

ChEMBL

Accessing big molecular data via
the web interface and API

James Blackshaw

ChEMBL Scientific Data Engineer

jblackshaw@ebi.ac.uk

& Melissa Adasme

Data Mining Scientist

adasme@ebi.ac.uk

<https://www.ebi.ac.uk/chembl/>
chembl-help@ebi.ac.uk

EMBL-EBI



Hybrid workshop on in silico Drug design, Jan 30, 2022



Training structure

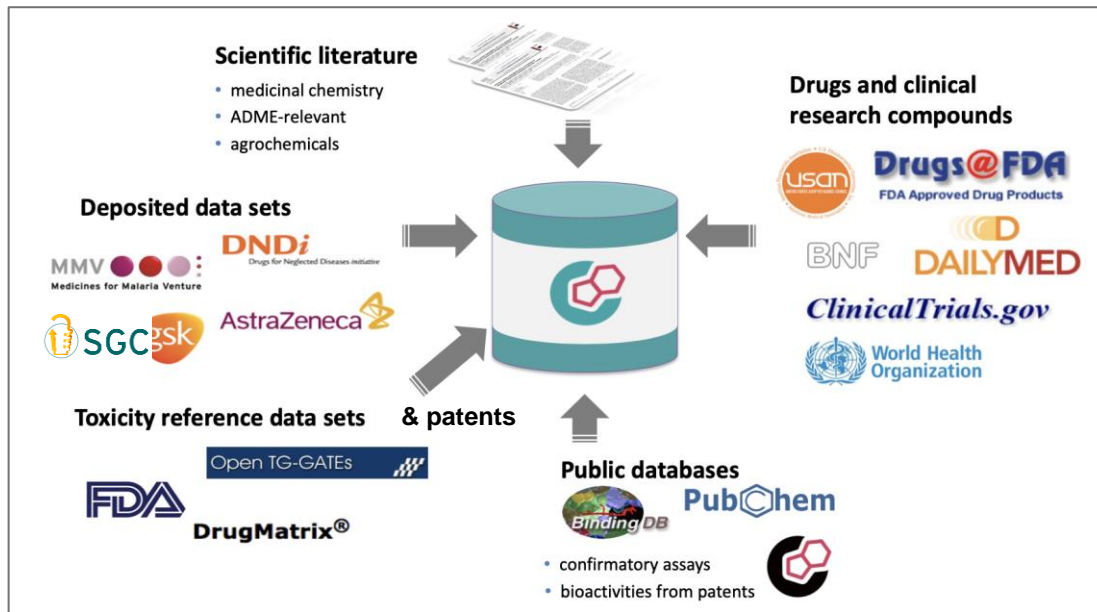
General introduction into ChEMBL (James Blackshaw) – 25 min

- What is ChEMBL and how is it structured?
- What data does ChEMBL contain?
- How are data extracted from scientific articles?
- How are the data in ChEMBL curated?
- How are drug data curated and annotated in ChEMBL?
- How can you use the ChEMBL GUI?

ChEMBL live tutorial (Melissa Adasme) – 25 min

Q&A – 10 min

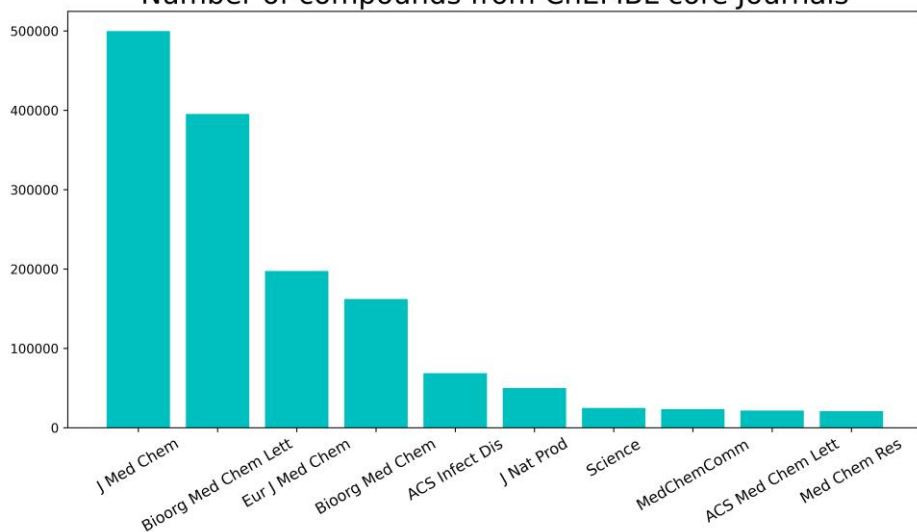
What is included in ChEMBL?



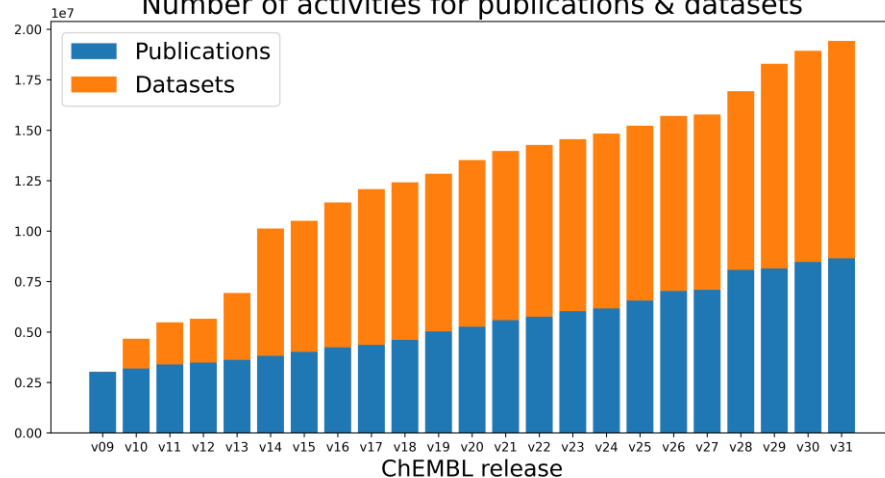
- “Open access data for drug discovery”.
- Manually curated, high quality, FAIR.
- The core of the database is the primary MedChem literature & donated data sets.
- ChEMBL covers >40 years of research (starting 1980).

Distribution of sources






Number of compounds from ChEMBL core Journals



Number of activities for publications & datasets



ChEMBL 32 Summary (Releasing in mid Feb)

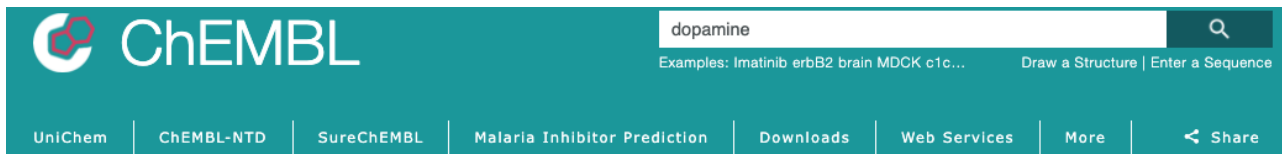
	Datasets	213
	Publications	86,364
	Targets	15,139
	Compounds	2,354,965
	Activities	20,038,828

Data statistics

Molecular properties	E.g. two-dimensional structures, calculated molecular properties (logP, molecular weight, Lipinski 'Rule of Five' parameters etc.)
Protein-target interactions	Information about how small molecules interact with their protein targets.
Cell and organism level data	Records on how these compounds affect cells and whole organisms. Information on absorption, distribution, metabolism, excretion and toxicity (ADMET) etc.
Bioactivity data	Binding constants and other activity results from deposited assays.
Curated drug and candidate set	A highly curated set of drugs and clinical candidates. These compounds are annotated with their known therapeutic targets and therapeutic indications.

ChEMBL entities

www.ebi.ac.uk/chembl



The image shows the ChEMBL search interface. At the top left is the ChEMBL logo. To its right is a search bar containing the text 'dopamine'. Below the search bar are examples: 'Examples: Imatinib erBB2 brain MDCK c1c...'. To the right of the examples are links: 'Draw a Structure | Enter a Sequence'. Below the search bar is a navigation menu with links: 'UniChem', 'ChEMBL-NTD', 'SureChEMBL', 'Malaria Inhibitor Prediction', 'Downloads', 'Web Services', 'More', and a 'Share' button with a left arrow icon.

[EBI](#) > [Databases](#) > [Chemical Biology](#) > [ChEMBL Database](#) > [Search Results](#) > dopamine

Search Results

All Results 8931 Compounds 84 Targets 54 Assays 7538 Documents 1255 Cells 0 Tissues 0

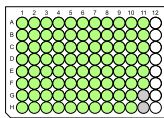
Compounds	Chemical structure and properties on which experimental data have been measured
Targets	Biological system in which bioactivity data is measured (e.g. protein, cell, organism)
Assays	The description of the specific experiment
Documents	Where the data are reported (e.g. scientific paper)
Cells	If applicable, the cell in which the bioactivity was measured
Tissues	If applicable, the tissue in which the bioactivity was measured

Literature data extraction

Reference



Assay



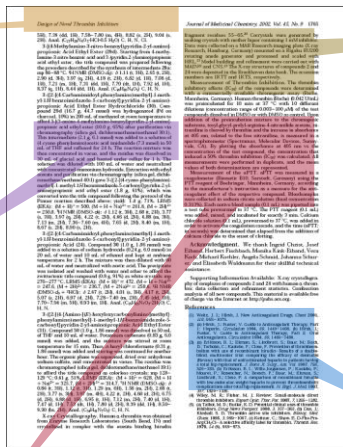
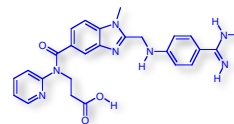
Target



Organism



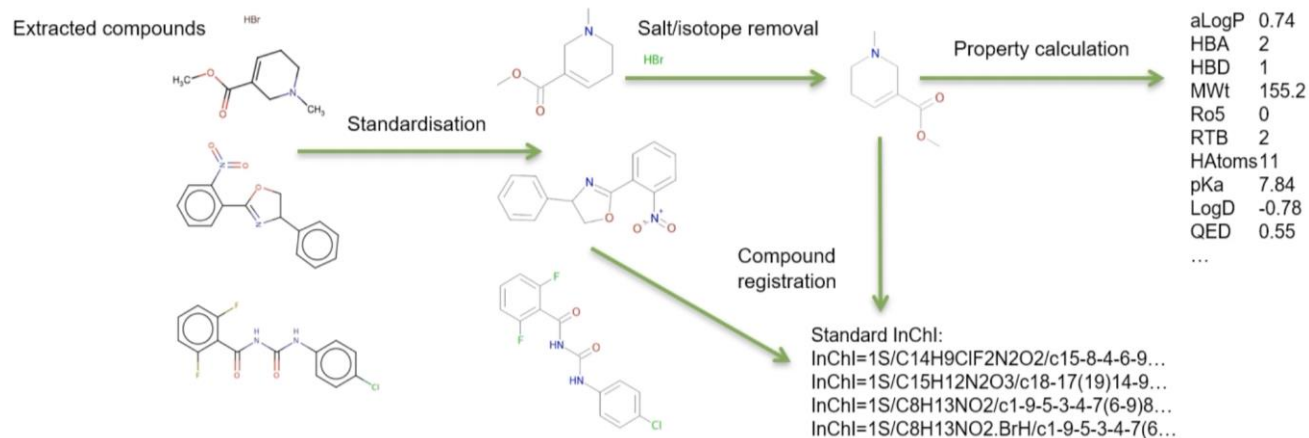
Compound



Literature data extraction

Manual extraction from the literature	Data are extracted from full text articles
Comprehensive	All quantitative bioactivity measurements are extracted
Multiple data types	These include: Target binding data Functional assays In-vivo efficacy data In-vitro and in-vivo ADME/pharmacokinetics Toxicology measurements

Chemistry curation



We standardise the structures.

[Automated](#) error checking, standardization, and salt stripping.

We calculate additional data.

Automated calculation step for compound properties.

ChEMBL curators correct any remaining issues.

Manual curation of invalid structures, name-structure mismatches, missing stereochemistry etc.



Activity curation



	ACTIVITY_ID	PCHEMBL_VALUE	STANDARD_RELATION	STANDARD_VALUE	STANDARD_UNITS	RELATION	VALUE	UNITS	DATA_VALIDITY_COMMENT
1	33501	7.57	=	27 nM		=	27 nM	(null)	
2	36800	(null)	=	19.8 ug.mL-1		=	19.8 ug mL-1	(null)	
3	859055	4.8	=	16000 nM		=	16 uM	(null)	
4	1273293	(null)	<=	100000 nM		<=	0.1 mM	(null)	
5	1699014	4.07	=	86000 nM		=	0.086 mM	(null)	

Standardised data

Depositor input

Mostly against protein targets

A large proportion of data is **dose-response measurements against protein targets**

E.g. IC50 – half maximal inhibitory concentration

pChEMBL values allow approximate comparisons

pChEMBL is calculated when:

- Relation is '='
- Units standardised to 'nM'

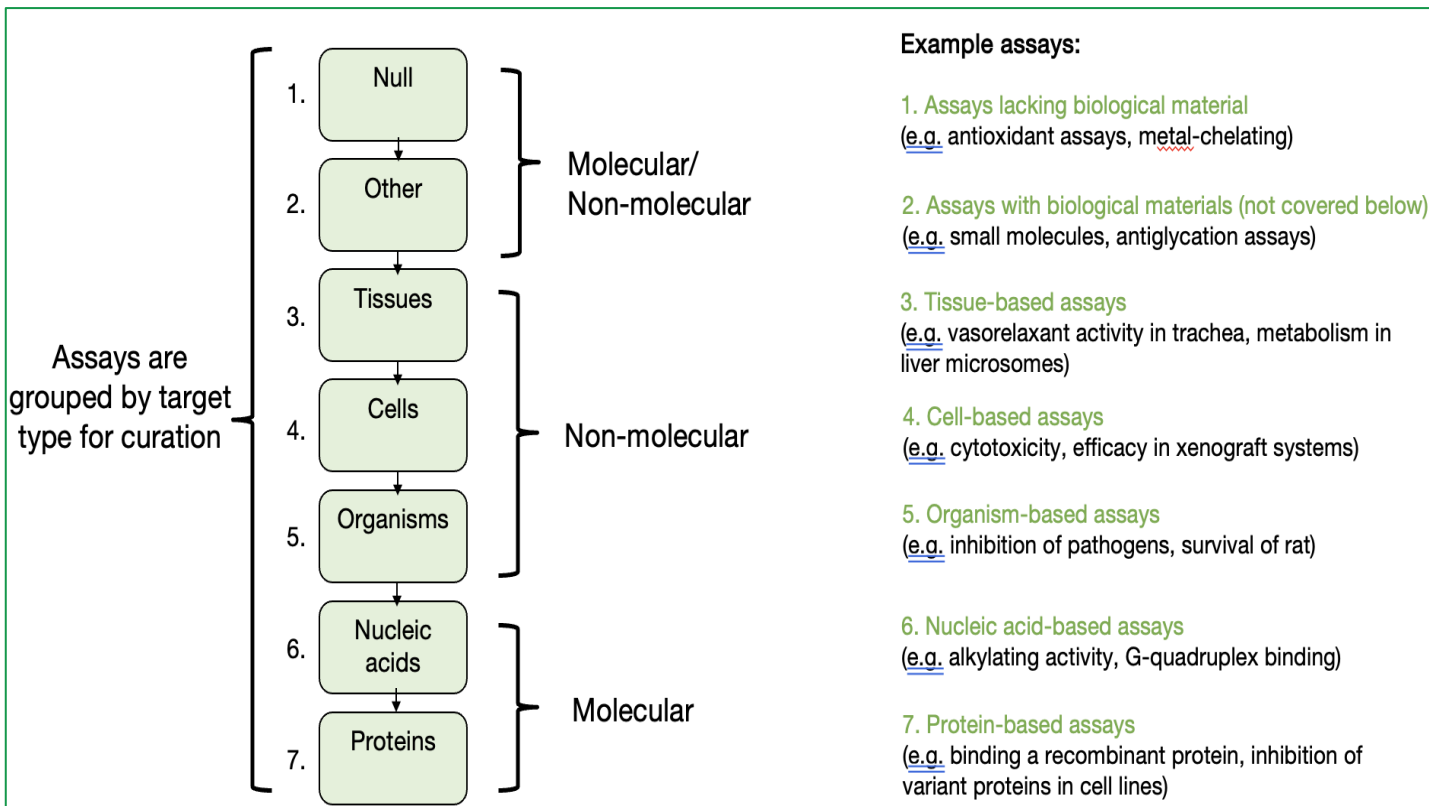
Assay Types

Assay type	Assay description	Example
F	Functional	Biological effect of a compound e.g., % cell death in a cell line, rat weight
A	ADME	ADME data e.g., t1/2, oral bioavailability
T	Toxicity	Toxicity of a compound e.g., cytotoxicity
B	Binding	Binding of a compound to a molecular target e.g., Ki, IC50, <u>Kd</u>
P	Physicochemical	Physicochemical properties of the compounds in the absence of biological material e.g., chemical stability, solubility
U	Unassigned	Cannot be classified into one of the above categories e.g., ratio of binding vs. efficacy

Bioassay curation

- Standardise activity types
e.g., IC-50, mean IC50, IC(50) -> IC50
- Antilog values
e.g., pKi -> Ki, -log IC50 -> IC50
- Standardise values/units
e.g., uM, 10⁻⁶ mol/L, pmol/mL -> nM
- Flag potentially incorrect data
e.g., out of range, non-standard units
- Calculate pChEMBL values
-log₁₀ (molar IC50, XC50, EC50, AC50, Ki, Kd or Potency)
- Flag potential duplicates
e.g., same compound, target, type & value
- Manual curation to fix erroneous data

Bioassay curation



Bioassay curation

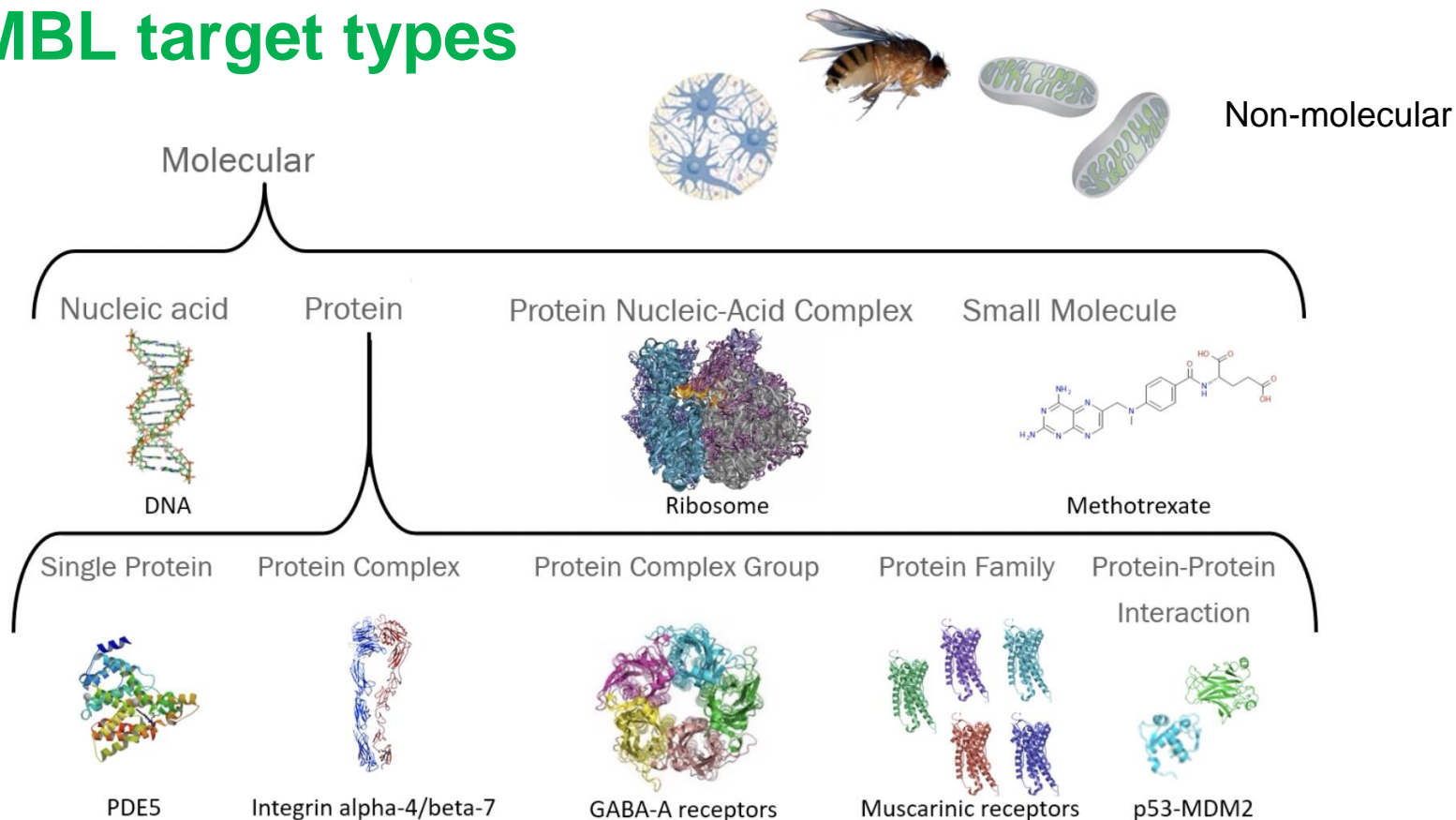
Curation Method

- **Target types** and **assay types** are both extracted from [assay description](#)
- Correct and standardize names for tissues, cell lines etc by using controlled vocabularies (e.g. UBERON, Cellosaurus)
- Assay formats are mapped to BAO ontology

Common curation tasks for the Assay type:

1. Curation of assay type 'U'
2. Re-mapping of toxicity assays from 'A' to 'T'
3. Re-mapping of GPCR assays measuring secondary messengers from 'B' to 'F'

ChEMBL target types



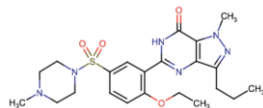
ChEMBL 31: ~8M bioactivities on protein targets; ~12M bioactivities on non-protein targets

Drugs and clinical candidates

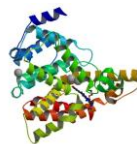
- Clinical Trials.gov
- DailyMed
- USAN applications
- FDA Orangebook
- ATC classification
- BNF
- INN applications
- FDA new drug approvals
- Withdrawn drugs
- Pharma pipelines



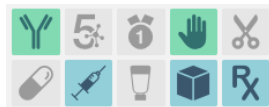
Disease annotation



Name & structure mapping



Mechanism annotation

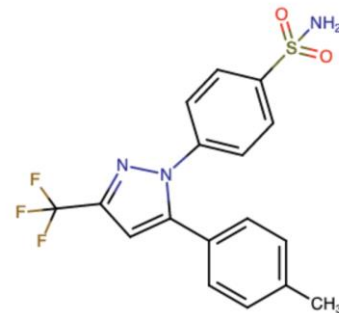


Drug property annotation

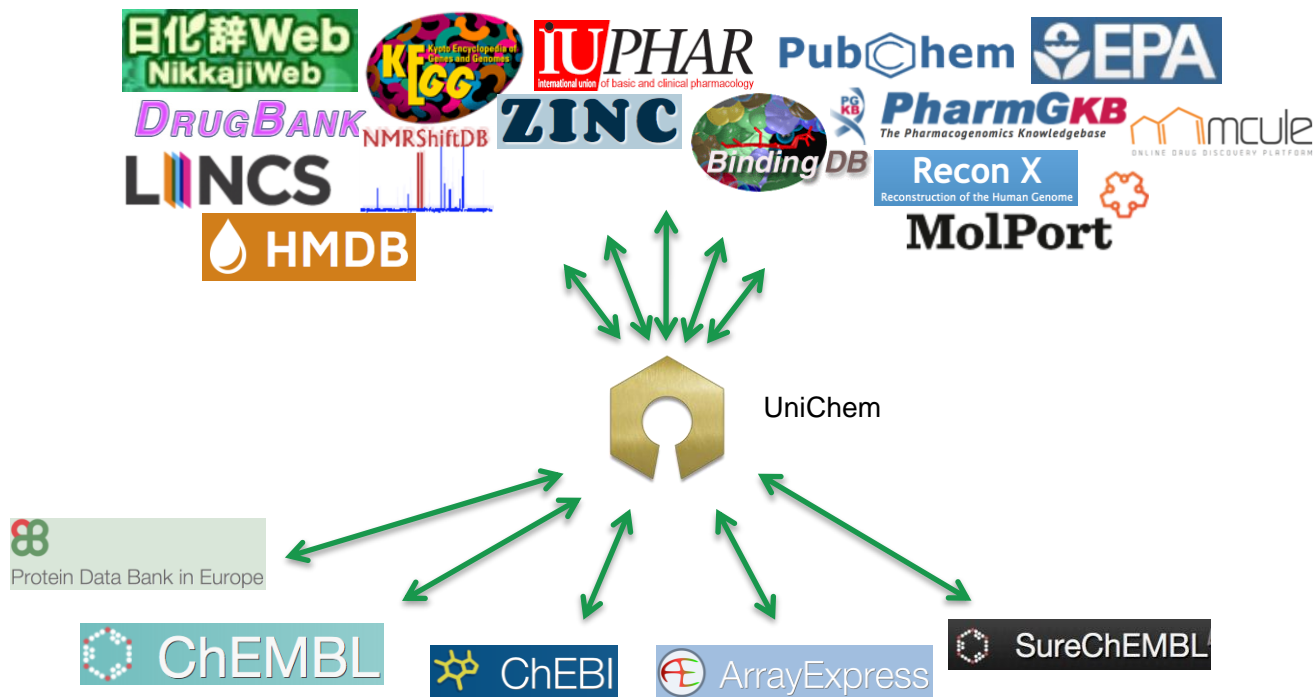


Drug properties

Molecule-based properties are assigned to all clinical-stage compounds (green icons)	<ul style="list-style-type: none">• Molecule type (small molecule, antibody, protein, oligonucleotide, oligosaccharide etc.)• Rule-of-five compliant• First in class• Chiral/racemic/achiral• Prodrug
Product-based properties are assigned to approved drugs (blue icons)	<ul style="list-style-type: none">• Oral/parenteral/topical administration• Black box warning/withdrawn• Availability type (prescription only, over the counter, discontinued, withdrawn)



UniChem generates ChEMBL linkouts



- UniChem maps database identifiers and structures for 170 million compounds from 40 different source databases

Cross-references from UniChem

Results rendered in
ChEMBL interface

```
---
aux_for_url: 0
base_id_url: https://www.ebi.ac.uk/chembl/db/compound/inspect/
base_id_url_available: 1
description: A database of bioactive drug-like small molecules and bioactivities abstracted from
name: chembl
name_label: ChEMBL
name_long: ChEMBL
src_compound_id:
  - CHEMBL490
src_id: 1
src_url: https://www.ebi.ac.uk/chembi/
---
aux_for_url: 0
base_id_url: http://www.drugbank.ca/drugs/
base_id_url_available: 1
description: A database that combines drug (i.e. chemical, pharmacological and pharmaceutical)
name: drugbank
name_label: DrugBank
name_long: DrugBank
src_compound_id:
  - DB00715
src_id: 2
src_url: http://drugbank.ca/
---
aux_for_url: 0
base_id_url: http://www.ebi.ac.uk/pdbe-srv/pdbechem/chemicalCompound/show/
base_id_url_available: 1
description: The European resource for the collection, organisation and dissemination of data on
name: pdbe
name_label: PDBe
name_long: PDBe (Protein Data Bank Europe)
src_compound_id:
  - 8PR
src_id: 3
src_url: http://www.ebi.ac.uk/pdbe/
---
aux_for_url: 0
base_id_url: http://www.guidetopharmacology.org/GRAC/LigandDisplayForward?ligandId=
base_id_url_available: 1
description: The IUPHAR (International Union of Basic and Clinical Pharmacology)/BPS (British Pharmacological Society)
name: gtpdb
name_label: Guide to Pharmacology
name_long: Guide to Pharmacology
src_compound_id:
  - 4790
src_id: 4
src_url: http://www.guidetopharmacology.org
---
```

Programmatic output
e.g., JSON, XML

Link from ChEMBL to
SureChEMBL

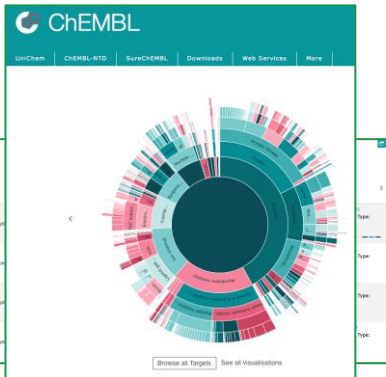
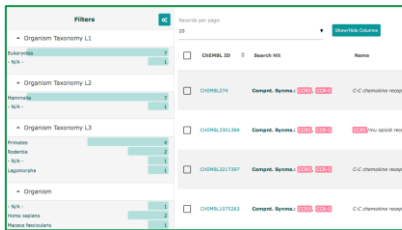
UniChem Cross References

New [View the UniChem Connectivity matches for ChEMBL490](#)

ACToR	110429-35-1 61869-08-7
Atlas	paroxetine
BindingDB	50331515 22416
ChEBI	7936
DrugBank	DB00715
DrugCentral	2068
eMolecules	1935652
EPA CompTox Dashboard	DTXSID3023425
FDA SRS	41VRH5220H
Guide to Pharmacology	4790
Human Metabolome Database	HMDB14853
IBM Patent System	BA9CB78987006F7DD38940BBB3C63F01
KEGG Ligand	C07415
LINCS	LSM-2843
Nikkaji	J259.859K
PDBe	8PR
PharmGKB	PA450801
PubChem	43815
PubChem: Drugs of the Future	12013591
PubChem: Thomson Pharma	14924270 14850948
SureChEMBL	SCHEMBL27799
ZINC	ZINC000000527386

Access to ChEMBL data

Website



Web Services

ChEMBL web services API live c

Version: 2.7.6

[chembl/api/data/activity/](https://chembl.org/api/data/activity/)

GET [chembl/api/data/activity/{ID}](#)

QUESTION Why is the H^+ concentration in the stomach so high?

Get <https://data.activity.sno.ac>

chembl/api/data/activity/search/

GET [chembl/ap/data/assay/](#)

GET [chembl/api/data/assay/:ID](#)

GET [chembl/api/data/assay/set/:IDs_](https://chembl.org/api/data/assay/set/:IDs_)

 [chembl/api/data/assay/search?](https://chembl.org/api/data/assay/search?)

GET [chembl/api/data/atc_class/](https://chembl.org/api/data/atc_class/)

GET [chembl/api/data/atc_class/:ID](#)

GET chembl/api/data/atc_class/set/1

100%

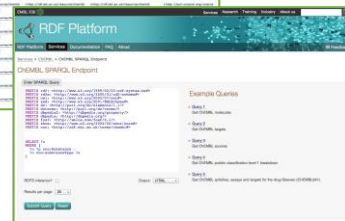
[illegible]

Semantic Web



PostgreSQL

Downloads



How should I access ChEMBL?

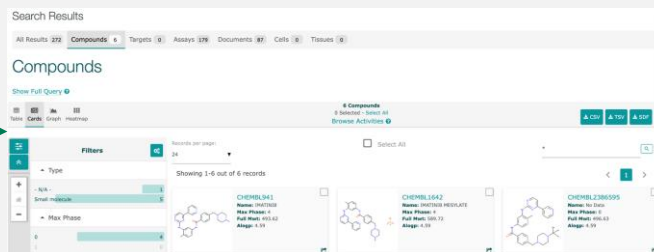
Use <u>API</u> if:	Need a repeatable query, or embed in a script? Don't want to download full database Want easy access to data via url weblink?
Use <u>Web Interface</u> if:	Only need to access small amounts of data, e.g. for one compound or one target For initial or exploratory investigation
Download <u>Database</u> if:	Need to run multiple queries with much input data? Have complex, bespoke queries across multiple tables? (Construct an sql query) Need data from a previous release? (API is for the current release only)

Web Interface Structure

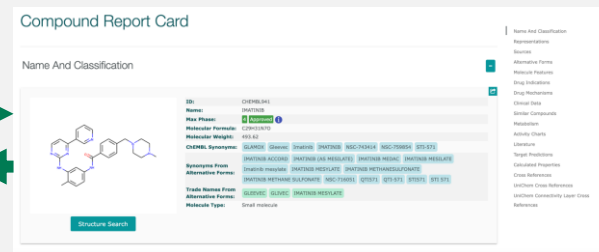
Landing Page



“Browse Items” Pages



Report Card Pages



Search by:

- Free Text.
- Chemical Structure.
- Identifiers.
- BLAST.

Explore all items in compounds, assays, etc.

- Explore datasets.
- See data distribution.
- Apply filters.
- See relationships among entities.

- See the details of one item.
- Start a structure search from the item (compounds).
- See the relationship with other entities.

Free Text Search

Enter a text term on the search bar

Results per entity.

View data in different presentations.

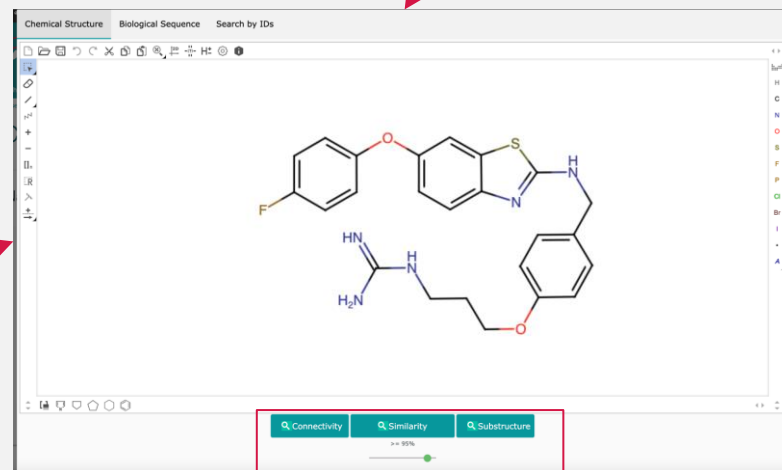
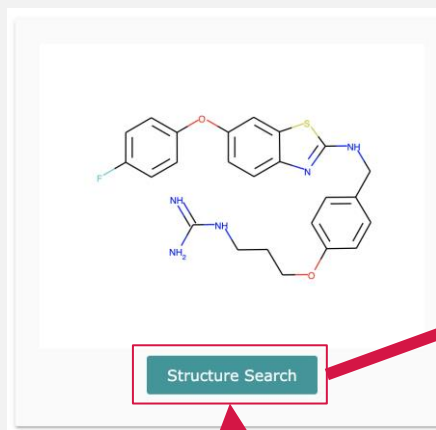
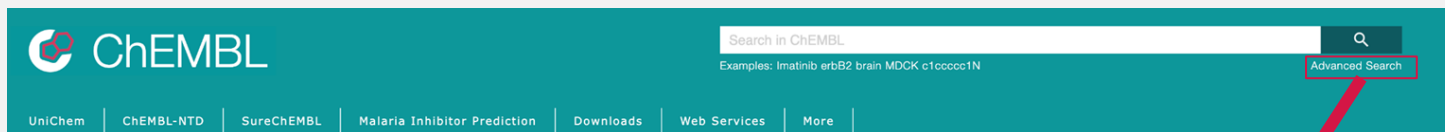
See distribution of data and apply filters.

Specific Items

The screenshot displays the ChEMBL Free Text Search interface. At the top, the ChEMBL logo is on the left, and a search bar on the right contains the text 'Imatinib'. Below the search bar, a navigation bar includes links for UniChem, ChEMBL-NTD, SureChEMBL, Malaria Inhibitor Prediction, Downloads, Web Services, and More. The main content area shows 'Search Results' for 'Imatinib', with tabs for All Results (272), Compounds (6), Targets (0), Assays (179), Documents (87), Cells (0), and Tissues (0). The 'Compounds' tab is selected, showing 6 compounds. Below this, there are options to 'Show Full Query' and buttons for 'Table', 'Cards', 'Graph', and 'Heatmap'. The 'Cards' view is active, displaying a list of compounds with their chemical structures and key properties. A sidebar on the left shows filters for 'Type' and 'Max Phase'. The 'Type' filter shows a distribution of 'Small molecule' (5) and 'N/A' (1). The 'Max Phase' filter shows a distribution of '0' (4), '1' (0), and '2' (0). The 'Cards' view shows three specific items: CHEMBL941 (Name: IMATINIB, Max Phase: 4, Full Mwt: 493.62, Alogp: 4.59), CHEMBL1642 (Name: IMATINIB MESYLATE, Max Phase: 4, Full Mwt: 589.72, Alogp: 4.59), and CHEMBL2386595 (Name: No Data, Max Phase: 0, Full Mwt: 496.63, Alogp: 4.59).

Chemical ID	Name	Max Phase	Full Mwt	Alogp
CHEMBL941	IMATINIB	4	493.62	4.59
CHEMBL1642	IMATINIB MESYLATE	4	589.72	4.59
CHEMBL2386595	No Data	0	496.63	4.59

Structure Search



https://www.ebi.ac.uk/chembl/compound_report_card/CHEMBL4301448/

Structure Search - Result

Edit search and start a new one.

Status of the search.

- Similarity and connectivity are immediate.
- Substructure is run by a job in LSF.

Similarity Search Results

Threshold: 40%
Query: N=C(N)NCCCCC1ccc(CN2nc3ccc(Oc4ccc(F)cc4)cc3a2)cc1

Status: Results Ready

Edit Search

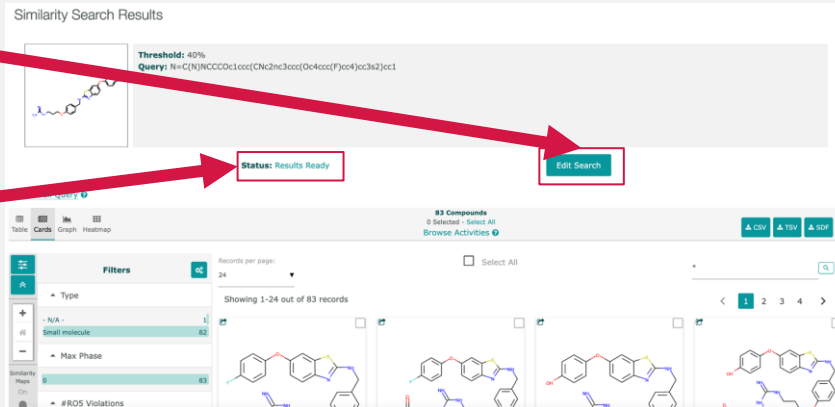
83 Compounds
0 Selected - Select All
Browse Activities

Table Cards Graph Heatmap

Filters

- Type
- N/A
- Small molecule
- Max Phase
- #ROS Violations

Showing 1-24 out of 83 records



Similarity Maps

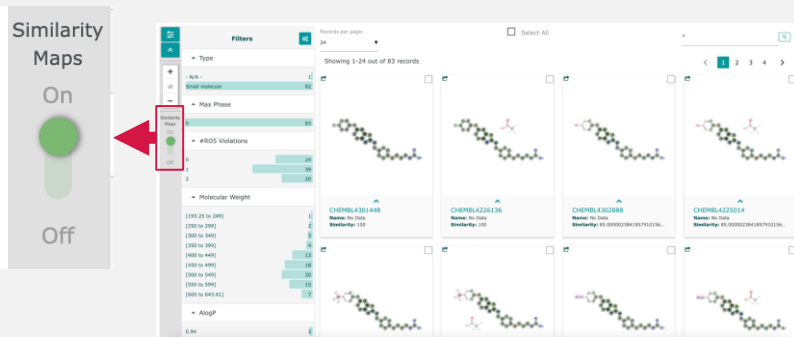
On

Off

Filters

- Type
- N/A
- Small molecule
- Max Phase
- #ROS Violations
- Molecular Weight
- Align

Showing 1-24 out of 83 records



Structure highlights and similarity maps.

Highlight

On

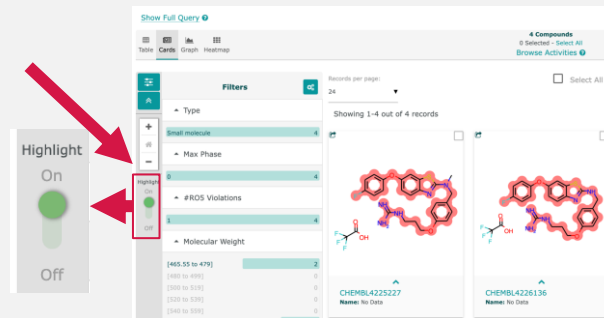
Off

Show Full Query

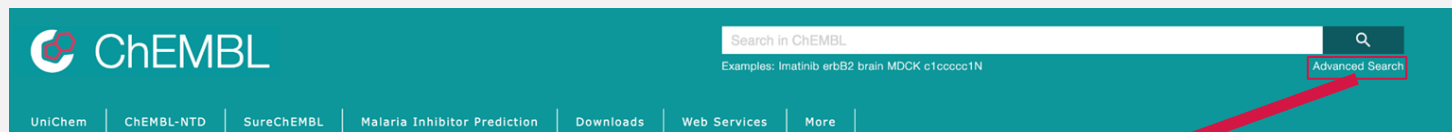
Filters

- Type
- Small molecule
- Max Phase
- #ROS Violations
- Molecular Weight

Showing 1-4 out of 4 records



Search by IDs



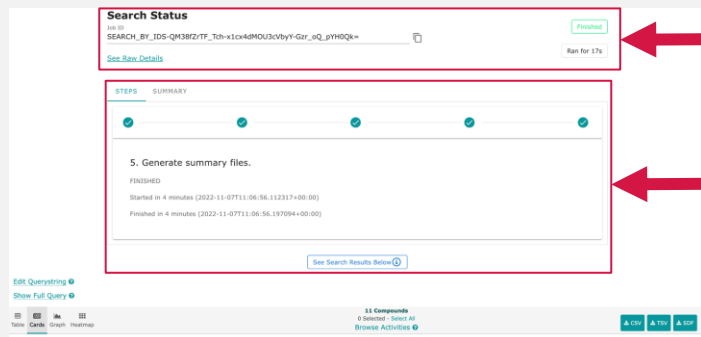
Select source and destination.

- From Molecule ChEMBL IDs to Compounds.
- From Target ChEMBL IDs to Targets.

Paste IDs or upload a file.

The image shows the 'Search by IDs' interface on the ChEMBL website. It is divided into two main sections: '1. Select the type of search:' and '2. Enter the identifiers:'. In the first section, there are two dropdown menus: 'From:' set to 'Molecule ChEMBL IDs' and 'To:' set to 'ChEMBL Molecules'. Below these are links: 'Provide Molecule ChEMBL IDs below' and 'Get ChEMBL Molecules'. In the second section, there are two tabs: 'PASTE IDS' (selected) and 'UPLOAD A FILE'. Under 'PASTE IDS', there is a 'Separator:' dropdown set to 'Automatic' and a '10 Items Identified' button. Below this is a text area containing a list of ChEMBL IDs: 'CHEMBL1905569, CHEMBL3659411, CHEMBL4104861, CHEMBL1502197, CHEMBL3931303, CHEMBL466111, CHEMBL323959, CHEMBL1255073, CHEMBL4288214, CHEMBL3480964'. At the bottom, there are two buttons: 'USE SAMPLE IDS' and 'SEARCH'.

Search by IDs



Search Status

Job ID: SEARCH_BY_IDS-QH3821TF_Tch-x1c48MDU3cVbYf-Gzr_iQ_pYH0Qk=

Finished

Run for 17s

[See Raw Details](#)

STEPS SUMMARY

5. Generate summary files.

FINISHED

Started in 4 minutes (2022-11-07T11:06:56.112317+00:00)

Finished in 4 minutes (2022-11-07T11:06:56.197094+00:00)

[See Search Results Below](#)

11 Compounds

0 Selected - Select All

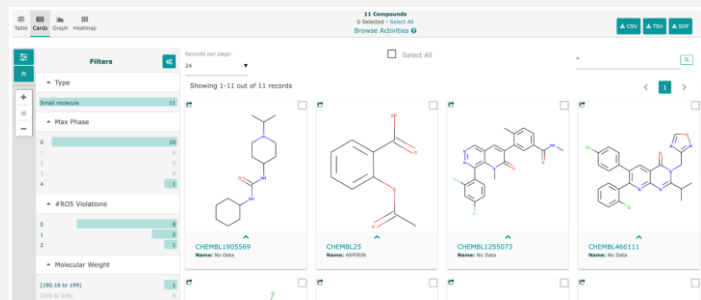
[Browse Activities](#)

[A Ctrl](#) [A Tab](#) [A Shift](#)

See the status of the search. It is run by a job in LSF.

Details of the process.

Details of the results.



Filters

Records per page: 24

Select All

Showing 1-11 out of 11 records

Small molecule

Max Phase

#RDS Violations

Molecular Weight

11 Compounds

0 Selected - Select All

[Browse Activities](#)

[A Ctrl](#) [A Tab](#) [A Shift](#)

CHMBL1905569

Name: No Data

CHMBL25

Name: 657606

CHMBL255072

Name: No Data

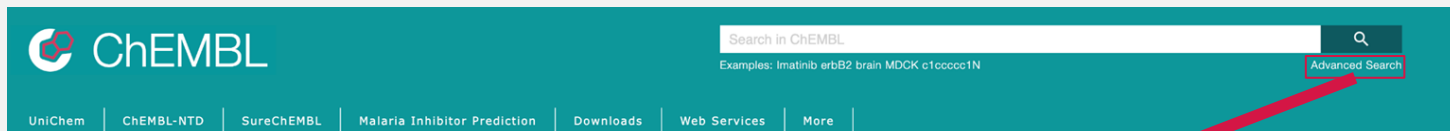
CHMBL466111

Name: No Data

Explore all results as usual.

STEPS SUMMARY		
Input		
Source Entity	MOLECULE_CHEMBL_IDS	Type of the items entered as input.
Destination entity	CHEMBL_COMPOUNDS	Type of the items mapped in ChEMBL from the input.
Separator	__AUTO__	Character used to separate the items.
Total items entered	11	Number of items entered in the input.
Unique items identified	11	Number of items identified after removing duplicates.
Items with match	11	Number of items that matched in ChEMBL.
Output		
Items included in results	11	Number of items included in the final results below.
Inactive ChEMBL IDs	0	Number of ChEMBL IDs that are inactive. They have not been included in the latest release.
Obsolete ChEMBL IDs	0	Number of ChEMBL IDs that are obsolete. They have been permanently deleted from ChEMBL.
ChEMBL IDs of a different entity	0	Number of ChEMBL IDs that belong to a different entity. They are not included in the results.
DOWNLOAD MATCHES FILE		

BLAST Search



Uses the EBI BLAST Search.

- Enter a sequence in any format supported by the EBI BLAST service.

Modify some parameters

The screenshot shows the ChEMBL BLAST search interface. At the top, there are three tabs: "Chemical Structure", "Biological Sequence" (which is selected), and "Search by IDs". Below the tabs, there are three numbered steps:

- 1. Enter or paste a sequence in any supported format:** This step is highlighted with a red box. It contains a large text area with a protein sequence:

```
>sp|P35858|HUMAN Insulin-like growth factor-binding protein complex acid labile subunit OS=Homo sapiens GN=IGFALS PE=1 SV=1
MALRRGGALALLSSWVAGPRSGLEGADPGTPEAGEGACPAACVCSYDDDADELSVFC
SSRNLTLPDGVPGGTQALWLDGNNLSSVPPAFAFNLSLGLNLQGGQLGSLFQALLG
LENLCHLERNQSLALGTFAFTPALASLSLNRLSRLEGLGFEGLGSLWDLKGNW
SLAVLPDAAFRGLGSLRELVIAGNRLAYLPALFSLGAELELDLSRNALRAIKANFVQ
LPRLLQKYLDRNLIAAVAPAGFLGLKALRWLDSLHNRVAGLLEDTFPGLLGLRLRLSHN
ATASLRPRTFKDLHFLEELQGHNRIRQLAERSFEGGLQLEVLTDHNLQLEVKAGAFGL
LTNAVMMNLSGNCLRNLPQVFRGLGKLSHLHLEGSCGRIRPHTFTGLSGLRRLFKDN
GLVGIEEQSLWGLAELELDLTSNQLTHPLHRLFQGLGKLEYLLSRNRLAELPADALGP
LQRAFVLDVSHNRLEALPNLSLAPLGRRLYLSRNNSLRITFTKPPGLERLWLEGNPWDC
GCPLKALRDFALQNPFAVPRFVQAICEGDDCQPPATYNNITCASPEVWGLDRDLSEA
HFAPC
```

 Below the text area are three links: "Use a example sequence", "Clear sequence", and "Upload a file".
- 2. Modify the parameters for **blastp** (optional):** This step is highlighted with a red box. It contains a link "Show Parameters".
- 3. Search in ChEMBL** This step is highlighted with a red box. It contains a button "Search in ChEMBL".

BLAST Search

Query:

```
>sp|P35858|ALS_HUMAN Insulin-like growth factor-binding protein complex acid labile subunit OS=Homo sapiens GN=IGFALS PE=1 SV=1
MALRKGGALALLLSWVALGPRSLLEGADPGTPGEAEGPACPAACVCSYDDDADELSVFC SSRNLTRLPGVPGGTQALWLDGNNLSSVPAAFQNLSSLGFLNLQGGQLGSLEPQALLG
LENLCHLHLERNQLRSLALGTFHTPALASGLSNNRLSRLEDGLFEGLSLWDLNLGWN SLAVLPDAAFRGLGSLRELVLAGNRLAYLQPAIFSGLAELRELDLSRNALRAIKANVFVQ
LPRLQKLYLDRNLIAAVAPAGFLGLKALRWLDLSHNRVAGLEDFTFPGLSLRLRLSHN AIASLRPRTFKDLHFLEELQLGHNRRQLAERSFEGLQGQLEVLTDHNLQEVKAGAFGLG
LTVNAVNNLSGNCRLNLPQVFRGLGKLHLSLHLEGSCLGRIPHTFTFPGLSLRLRLFKDN GLVGIEEQSLWGLAELELDLTSNQLTHLPHRLFGQLGKLEYLLSRNLRAELRAD
LQRAFWLDVSHNRLEALPNSLLAPLGRRLSLRNNSLRTFTTQPPGLERLWLEGNPWDC GCPLKALRDFALQNPSAVPRFVQAICEGDDCOPPAYTYNNITCAVYVGLDRLDLSEA HFAPC
```

This page only shows the best alignment per each hit in the results. Click [here](#) to view the job submission to the EBI BLAST Search.

Status: **Results Ready** [Edit Search](#)

[Show Full Query](#)

88 Targets
0 Selected - Select All
[Browse Activities](#)

Records per page: 20 [Show/Hide Columns](#)

Showing 1-20 out of 88 records

<input type="checkbox"/>	E-Value	Positives %	Identities %	Score (bits)	Score	Length	ChEMBL ID	Name	UniProt Accessions	Type	Organism	Compounds
<input type="checkbox"/>	5.2e-54	44.5	32.6	192.971	489	581	CHEMBL4295907	Leucine-rich repeat-containing protein 15	Q8TF66	SINGLE PROTEIN	Homo sapiens	<div>1</div> By Mol. Wt.: <div></div>

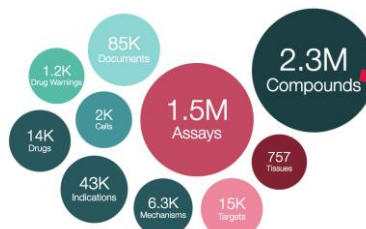
Go to EBI - BLAST search results

Edit Search

See progress of Search

Results in Targets

'Browse' Pages



View data in different ways.

See data distribution and filter by several properties.

Type	
- N/A -	369155
Antibody	974
Cell	47
Enzyme	118
Gene	77
Oligonucleotide	170
Oligosaccharide	92
Protein	22682
Small molecule	1920366
Unclassified	4
Unknown	18015

Browse Compounds

Edit Querystring

Show Full Query

Table Cards Graph Heatmap

Type	
- N/A -	369155
Antibody	974
Cell	47
Enzyme	118
Gene	77
Oligonucleotide	170
Oligosaccharide	92
Protein	22682
Small molecule	1920366
Unclassified	4
Unknown	18015

Max Phase	
0	2322126
1	1479
2	2346

Go to related entities from a selection of items.

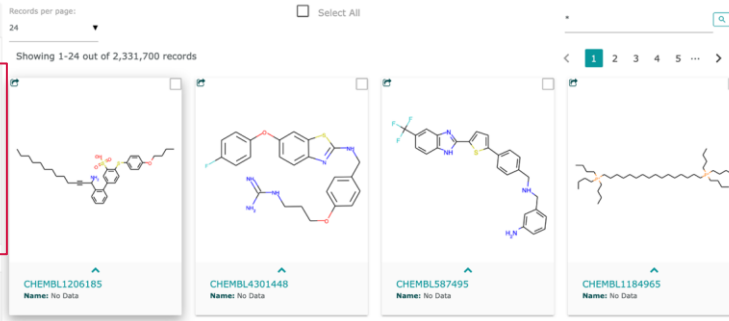
2,331,700 Compounds
3 Selected - Clear Selections
Browse Activities ?
Browse Drugs
Browse Drug Mechanisms
Browse Drug Indications

Download data

CSV TSV SDF

2,331,700 Compounds
0 Selected - Select All
Browse Activities ?

CSV TSV SDF



Report Card Pages

Report Cards for Compounds, Targets, Assays, Documents, Cell Lines and Tissues.

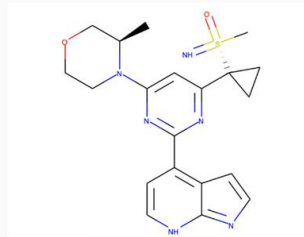
Search Results

All Results 272 Compounds 6 Targets 0 Assays 179 Documents 87 Cells 0 Tissues 0

Basic Information
of the item.

Compound Report Card

Name And Classification



Structure Search

ID: CHEMBL4285417
Name: CERALASERTIB
Max Phase: 2 Phase II
Molecular Formula: C20H24N6O2S
Molecular Weight: 412.52
ChEMBL Synonyms: ATR KINASE INHIBITOR AZD6738 AZD6738 AZD-6738 CERALASERTIB
Molecule Type: Small molecule

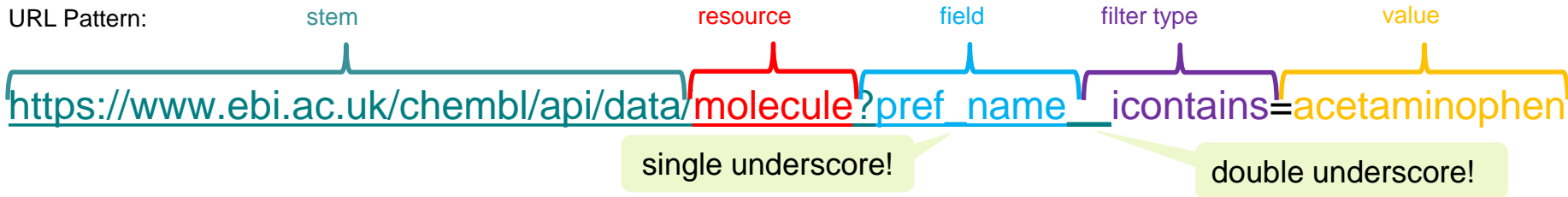
- Name And Classification
- Representations
- Sources
- Alternative Forms
- Molecule Features
- Drug Indications
- Drug Mechanisms
- Clinical Data
- Activity Charts
- Literature
- Target Predictions
- Calculated Properties
- Cross References
- UniChem Cross References
- UniChem Connectivity Layer Cross
- References

Summary of
the sections of
the page.



Formatting API calls

URL Pattern:



Filters:

- contains (iconcontains)
- startswith (istartswith)
- regex (iregex)
- isnull ...

(case insensitive)

Formatting:

Ordering -

https://www.ebi.ac.uk/chembl/api/data/target?order_by=-pref_name

insert '-' here
for descending

Formats -

- XML(default), JSON, YAML, (SVG, SDF)

<https://www.ebi.ac.uk/chembl/api/data/target.json>

<https://www.ebi.ac.uk/chembl/api/data/target?format=json>

Full list at: <https://chembl.gitbook.io/chembl-interface-documentation/web-services/chembl-data-web-services>

Accessing data: website versus programmatically

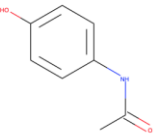
ChEMBL web interface

acetaminophen

Examples: Imatinib erBB2 brain MDCK c1ccccc1N

Compound Report Card

Name And Classification



Structure Search

ID: CHEMBL112
Name: ACETAMINOPHEN
Max Phase: 4 Approved
Molecular Formula: C8H9NO2
Molecular Weight: 151.16
ChEMBL Synonyms: Acetaminophen, ACETAMINOPHEN, APAP, NSC-109028, NSC-3991, PARACETAMOL, ARDINE, Acephen, ALVEDON, ANGIERS, CALPOL, CALPOL SIX PLUS, CALPOL SIX PLUS FASTMELTS, CHILD LEMSI, CUPANOL, DAFALGAN, Distri, DISPROL, DISPROL INFANT, DISPROL JNR, DISPROL PAED, FENNINGS, GALFAMOL, HEDEX, INFADROPS, Infanta's Feveral, Injectapap, JUNIOR PARAPAE, MANDANOL, MEDINOL, MEDINOL FOR CHILDREN, MEDINOL OVER 6, MEDINOL UNDER 6, MEDISSED PLAIN, MIRADOL, Neopap, OPIRHEV, PALDESIC, PANADOL, PANADOL ACTIFAST, PANADOL ACTIFAST SOLB, PANADOL ADVAN, PANADOL JNR, PANADOL OA, PANALEVE, PANALEVE 6 PLUS, Paracetamol, PARAMIN, PARAPAE JNR, PARAPAE SIX PLUS, PARAVICT, PERFALGAN, Phenaphen, PLACIDEX, RIMADOL, SALZONE, TIXYNOL, TRAMIL SOD, Tylenol
Trade Names:
Molecule Type: Small molecule

- Search for acetaminophen
- View the interactive compound report card

ChEMBL web services

<https://www.ebi.ac.uk/chembl/api/data/molecule/CHEMBL112>

Stem

Resource

Entity

```
<molecule>
  <atc_classifications>
    <level5>N02BE51</level5>
    <level5>N02BE01</level5>
    <level5>N02BE71</level5>
  </atc_classifications>
  <availability_type>2</availability_type>
  <biotherapeutic>
  <black_box_warning>1</black_box_warning>
  <chebi_par_id>46195</chebi_par_id>
  <chirality>2</chirality>
</cross_references>
  <molecule>
    <xref_id>acetaminophen</xref_id>
    <xref_name>acetaminophen</xref_name>
    <xref_src>DailyMed</xref_src>
```

- Direct URL to access compound data
- Computer-readable output

Best practice ideas

First steps:	Use the web interface for initial exploration, then progress onto the API
Get started with API:	Look at some example URLs: e.g. return all compounds: https://www.ebi.ac.uk/chembl/api/data/molecule
Check data output	Remember to spot check the API output e.g. For one compound: have I captured all relevant data? Are activity data for both parent and salt drug forms present?
Schema:	A schema description is available at e.g. https://www.ebi.ac.uk/chembl/api/data/activity/schema (although it can be easier to examine an actual example instead)
Try it out interactively:	Use our interactive API documentation https://www.ebi.ac.uk/chembl/api/data/docs



www.ebi.ac.uk/chembl



www.ebi.ac.uk/unichem/
www.ebi.ac.uk/unichem/beta/

- **Webinars:**
 - ChEMBL Webinar: <https://www.ebi.ac.uk/training/events/guide-explore-drug-compounds-and-their-biological-targets-using-chembl/>
 - ChEMBL quick tour: <https://www.ebi.ac.uk/training-beta/online/courses/chembl-quick-tour/>
- **Documentation:**
 - ChEMBL web services: <https://chembl.gitbook.io/chembl-interface-documentation/web-services>
 - UniChem documentation: <https://chembl.gitbook.io/unichem/> , Beta: <https://chembl.gitbook.io/unichem/unichem-2.0/>
- **FTP sites:**
 - UniChem: <ftp://ftp.ebi.ac.uk/pub/databases/chembl/UniChem/data/>
 - ChEMBL: <https://ftp.ebi.ac.uk/pub/databases/chembl/ChEMBLdb/releases/>
- **Contact us:** chembl-help@ebi.ac.uk, unichem@ebi.ac.uk
- **Blog:** <https://chembl.blogspot.com/>
- **FAQs:** <https://chembl.gitbook.io/chembl-interface-documentation/frequently-asked-questions>
- **For news and data releases subscribe to:** <http://listserver.ebi.ac.uk/mailman/listinfo/chembl-announce>

Acknowledgements

Chemical Biology Team

- Andrew R. Leach (Team Leader)
- **Barbara Zdrazil (Group Coordinator)***
- Eloy Felix (Technical Lead)
- **Fiona Hunter (Drug Data Leader)***
- Ricardo Arcila
- Melissa Adasme
- James Blackshaw
- Nicolas Bosc
- Sybilla Corbett
- Marleen De Veij
- Harris Ioannidis
- Tevfik Kiziloren
- Emma Manners
- M. Paula Magarinos
- **David Mendez***
- Juan F. Mosquera



* Contributed slides to these training materials



Open Targets

