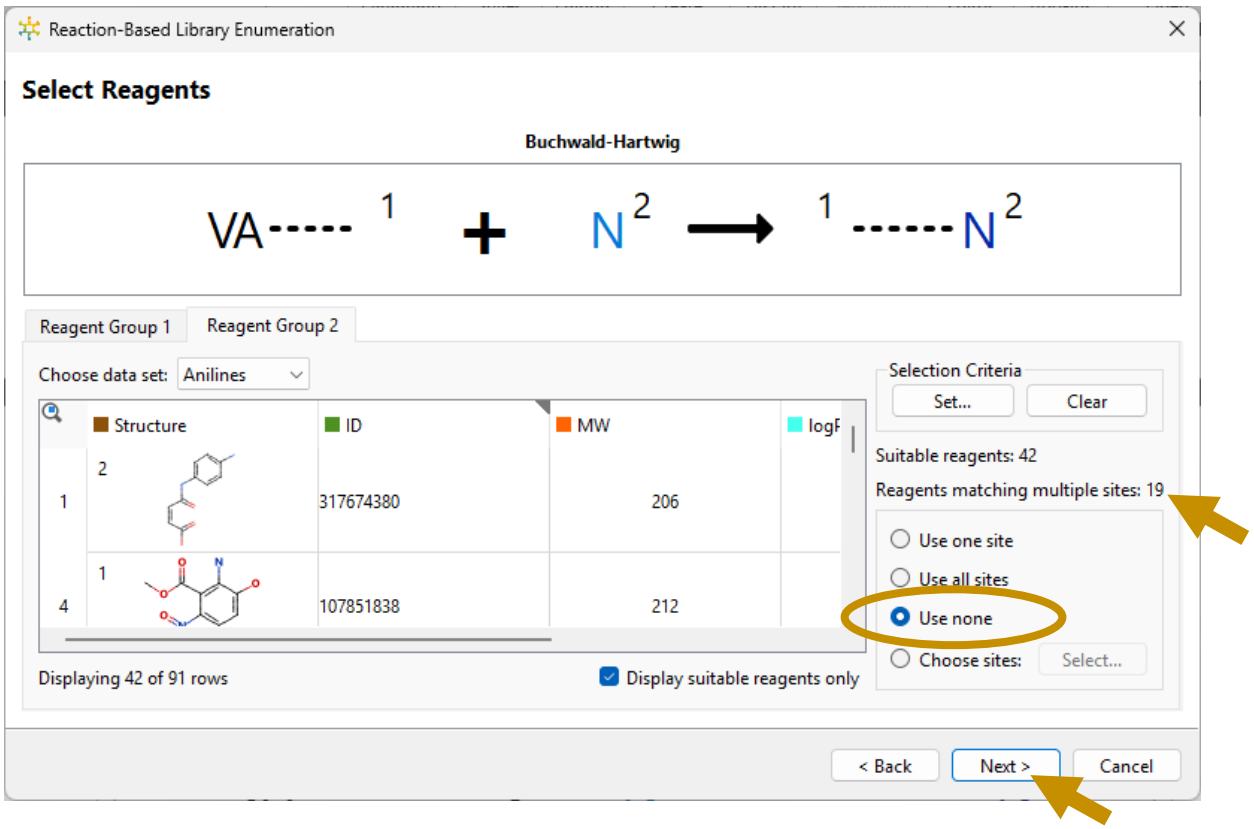
For the anilines, we're going to specify selection criteria based on in-house availability.

- In the **Selection Criteria** section on the right, click the **Set** button.
- From the drop-down menu on the left, choose **In-house**.
- Tick the **Yes** option to indicate that we only wish to keep reagents that are available in-house.
- Click the **OK** button.



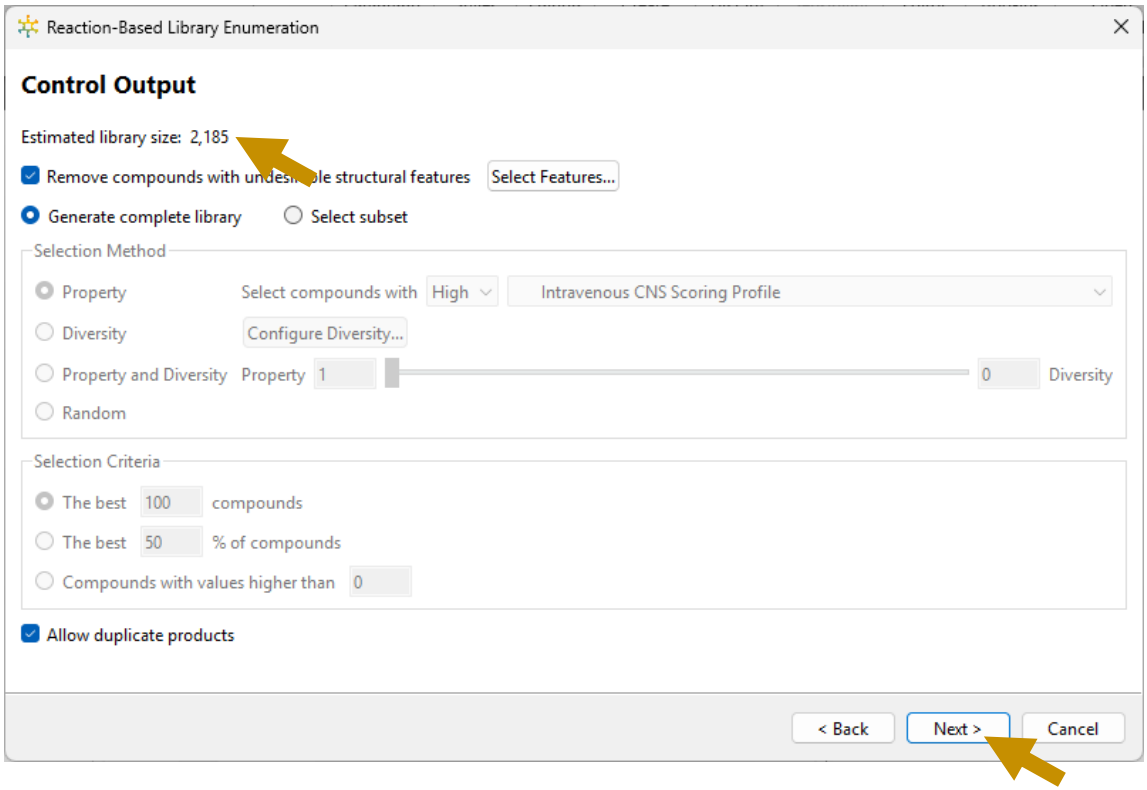
There are 19 reagents with multiple sites, so for this set, we'll avoid using these 19 reagents.

- Select the **Use none** option and then click the **Next** button.

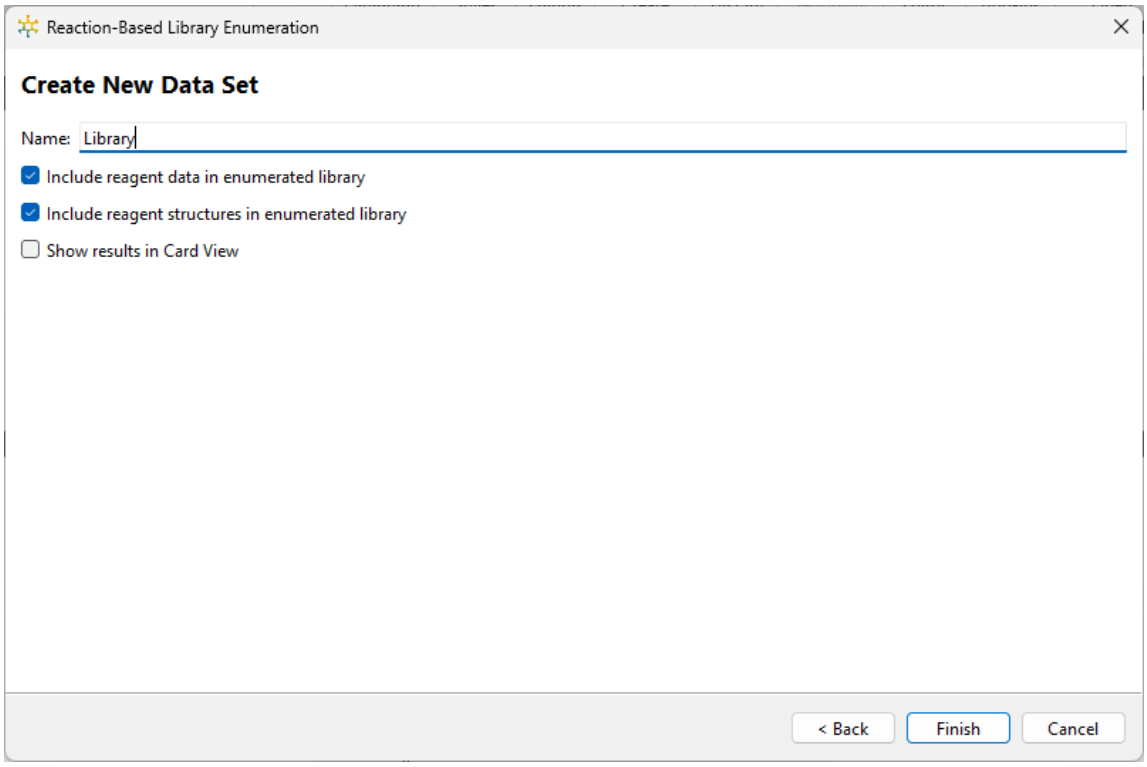


The **Control Output** page shows us that there will be approximately 2,185 products generated using the reagents we have selected. With this being a small library, it is no problem to generate the complete set; however, it would be easy to specify reagent sets that produce libraries with many thousands, or even millions, of products when combined. In such a scenario, it may be useful to choose the Select subset option, after which you can choose how to select the products based upon a property, diversity, or a score combining multiple properties to be optimised simultaneously.

- Click the **Next** button.



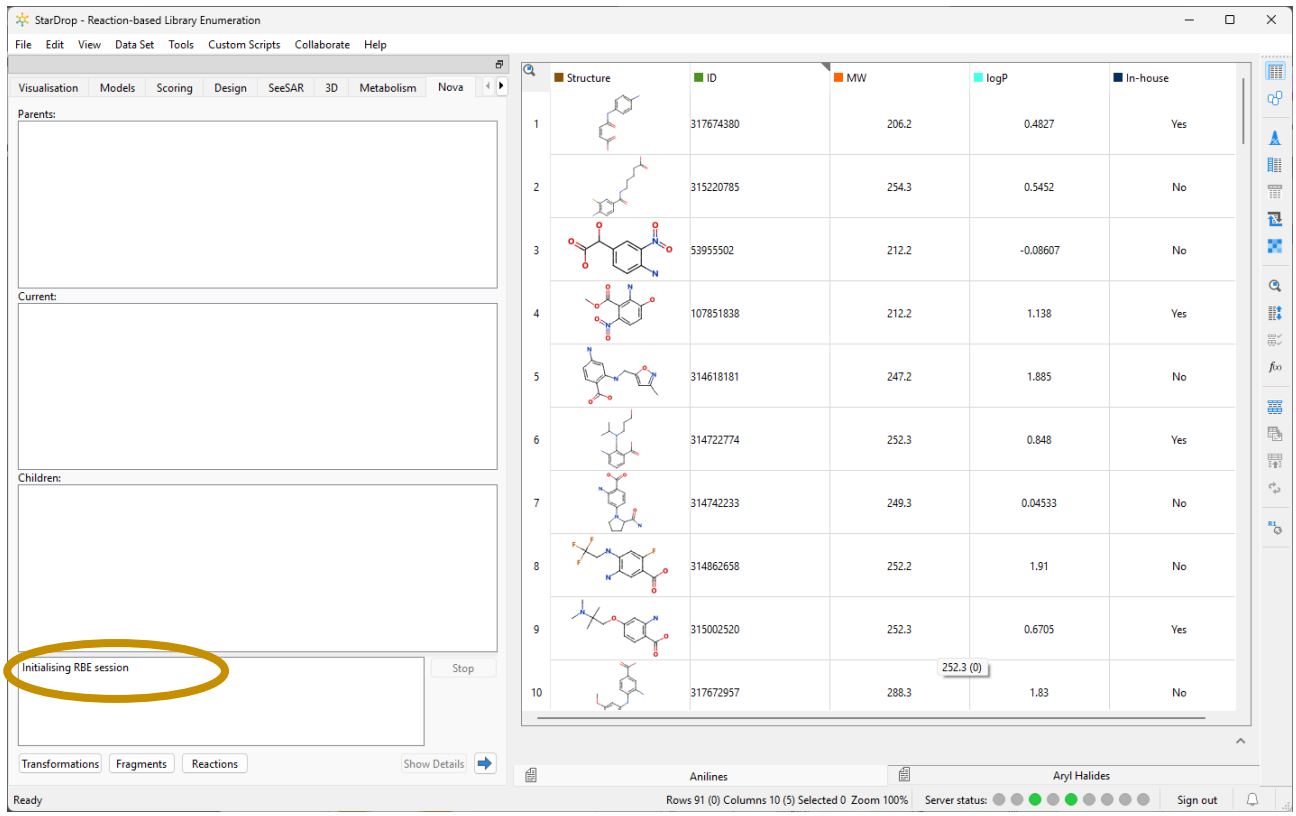
- Give the new data set a name (in this case, we have called it “Library”).



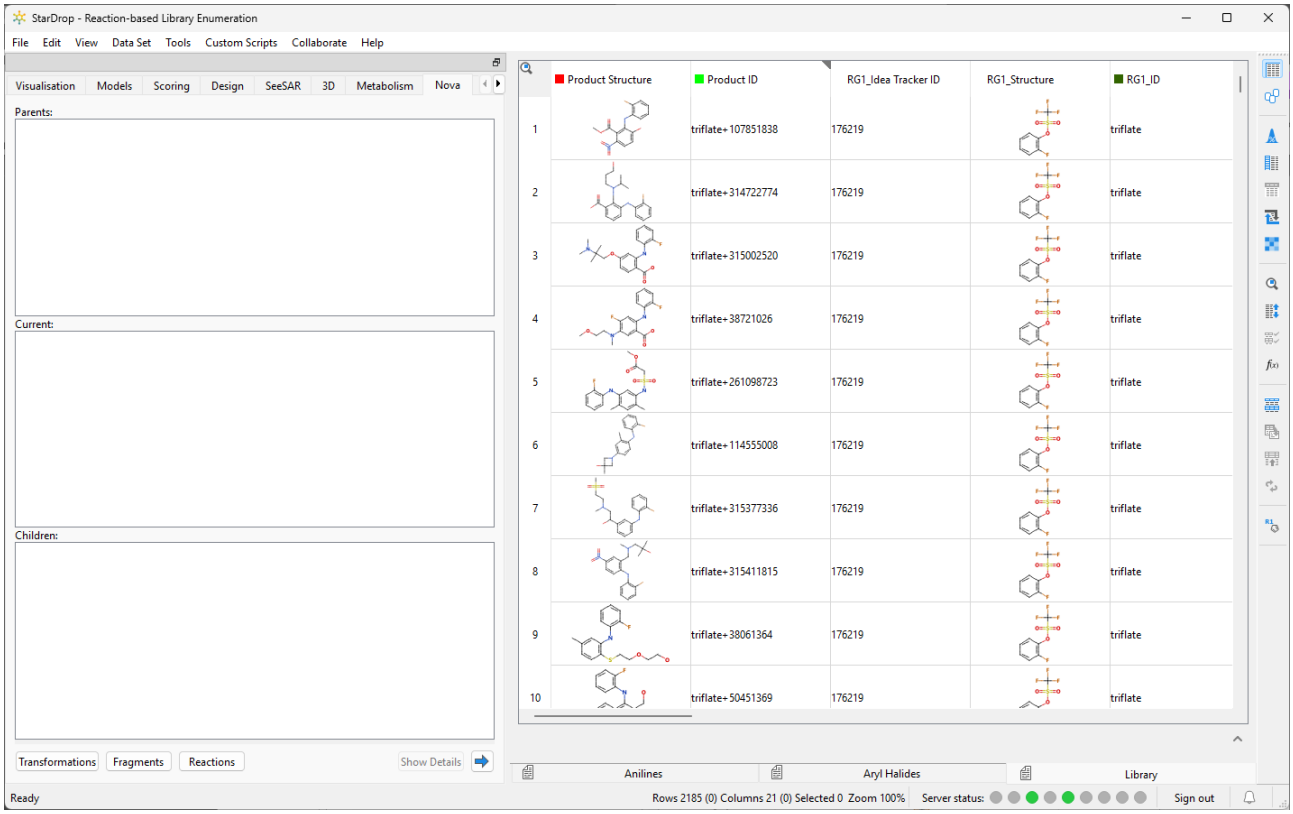
By default, all data associated with reagents, as well as their structures, will be included in the new library. If you are generating a library with many thousands of products, then you might wish to untick the Include reagent structures in enumerated library option, because for large sets, the additional structures in every row may result in a significantly larger file size when you save the StarDrop project.

- Click the **Finish** button.

While the library is being enumerated, StarDrop will provide a progress indicator within the Nova area. You can continue to use StarDrop while the enumeration takes place.



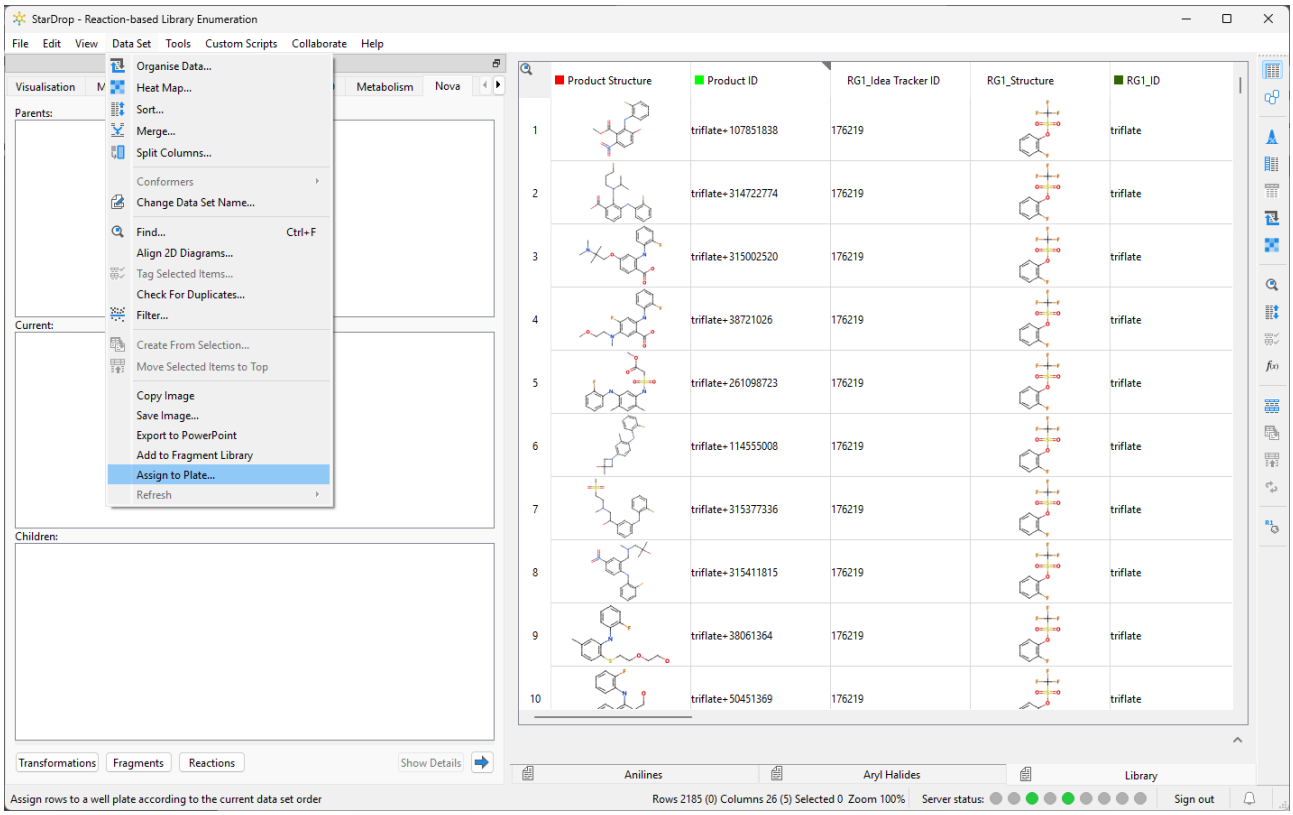
When the process has finished, the new data set will be displayed.



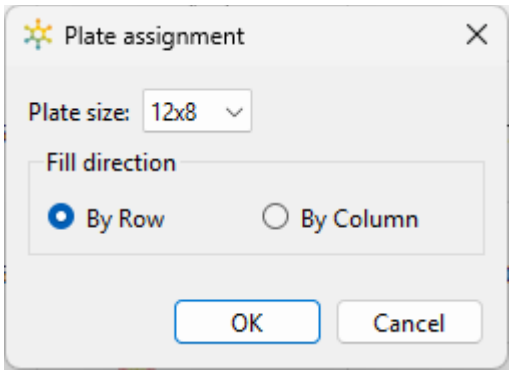
Each product in the data set is given an automatic ID based on the combination of the names of the reagents used. As you scroll across the data set, for each product, you will see all the properties from its associated reagents, prefixed by RG1_ for reagent 1, RG2_, etc.

To prepare this data set for synthesis, we will assign plate IDs to all the products.

- From the **Data Set** menu, choose **Assign to Plate**.



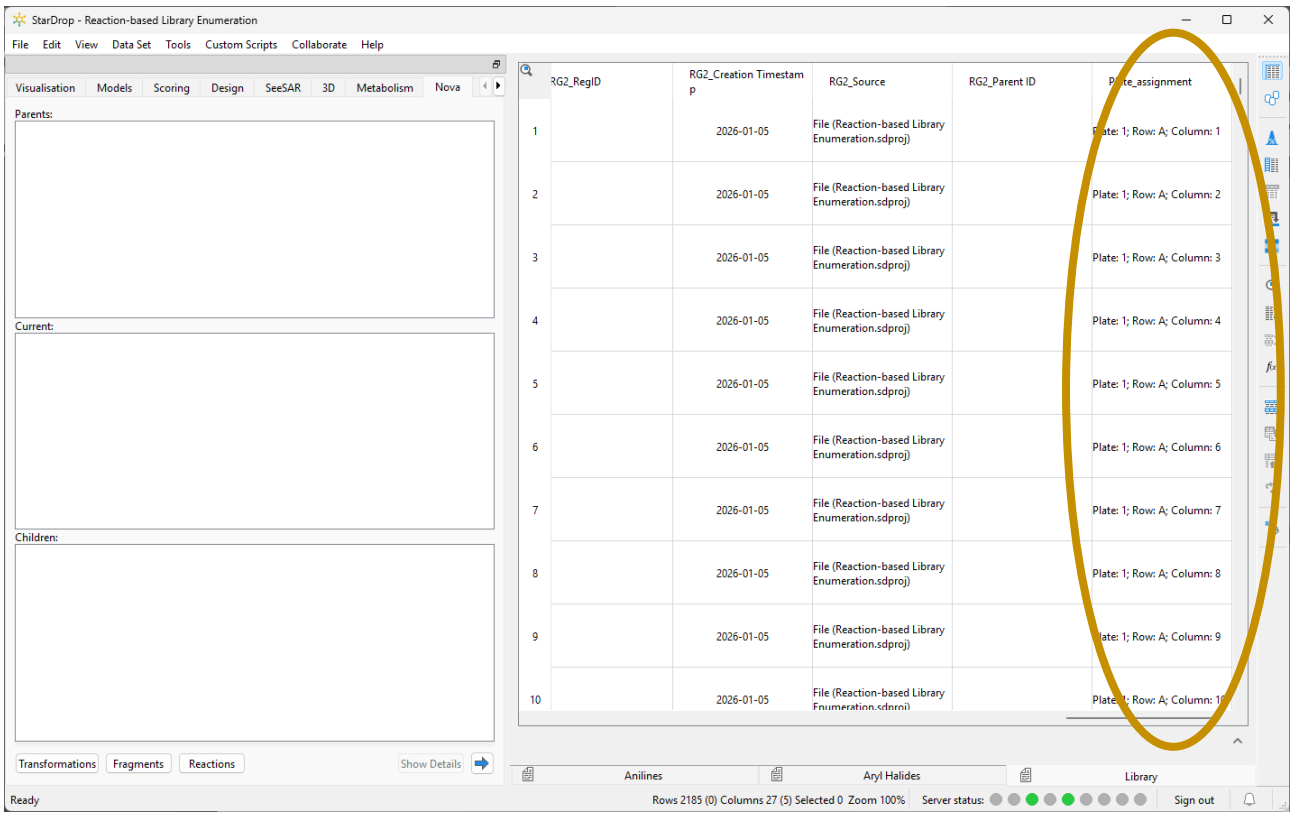
The dialog that appears enables you to choose between a 96 (12x8), 384 (24x16), and 24 (6x4) well plate.



In addition, you can choose whether to fill the plate **By Row** or **By Column**.

- In this case, we will use the default options, so click the OK button.

A new plate assignment column is added to the data set. Each entry shows the plate number, the row (indicated by a letter) and the column (indicated by a number).



All StarDrop's capabilities can be used to further analyse or visualise the resulting library. You can also export the data set easily via the File menu as a CSV or text file when you're ready to send the library for synthesis.

Conclusion

With this worked example, we have demonstrated a reaction-based library enumeration using StarDrop's Nova module to generate a library of virtual compounds based on pre-defined sets of available reagents. Additional features of the Nova module can be found in examples that include Matched Series Analysis, Scaffold Hopping via R-Group Clipping, and the Automatic Generation of New Compound Ideas.

If you have any questions, please feel free to contact support@optibrium.com.