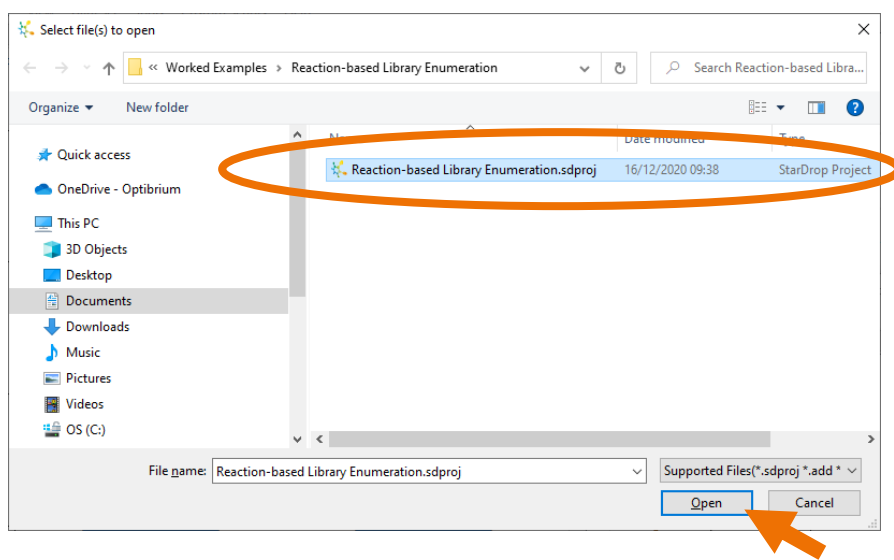


Worked Example:

Reaction-based Library Enumeration

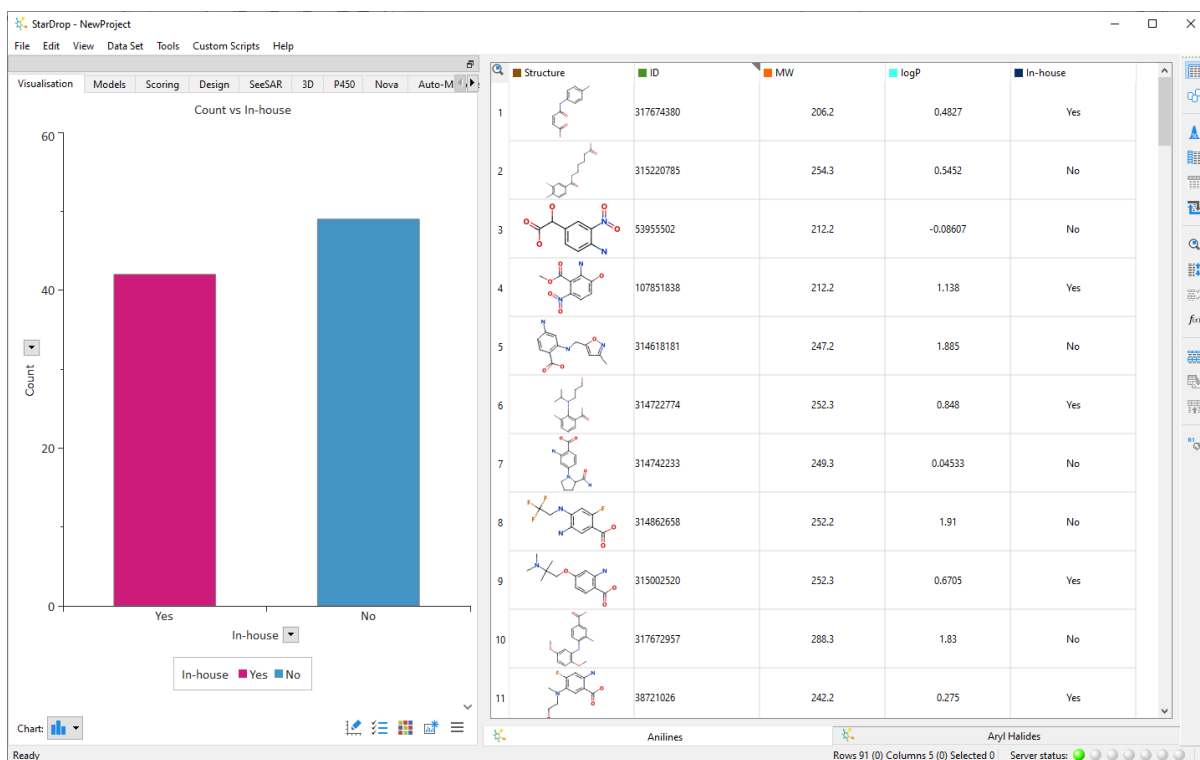
In this example, we are going to use the reaction-based library enumeration feature in StarDrop's Nova module to generate a library of virtual compounds. This will be based on pre-defined sets of reagents that will be used to generate products using well-known reactions.

- In StarDrop, open the project **Reaction-based Library Enumeration.sdproj** by selecting **Open** from the **File** menu.



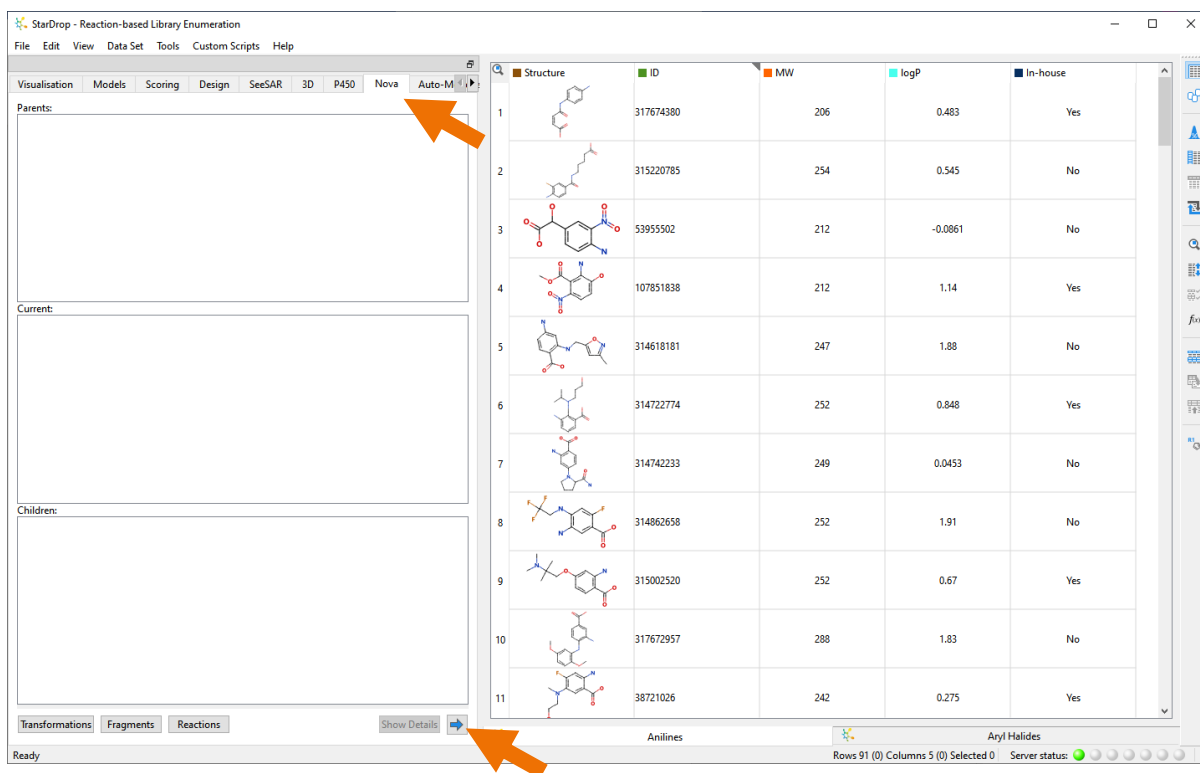
The project contains two reagent data sets. The current data set contains a set of anilines and these have been categorised, indicating whether they are available “in-house”.






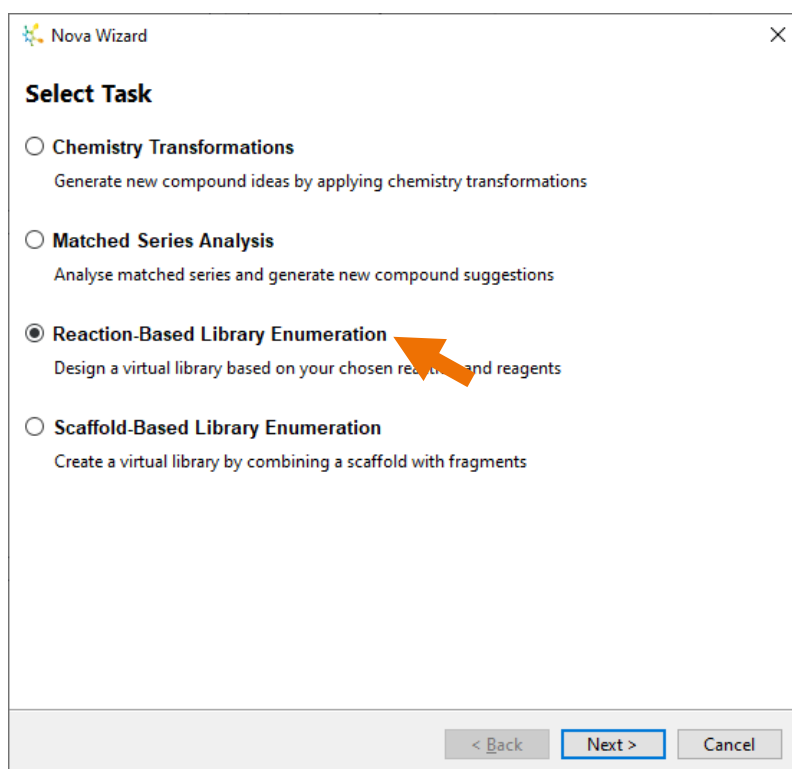
We're going to enumerate a library limiting the choice of anilines to those available in-house.

- Click on the **Nova** tab.

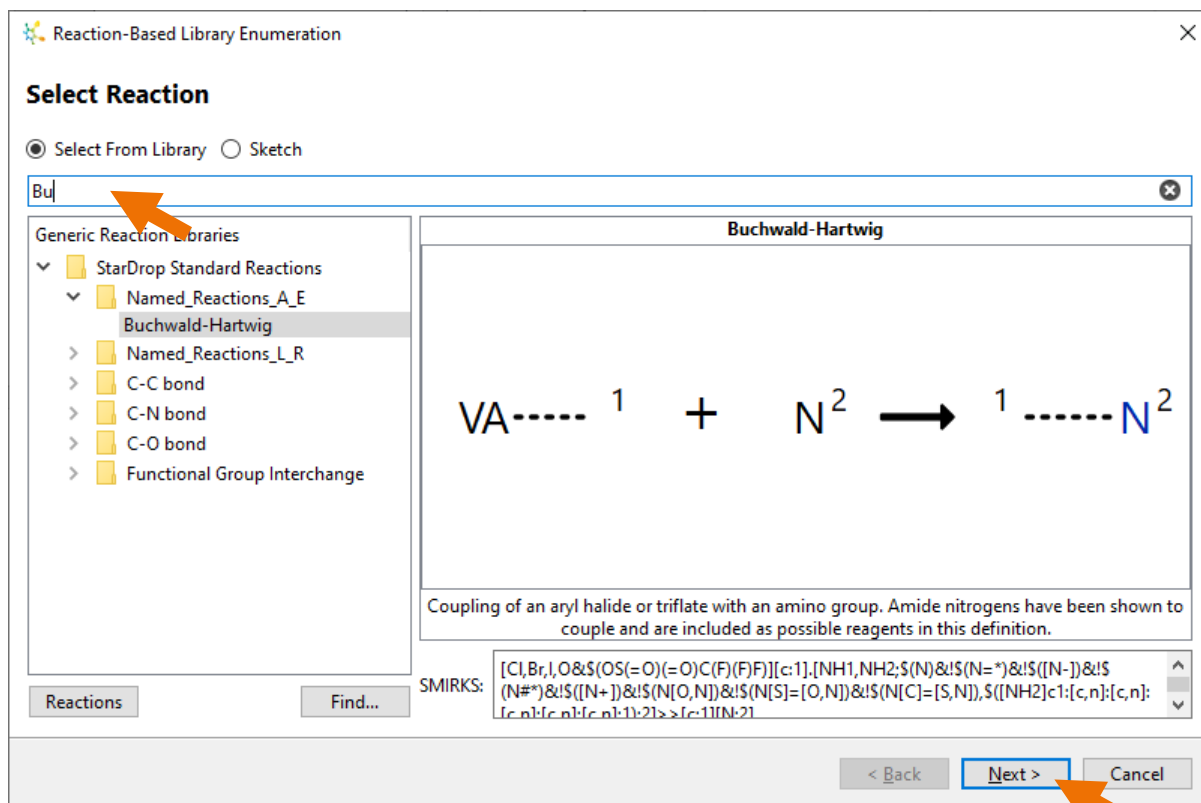


- Click on the  button to start the Nova Wizard.

- 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 contact stardrop-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're going to use the Buchwald-Hartwig reaction, which you can find easily by starting to type 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 wizard, you can see the chosen reaction displayed, and below it a series of tabs, one for each reagent in the reaction.

Reaction-Based Library Enumeration

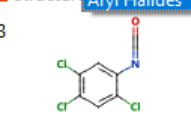
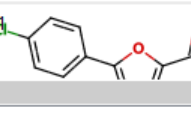
Select Reagents

Buchwald-Hartwig

$$VA \cdots \bullet^1 + N^2 \longrightarrow \bullet^1 \cdots N^2$$

Reagent Group 1 | Reagent Group 2

Choose data set: Aryl Halides

Structure	logP	MW
	3.7	513514
	2.84	507297

Displaying 100 of 101 rows

Display suitable reagents only

Selection Criteria

Suitable reagents: 100

Reagents matching multiple sites: 5

Use one site
 Use all sites
 Use none
 Choose sites: Select...

< Back | Next > | Cancel

- For reagent group 1, select the **Aryl Halides** data set.

This data set contains 101 reagents, but only 100 of them are suitable, as shown to the right. 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).

Reaction-Based Library Enumeration

Select Reagents

Buchwald-Hartwig

VA---1 + N² → 1---N²

Reagent Group 1 | Reagent Group 2

Choose data set: Aryl Halides

Structure	ID	logP	MW
3 1	513514	3.7	
2 2	507297	2.84	

Displaying 100 of 101 rows

Display suitable reagents only

Selection Criteria

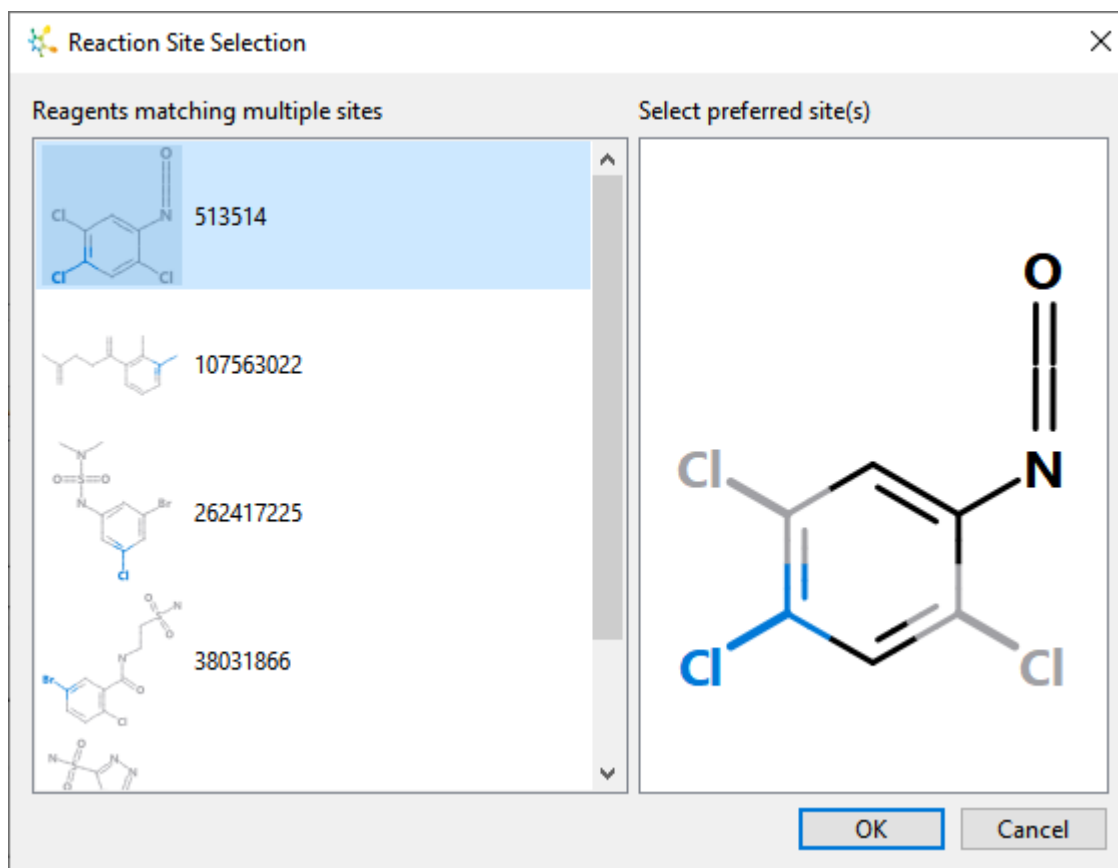
Suitable reagents: 100
Reagents matching multiple sites: 5

Use one site
 Use all sites
 Use none
 Choose sites: Select...

< Back | Next > | Cancel

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**. In this case, we will **Choose sites**, which enables us to specify the regioselectivity.

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



Each of the reagents 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 take part 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.

Reaction-Based Library Enumeration

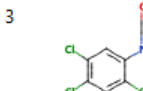
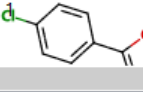
Select Reagents

Buchwald-Hartwig

$$VA \cdots 1 + N^2 \longrightarrow 1 \cdots N^2$$

Reagent Group 1 Reagent Group 2

Choose data set: Aryl Halides

Structure	ID	logP	MW
	513514	3.7	
	507297	2.84	

Displaying 100 of 101 rows Display suitable reagents only

Selection Criteria

Set... Clear

Suitable reagents: 100
Reagents matching multiple sites: 5

Use one site
 Use all sites
 Use none
 Choose sites: Select...

< Back Next > Cancel

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

Selection Criteria

Keep reagents where:

All criteria are met
 At least one criterion is met
 No criteria are met

Structure
 contains any of

 Select...

contains any of
 contains none of
 is valid
 is not valid

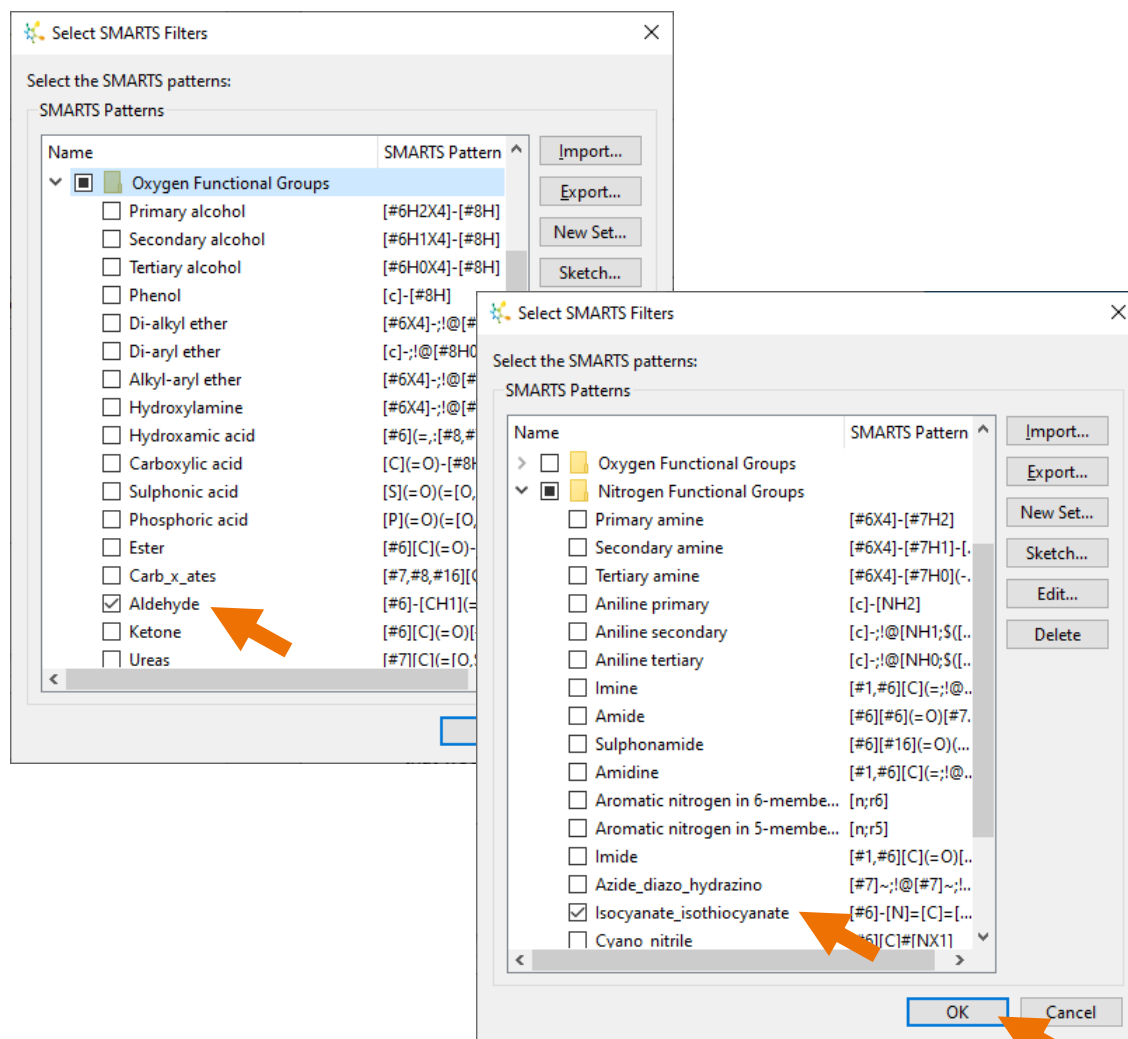
Add Clear

OK

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 dialogue provides a large number of 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 dialogue will now show these groups to be avoided. Note that you can define multiple selection criteria based on and 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 left 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**.

Reaction-Based Library Enumeration

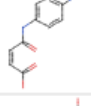

Select Reagents

Buchwald-Hartwig

$$VA \text{-----}^1 + N^2 \longrightarrow 1 \text{-----}N^2$$

Reagent Group 1 Reagent Group 2

Choose data set: Anilines

Structure	MW	logP
	317674380	206
	315220785	254

Displaying 91 of 91 rows Display suitable reagents only

Selection Criteria

Set... Clear

Suitable reagents: 91
Reagents matching multiple sites: 46

Use one site
 Use all sites
 Use none
 Choose sites: Select...

< Back Next > Cancel

For the anilines, we're going to specify selection criteria based upon 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.

Selection Criteria

Keep reagents where:

All criteria are met
 At least one criterion is met
 No criteria are met

Value	one of	<input checked="" type="checkbox"/> Yes <input type="checkbox"/> No
In-house		

Add Clear

OK

- Click the **OK** button.

Reaction-Based Library Enumeration

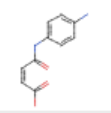
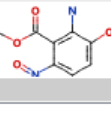
Select Reagents

Buchwald-Hartwig

$$VA \text{-----} 1 + N^2 \longrightarrow 1 \text{-----} N^2$$

Reagent Group 1 Reagent Group 2

Choose data set: Anilines

	Structure	ID	MW	logP
2		317674380	206	
1		107851838	212	

Displaying 42 of 91 rows Display suitable reagents only

Selection Criteria

Set... Clear

Suitable reagents: 42
Reagents matching multiple sites: 19

Use one site
 Use all sites
 Use none
 Choose sites: Select...

< Back **Next >** Cancel

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.

Reaction-Based Library Enumeration

Control Output

Estimated library size: 2,185

Remove compounds with undesirable structural features Select Features...

Generate complete library Select subset

Selection Method

Property Select compounds with High Intravenous CNS Scoring Profile
 Diversity Configure Diversity...
 Property and Diversity Property 1 Diversity 0
 Random

Selection Criteria

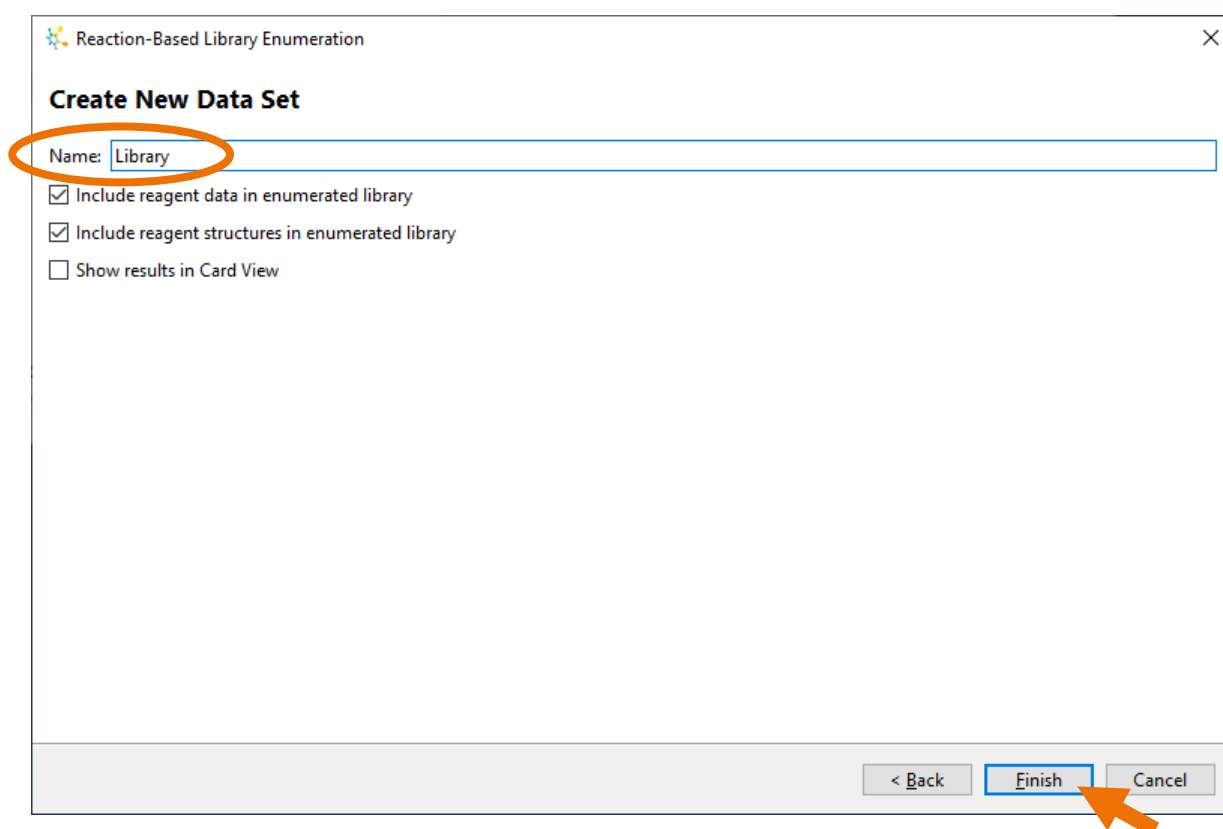
The best 100 compounds
 The best 50 % of compounds
 Compounds with values higher than 0

Allow duplicate products

< Back **Next >** Cancel

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.



Reaction-Based Library Enumeration

Create New Data Set

Name:

Include reagent data in enumerated library

Include reagent structures in enumerated library

Show results in Card View

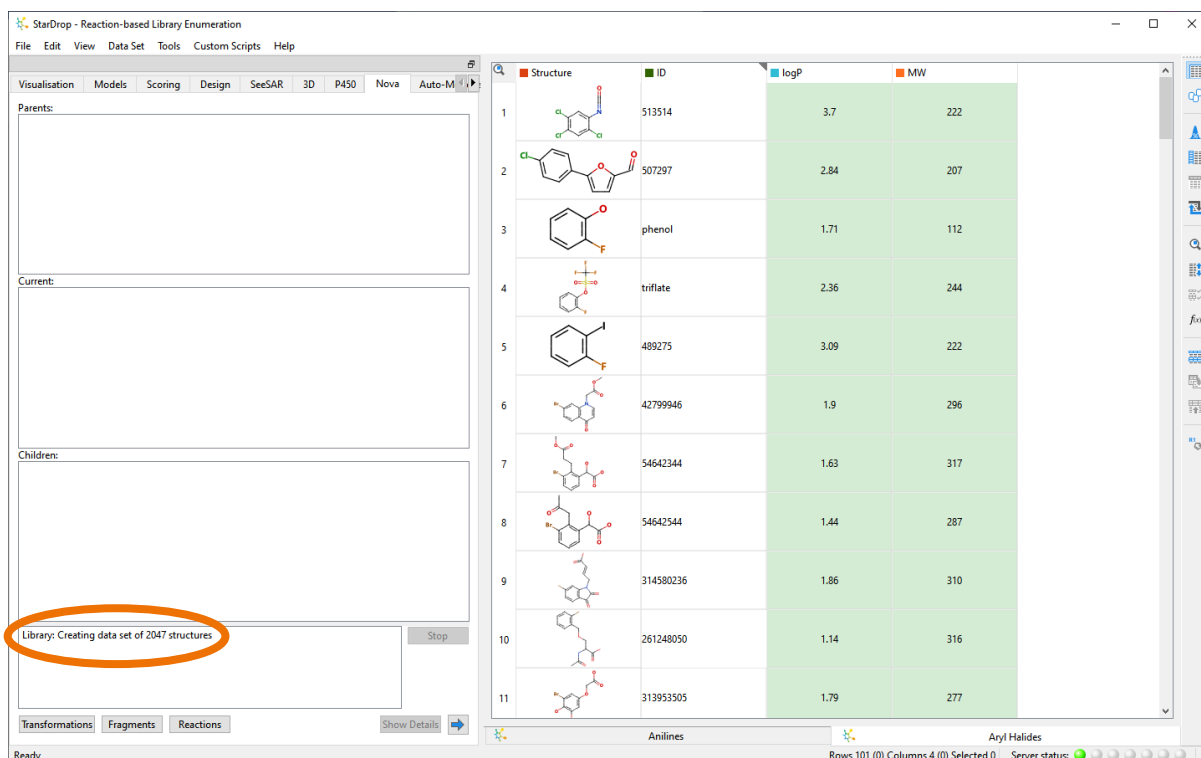
< Back **Finish** Cancel

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

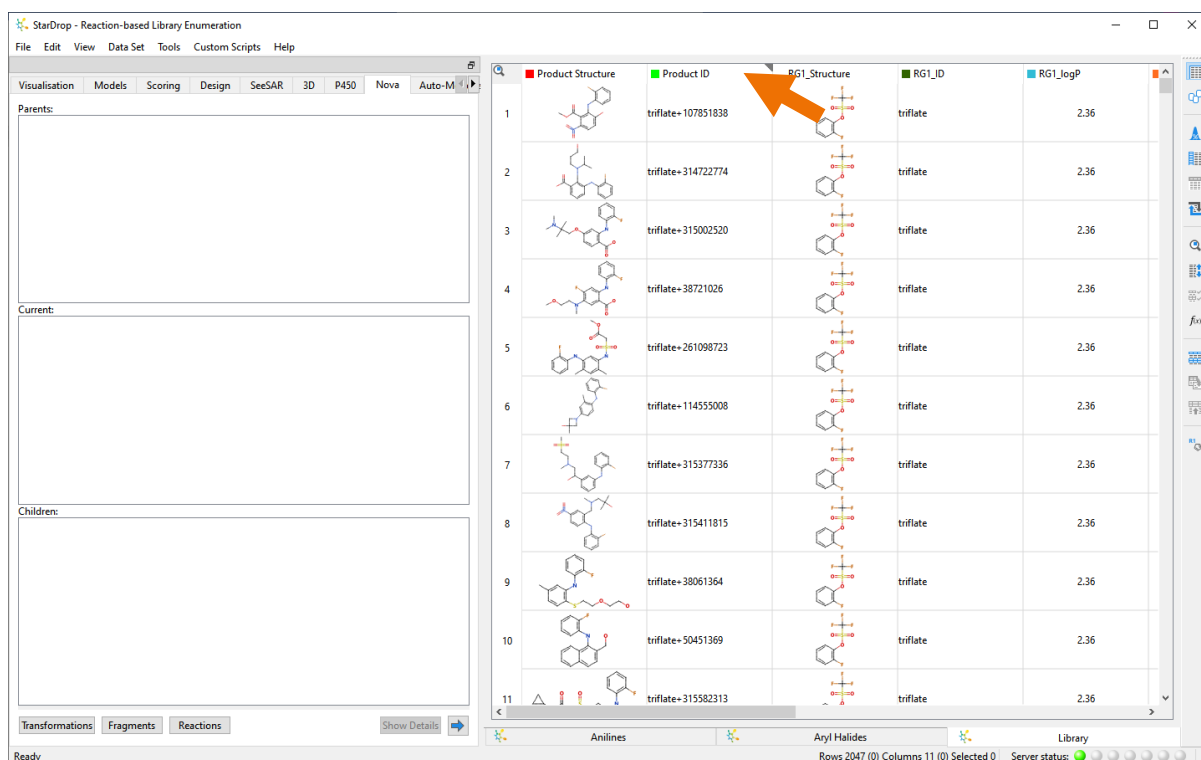
By default, all data associated with reagents, as well as their structures, will be included in the new library. If you are generating a library with many thousands of products, then you might wish to untick the **Include reagent structures in enumerated library** option, because for large sets the additional structures in every row may result in 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_, RG2_, etc.

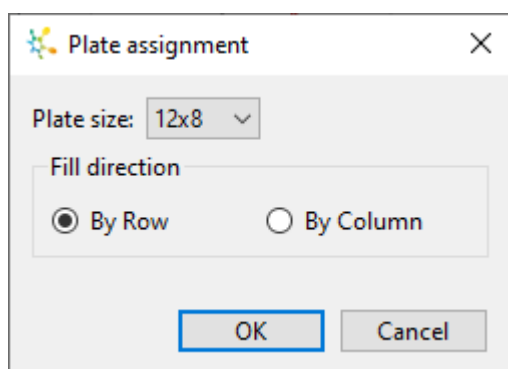
To prepare this data set for synthesis, we will assign plate IDs to all the products.

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

The screenshot shows the StarDrop interface. The 'Data Set' menu is open, and 'Assign to Plate...' is highlighted with an orange arrow. The main window displays a table of chemical products with the following columns: Product Structure, Product ID, RG1_Structure, RG1_ID, and RG1_LogP. The table contains 11 rows of data, each with a chemical structure, a product ID (e.g., triflate-107851838), a chemical structure (RG1_Structure), a plate ID (RG1_ID, all 'triflate'), and a logP value (all 2.36).

	Product Structure	Product ID	RG1_Structure	RG1_ID	RG1_LogP
1		triflate-107851838		triflate	2.36
2		triflate-314722774		triflate	2.36
3		triflate-315002520		triflate	2.36
4		triflate-38721026		triflate	2.36
5		triflate-261098723		triflate	2.36
6		triflate-114559008		triflate	2.36
7		triflate-315377336		triflate	2.36
8		triflate-315411815		triflate	2.36
9		triflate-38061364		triflate	2.36
10		triflate-50451369		triflate	2.36
11		triflate-315582313		triflate	2.36

The dialogue 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).

StarDrop - Reaction-based Library Enumeration

File Edit View Data Set Tools Custom Scripts Help

Visualisation Models Scoring Design SeeSAR 3D P450 Nova Auto-M

Parents:

Current:

Children:

Transformations Fragments Reactions Show Details

	RG2_ID	RG2_MW	RG2_LogP	RG2_In-house	Plate_assignment
1	107851838	212	1.14	Yes	Plate: 1; Row: A; Column: 1
2	314722774	252	0.848	Yes	Plate: 1; Row: A; Column: 2
3	315002520	252	0.67	Yes	Plate: 1; Row: A; Column: 3
4	38721026	242	0.275	Yes	Plate: 1; Row: A; Column: 4
5	261098723	272	0.86	Yes	Plate: 1; Row: A; Column: 5
6	114555008	192	1.1	Yes	Plate: 1; Row: A; Column: 6
7	315377336	272	0.109	Yes	Plate: 1; Row: A; Column: 7
8	315411815	253	1.73	Yes	Plate: 1; Row: A; Column: 8
9	38061364	227	1.27	Yes	Plate: 1; Row: A; Column: 9
10	50451369	173	1.27	Yes	Plate: 1; Row: A; Column: 10
11	315582313	266	0.605	Yes	Plate: 1; Row: A; Column: 11

Ready Anilines Aryl Halides Library Rows 2047 (0) Columns 12 (0) Selected 0 Server status

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