

ChEMBL Plug-in for StarDrop

Introduction

ChEMBL (<https://www.ebi.ac.uk/chembl/>) is probably the foremost manually curated database of bioactive molecules with drug-like properties. It brings together chemical, bioactivity and genomic data to aid the translation of genomic information into effective new drugs.

This script enables you to search ChEMBL directly from StarDrop, returning the data pre-formatted and ready to use for tasks such as:

- Identifying potential off-target activities for your compounds
- Understanding SAR around your compounds or targets
- Building QSAR models to make predictions for new compounds

Please note that any compound structures and target information used in a query will be passed to the ChEMBL servers. These data will be sent using a secure connection (i.e. encrypted and transmitted via HTTPS) and you can view the [ChEMBL Data Protection Policy](#).

Version

This script is for StarDrop 6.5 or more recent version for either Windows® or Mac®. To find out which version you have installed, start StarDrop and select the **Help->About** menu

Installation

Windows

To install this script on Windows:

- Ensure that you have saved your work and close StarDrop
- For 64-bit Windows, download the file [StarDrop Script - ChEMBL \(64-bit\).exe](#)
- When the download is complete, double-click the file to run it and follow the instructions (we recommend that you accept the default options provided)
- Restart StarDrop

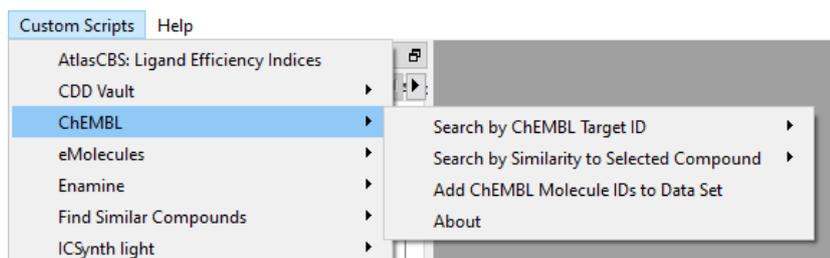
Mac OSX

To install this script on Mac OSX:

- Ensure that you have saved your work and close StarDrop
- Download the file [StarDrop Script - ChEMBL \(OSX\).zip](#)
- In your user area, navigate to the StarDrop folder (~StarDrop/)
- Extract the contents of the zip file into this folder overwriting any existing files
- Restart StarDrop

Using the ChEMBL Plug-in

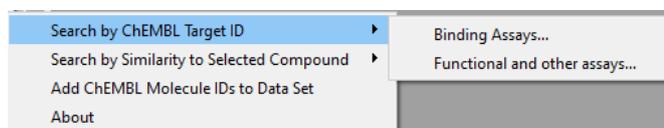
After following the installation instructions and restarting StarDrop, you will find a menu item **ChEMBL** under the **Custom Scripts** menu in StarDrop. This will give you access to three approaches you can use to search ChEMBL.



Search by ChEMBL Target ID

This option enables you to search for compound activity data associated with a given target. The ChEMBL Target ID is a unique identifier given to each molecular target in the ChEMBL database and you can search for a target in the [ChEMBL web interface](#) to find the ChEMBL Target ID.

The types of assay in which you can search are divided into two broad categories: “Binding Assays” and “Functional and other assays”. For more information on assay types, please see the notes below.



After choosing the assay type, you can specify a list of up to three ChEMBL Target IDs and click the **OK button** to begin the search.

The search will return a data set in which the first two columns display the structure and ChEMBL Molecule ID for every compound where there is a measured activity against the ChEMBL Target ID(s) entered. The target name and organism will also be displayed in the next two columns.

Following this, for **Binding assays**, aggregated and filtered pChEMBL values will be displayed for each data type for which an unqualified measurement has been reported (e.g. IC₅₀, K_i, K_D). Finally, for each measurement type and unit (e.g. IC₅₀ in μM), two columns are displayed containing aggregated and filtered values (see notes below) for that specific type and unit of data along with links to the corresponding assays descriptions in ChEMBL.

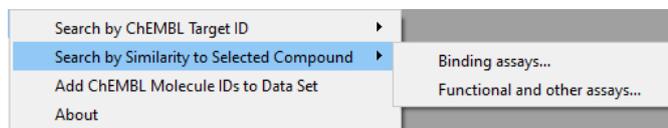
The data for **Functional and other assays** are more heterogeneous and therefore only columns containing the aggregated and filtered values for each specific type and unit of data, along with links to the corresponding assay descriptions in ChEMBL, are displayed.

These searches can sometimes take a long time to complete due to there being many results. You can stop the search at any time by closing the progress dialogue, at which point any results already retrieved will be displayed in a StarDrop data set.

Search by Similarity to Selected Compound

This option enables you to search for compounds in ChEMBL that are similar to a selected compound in your data set. Before choosing this option, please select a row within your data set.

The types of assay are again divided into two broad categories: “Binding Assays” and “Functional and other assays”. For more information on assay types, please see the notes below.



After choosing an option, you will be prompted to enter a similarity value (between 0.7 and 1). Click the **OK button** to begin the search.

Any compounds in ChEMBL with greater similarity to the selected compound than the value specified will be returned in a new data set, along with the ChEMBL Molecule ID and the similarity value. Following these will be columns summarising the measured activities against each target for which data are available for any of the similar compounds found.

For **Binding assays**, aggregated and filtered pChEMBL values will be displayed for each target for which an unqualified measurement has been reported (e.g. IC₅₀, K_i, K_D). Finally, for each measurement type and unit (e.g. IC₅₀ in µM) for each target, two columns will be displayed containing aggregated and filtered values (see notes below) for that specific type and unit of data and, along with links to the corresponding assays descriptions in ChEMBL.

The data for **Functional and other assays** are more heterogeneous and therefore only columns containing the aggregated and filtered values for each specific type and unit of data, along with links to the corresponding assay descriptions in ChEMBL, are displayed.

These searches can sometimes take a long time to complete, particularly for lower similarity values. You can stop the search at any time by closing the progress dialogue, at which point any results already retrieved will be displayed in a StarDrop data set.

Add ChEMBL Molecule IDs to Data Set

This option will search for each structure in your data set and, if the compound exists within the ChEMBL database, return the corresponding ChEMBL Molecule ID. The ChEMBL ID(s) for each compound, if present, will be added as a new column in your data set. Each ID is linked to the corresponding record on the ChEMBL web site.

Notes on Data Returned

Assay Type

When you search by ChEMBL Target ID or using a selected compound, you can choose the assay types to search, either “Binding” or “Functional and other” (all others). These assay types are summarized in the excerpt from the [ChEMBL documentation](#) below.

What is the Assay Type?

Binding (B) - Data measuring binding of compound to a molecular target, e.g. K_i, IC₅₀, K_d.

Functional (F) - Data measuring the biological effect of a compound, e.g. % cell death in a cell line, rat weight.

ADMET (A) - ADME data e.g. t_{1/2}, oral bioavailability.

Toxicity (T) - Data measuring toxicity of a compound, e.g., cytotoxicity.

Physicochemical (P) - Assays measuring physicochemical properties of the compounds in the absence of biological material e.g., chemical stability, solubility.



Unclassified (U) - A small proportion of assays cannot be classified into one of the above categories e.g., ratio of binding vs efficacy.

Filtering and Aggregation Rules

Where there are multiple bioactivity or pChEMBL values for a target of the same type (e.g. IC₅₀ or K_i) and units, the result displayed will be the arithmetic mean of the values. When averaging data which contains qualifiers (i.e. <, >, ≤, ≥), a null value will be shown unless the qualifiers are the same on each measurement.

Activity Comments

Bioactivities in ChEMBL will sometimes have an activity comment. If there is an activity comment and it is not 'active', the bioactivity value is discarded. This is because the activity comment is often something negative such as 'inactive', 'not active' or 'inconclusive'. The activity comment can also be an ID, which allows for a link to other databases. These IDs are numbers, therefore if the comment is numerical, the bioactivity value is not discarded.

Disconnected Molecules

Molecules which are disconnected, for example molecules with salt adducts, will not be recognised by the ChEMBL web server. For these molecules, you will see an error.