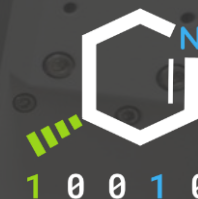
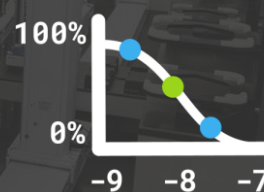


# DATA PLATFORMS FOR CHEMICAL BIOLOGY

● ● ● DEVELOPED AT CZ-OPENSREEN

Ctibor Škuta



1st Czech Cheminformatics Meeting 21/02/2023



 CZ-OPENSREEN

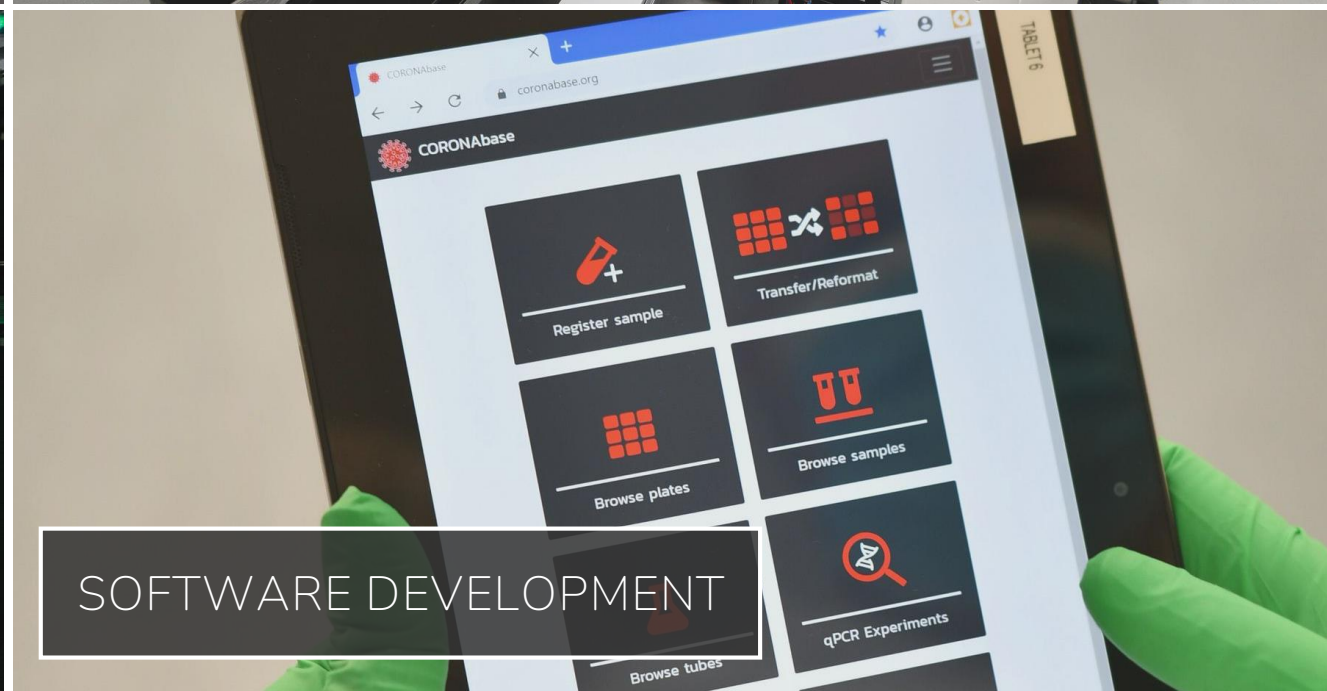
NATIONAL INFRASTRUCTURE FOR CHEMICAL BIOLOGY  
*openscreen.cz*



HTS/HCS SCREENING



COMPOUND MANAGEMENT



SOFTWARE DEVELOPMENT

# SOFTWARE DEVELOPMENT AT CZ-OPENSOURCE



A hub for the integration of high-quality bioactive compound sets enabling their analysis and comparison

[probes-drugs.org](http://probes-drugs.org)



An intuitive, cross-platform animal tracking database with advanced visualization, reporting and management features

[zebrabase.org](http://zebrabase.org)



European Chemical Biology Database (ECBD) is a central data hub for data generated within the EU-OPENSOURCE network

[ecbd.eu](http://ecbd.eu)



A comprehensive, cross-platform database for PCR diagnostics, which has been developed and employed at the IMG during the pandemic

[coronabase.org](http://coronabase.org)



A laboratory information management system (LIMS) - from sample acquisition to results reporting

not public yet



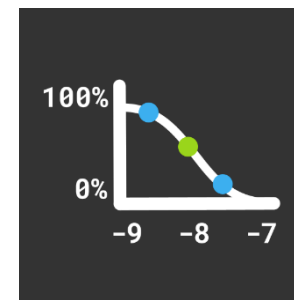
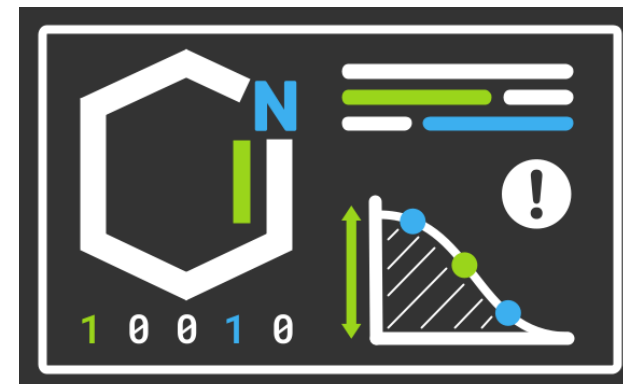
A cross-platform inventory application for biological samples management

not public yet

Other independent libraries: InChIlib, ChemSpace.js, RackScanner, printwes...

# DATA PLATFORMS IN CHEMICAL BIOLOGY

- Compound-centric systems with „chemical intelligence“
- Based on their purpose/users contain various sets of functions/data
- Public data integration platform (P&D) X Data hub for a large research consortium (ECBD) X Laboratory information management system (ScreenX)
- Common functionality – browse/detail views, filtering system, structure-based queries, data export



## OUR TECHNOLOGY STACK

Python/Django, JavaScript/Vue/Quasar, RDkit, PostgreSQL/RDkit cartridge, HTML, CSS, etc.





## ●●● A SHORT HISTORY OF THE PROBES & DRUGS PORTAL

- Data platform for bioactive compounds selection for a wide scientific community
- Started as a CZ-OPENSCREEN in-house project that would simplify our bioactive library updates selection
- High-quality bioactive compound sets scattered across databases, web-pages, publications, vendor sets
- Our goal was to put all of the relevant sets together and provide a user-friendly interface with tools that would allow anyone to work with the set
- P&D released in 2017 with ~30,000 compounds representing 29 sets

Skuta C, Popr M, Muller T, Jindrich J, Kahle M, Sedlak D, Svozil D, Bartunek P. (2017) Probes & Drugs Portal: an interactive, open data resource for chemical biology. *Nature Methods* 14(8), pp.759-760.



# PROBES & DRUGS PORTAL IN 2023

- P&D integrates 79 compound sources together with established public data resources (ChEMBL, BindingDB, GtoPDB, DrugBank, Drugcentral, Reactome and others)
- P&D probe-likeness score integrated together with scores from other probe resources Chemical Probes Portal and Probe Miner
- Live set of High-quality Chemical Probes (updated with each new P&D version)

Currently, the most comprehensive resource in the field of chemical probes

Škuta, C.; Southan, C.; Bartůněk, P. "Will the Chemical Probes Please Stand Up?" *RSC Med. Chem.* 2021, 12(8), 1428–1441.

**Probes & Drugs** 04.2022  
105,223 COMPOUNDS  
5,114 APPROVED DRUGS  
4,636 PROBES

PROBE TYPES: 1,134 experimental, 3,670 calculated, 1,154 P&D approved, 714 high-quality, 249 obsolete, 166 free of charge

Search:

Examples of complex queries:

1. What chemical probes are also approved drugs?
2. Which approved drugs bind to Glucocorticoid and Mineralocorticoid receptor with at least 1 nM potency?
3. What potent GPCR ligands are also potent Kinase ligands?
4. What chemical probes are highly-rated at the Chemical Probes Portal as well as Probes & Drugs?
5. What is the intersection of compounds in DrugBank, DrugCentral and ChEMBL Approved Drugs set?
6. What chemical probes are associated with at least one structural alert (PAINS, nuisance compounds, obsolete structure)?
7. What potent ligands of Estrogen receptor are dissimilar to Estradiol?
8. What chemical probes are, based on available data, meeting the biochemical potency (100nM) and on target cell-based potency (1uM) criteria?

Latest news: Jan. 6, 2023. P&D 04.2022 has surpassed 100k standardized compounds. At the beginning of the year 2023, we released a new version of the Probes... (read more)

Oct. 13, 2022. P&D 03.2022 released with extended downloads, intersection matrix and free chemical probes. This year's third version of the Probes & Drugs portal (ver. 03.2022) was released with... (read more)

Follow P&D on Twitter

Tweets from @probesanddrugs

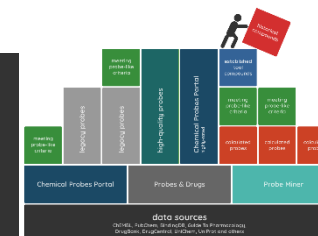
Probes & Drugs Retweeted Target 2035 @target2035 - Jan 12

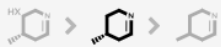
Congratulations to Aled (Order of Canada 2022), Cheryl (who is one of the top 1% of highly cited scientists, AAAS Fellow, Fellow of the Royal Society of Canada, past and present staff) #chemicalprobes

Probes & Drugs @probesanddrugs - Jan 6

Replying to @probesanddrugs @Boehringer and @thesgfrankfurt

On the occasion of adding a new drug resource, a table demonstrating the current situation in the field of drug structures (not only approved, AD) -





COMPOUND

NAMES & IDs

TAGS & SETS

PROBE SCORES

TARGETS & PATHWAYS

TABS

VIS

SEARCH

FILTERS

Search bar with icons

Search...

85082 AND NOT OR XOR

Compound Set 669

High-quality chemical probes HQCP

669 AND NOT OR XOR

Score 289

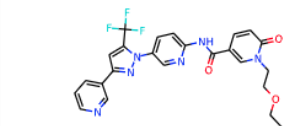
Cells score (Chemical Probes.org)

75

100

289 AND NOT OR XOR

Similarity in % 219



30

100

286

Compound details for GSK2334470

Names 4: GSK 2334470, GSK2334470, 1227911-45-6, External IDs 16, PubChem 46215815, ChEMBL CHEMBL1765740, GtoPdb 8008, ChemSpider 26346147, P&D IDs PD010630

Tags: Availability available, Compound Type probe, Probe type calculated probe experimental probe P&D approved, Probe selectivity protein-selective, Compound Sets 21: AdooQ Bioactive Compound Library, Axon Medchem Screening Library, Cayman Chemical Bioactives

Compound details for VENETOCLAX

Names 15: ABT-199, ABT 199, 1257044-40-8, External IDs 33, PubChem 49846579, ChEMBL CHEMBL3137309, GtoPdb 8318, ChemSpider 29315017, P&D IDs PD003393

Tags: Availability available, Compound Type drug probe, Drug Status approved investigational, Approved by FDA, PMDA, EMA, First approval 2016, Probe type

Compound details for S63845

Names 1: S63845, 1799633-27-4, External IDs 11

Tags: Availability available, Compound Type probe, Probe type experimental probe

Probe scores for GSK2334470

CELL-LINES & ORGANISMS

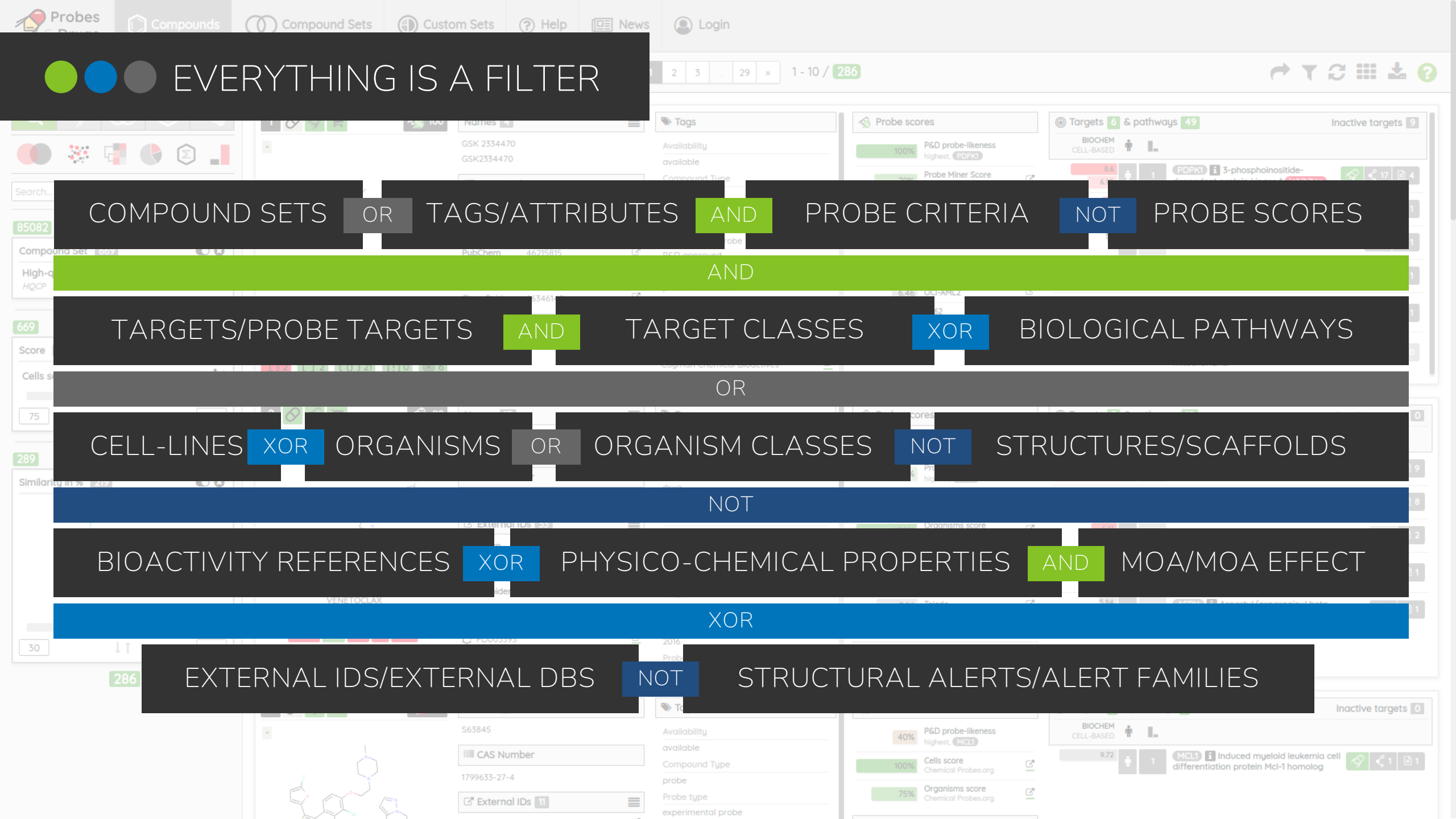
Probe scores for VENETOCLAX

Probe scores for S63845

Targets & pathways for GSK2334470

Targets & pathways for VENETOCLAX

Targets & pathways for S63845



# EVERYTHING IS A FILTER

COMPOUND SETS OR TAGS/ATTRIBUTES AND PROBE CRITERIA NOT PROBE SCORES

AND

TARGETS/PROBE TARGETS AND TARGET CLASSES XOR BIOLOGICAL PATHWAYS

OR

CELL-LINES XOR ORGANISMS OR ORGANISM CLASSES NOT STRUCTURES/SCAFFOLDS

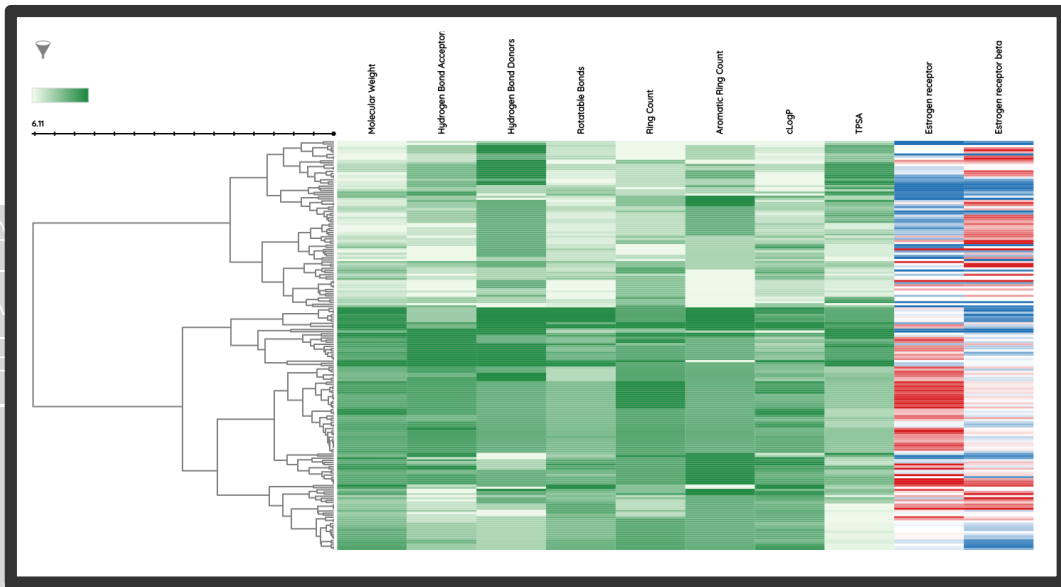
NOT

BIOACTIVITY REFERENCES XOR PHYSICO-CHEMICAL PROPERTIES AND MOA/MOA EFFECT

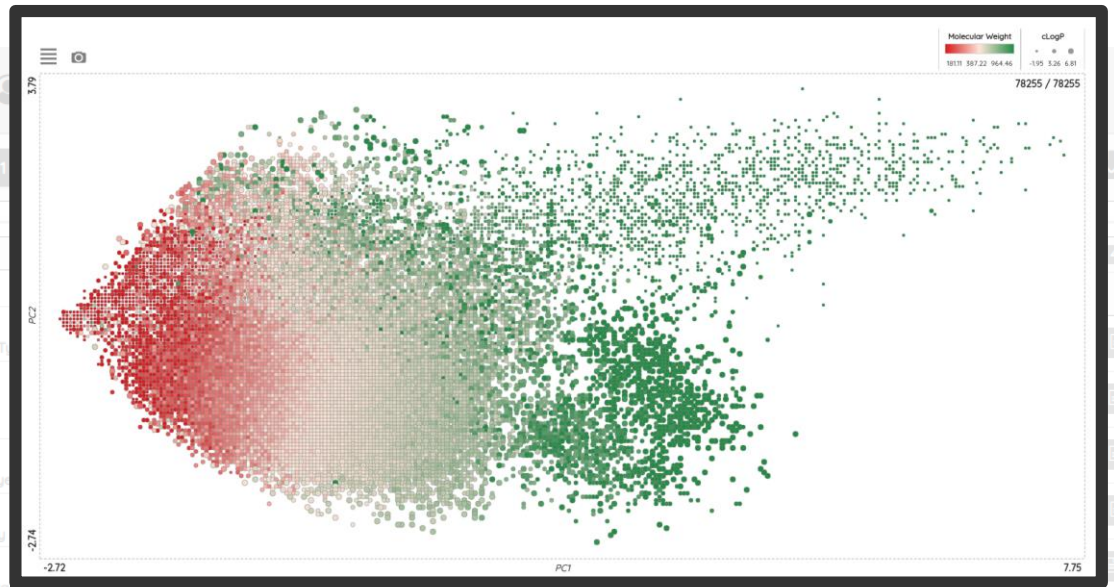
XOR

EXTERNAL IDS/EXTERNAL DBS NOT STRUCTURAL ALERTS/ALERT FAMILIES





CLUSTER HEATMAPS



CHEMICAL SPACE

Number of targets: [slider]

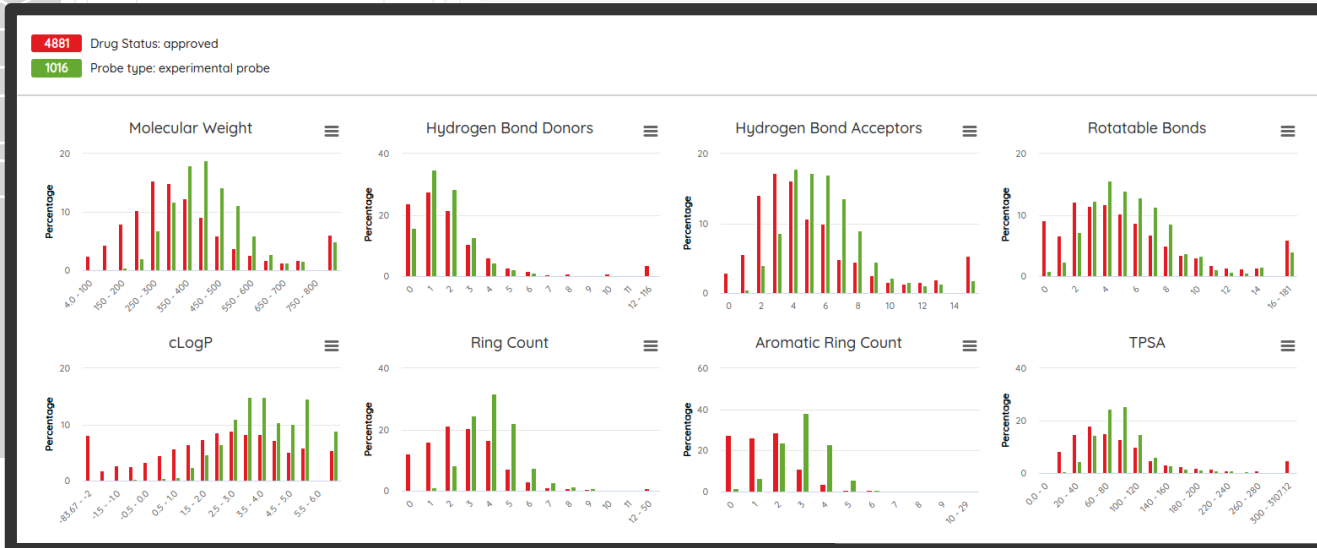
Tags: Availability: available, Compound Type: probe, drug, Drug Status: approved, Withdrawn year: 2001, Approved by: FDA, First approved: 2001

Names & IDs: [input fields]

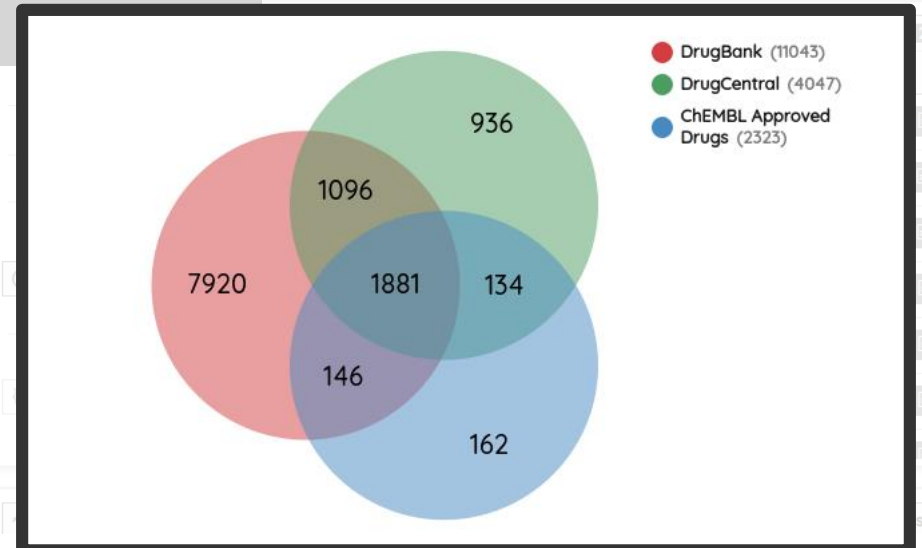
Tags & Sets: [input fields]

Organism: [input fields]

Search results: CA1 Carbonic anhydrase 1, PDGFRB Platelet-derived growth factor



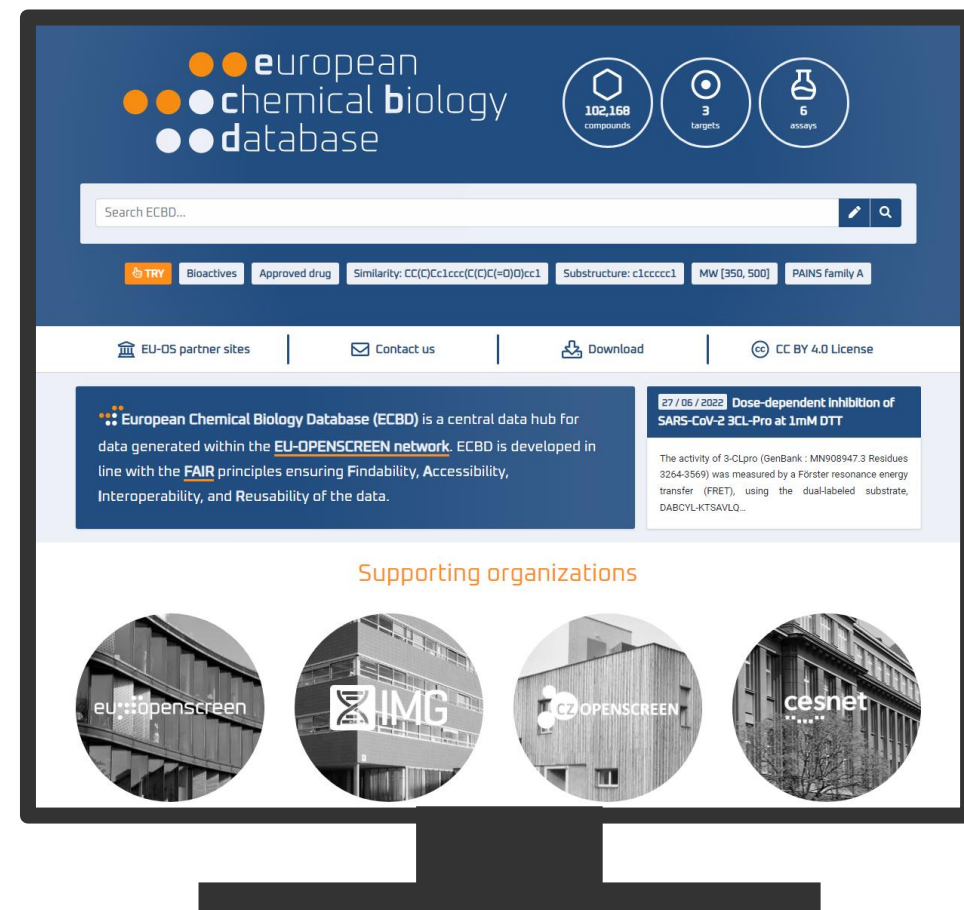
PHYSICAL-CHEMICAL PROPERTIES DISTRIBUTION



VENN DIAGRAMS

# EUROPEAN CHEMICAL BIOLOGY DATABASE (ECBD)

- Central repository for data generated within the EU-OPENSOURCE network
- Web interface for data upload, browse, analysis and export
- Public and private data (under embargo up to 3 years)
- Data in ECBD must be **FAIR** (Findable, Accessible, Interoperable and Reusable)
- Ontologies, established IDs and formats
- Web UI, exports, database dumps (public, private) and API
- Creative Commons license [Attribution 4.0 International \(CC BY 4.0\)](#)





# ECBD DATA UPLOAD

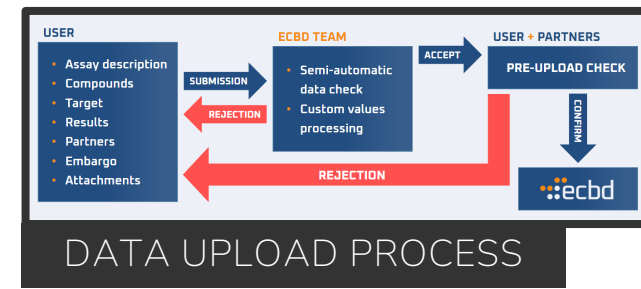
The screenshot shows the 'Upload data' form in the ECBD system. It is divided into several sections: 'Assay', 'Intended target', 'Target context', 'Compounds', 'Endpoint', 'Embargo time', 'Collaborators', and 'Attachments'. Each section contains various input fields, dropdown menus, and buttons for selection and submission.

UPLOAD FORM

The screenshot shows the 'Result' form in the ECBD system. It features several input fields for 'Endpoint count', 'Concentration unit', and 'Time unit'. A callout box labeled 'ONTOLOGY-BASED FIELDS' points to these fields, with a note '> 0.5, manual validation'. Below these are sections for 'Activity determination method', 'Endpoint', 'Endpoint file', 'Results file', 'Embargo time', and 'Collaborators'. The 'Endpoint' section includes a 'Unit' dropdown and a 'Preferred' checkbox. The 'Embargo time' section has a dropdown menu set to '-- CHOOSE --'. The 'Collaborators' section has 'Add' and 'Remove' buttons.

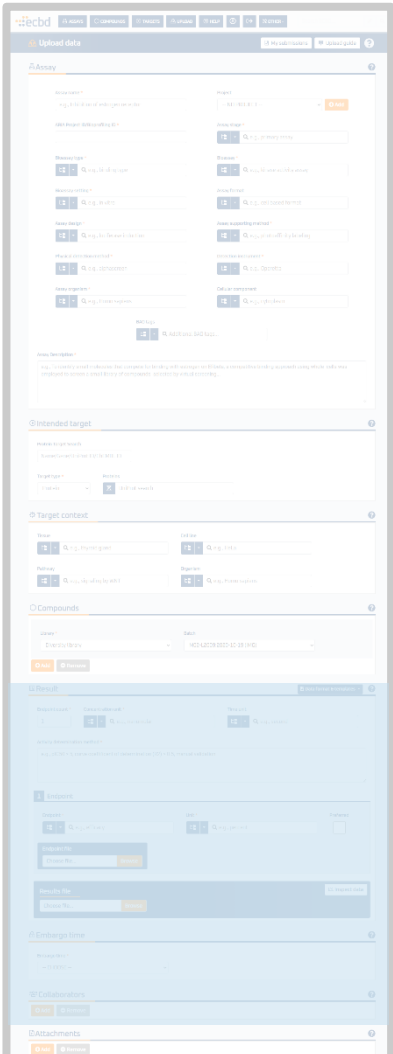
The screenshot shows the 'Select Assay design' dialog box. It includes a search bar for 'Search ontology' and a list of assay methods. The method 'chloramphenicol acetyltransferase induction' is highlighted. To the right, there is a detailed description of this method, including its ID (BAO\_000494) and an 'alternative term' section listing 'reporter gene method' as a subclass.

ONTOLOGY TERM SELECTION



DATA UPLOAD PROCESS





UPLOAD FORM

## Select Assay design

Search ontology

- binding assessment method
- cell cycle progression assessment method
- cell movement measurement method
- chromatin accessibility method
- conformation determination method
- cytokine quantitation method
- enzyme activity measurement method
- epigenetic modification detection method
- gene expression detection method
  - nucleic acid identification and quantitation method
  - reporter gene method
    - beta galactosidase induction
    - beta lactamase induction
    - chloramphenicol acetyltransferase induction
    - fluorescent protein induction
    - luciferase induction
- in vivo assay method
- kinase competitive binding method
- membrane potential measurement method
- molecular abundance method
- molecular redistribution determination method
- morphology assessment method
  - cellular morphology assessment method
  - in situ immunoassay
    - immunocytochemistry
    - immunohistochemistry
    - sub-cellular morphology assessment method
  - phosphoprotein detection method
    - phosphoprotein-specific antibody-coated bead based method
- size separation method
- viability measurement method

ONTOLOGY TERM SELECTION

# chloramphenicol acetyltransferase induction

ID: BAD\_0000494

*Chloramphenicol acetyltransferase (CAT) gene is attached to the regulatory sequence of a gene of interest. When induced to express, CAT breaks down the antibiotic chloramphenicol resulting in a resistance against that antibiotic. The effect of a perturbagen on the expression of the gene of interest could easily be monitored by the gain of antibiotic resistance by the cells.*

## alternative term

CAT induction

## Subclass of:

- [reporter gene method](#)

assay design

ontology

# chloramphenicol acetyltransferase induction

ID: BAD\_0000494

*Chloramphenicol acetyltransferase (CAT) gene is attached to the regulatory sequence of a gene of interest. When induced to express, CAT breaks down the antibiotic chloramphenicol resulting in a resistance against that antibiotic. The effect of a perturbagen on the expression of the gene of interest could easily be monitored by the gain of antibiotic resistance by the cells.*

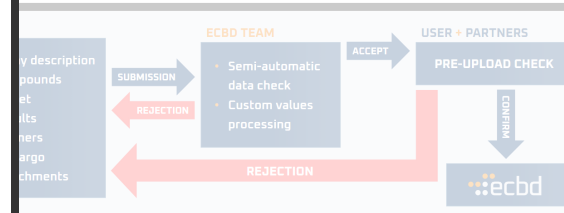
## alternative term

CAT induction

## Subclass of:

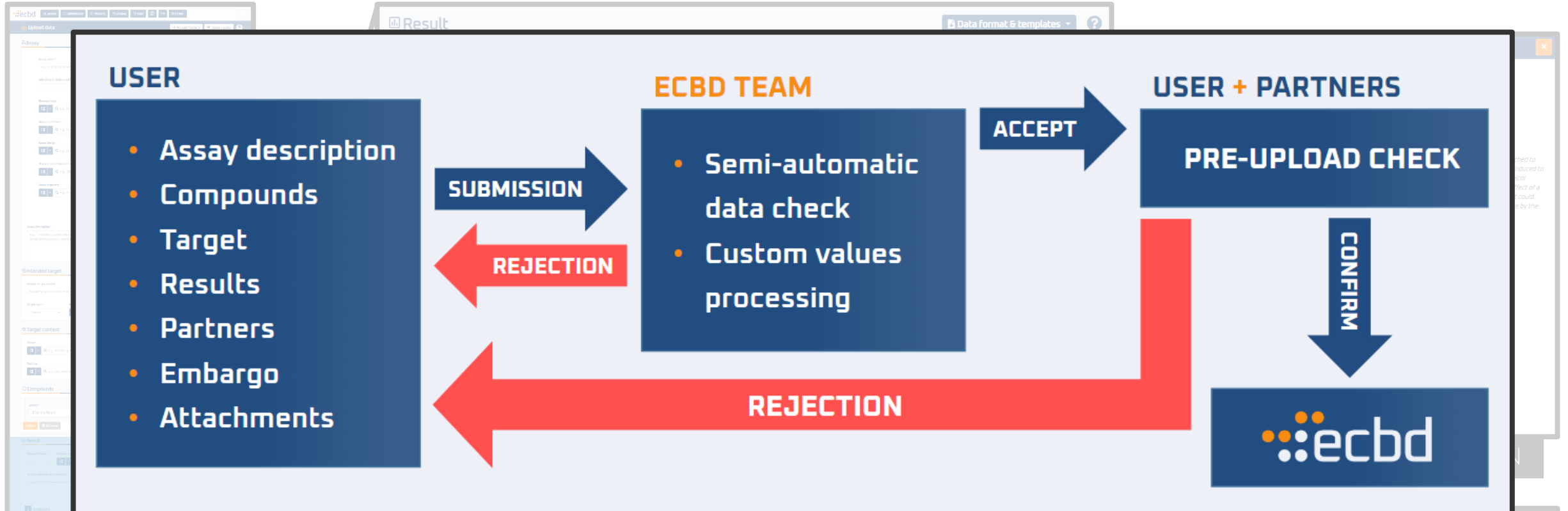
- [reporter gene method](#)

## ONTOLOGY TERM SELECTION



## DATA UPLOAD PROCESS

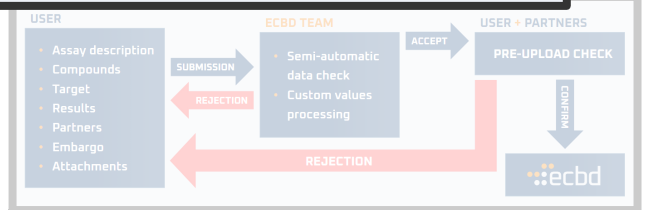
# ECBD DATA UPLOAD



DATA UPLOAD PROCESS

The screenshot shows the 'Upload form' interface. It includes a section for 'Embargo time' with a dropdown menu currently set to '-- CHOOSE --'. Below this is a 'Collaborators' section with 'Add' and 'Remove' buttons. The interface is part of a larger web application with various input fields and buttons.

UPLOAD FORM



DATA UPLOAD PROCESS

Assays 1 4

Search in assays

CURRENT FILTERS 2

- Assay stage 3
  - primary assay
- Bioassay 2
  - cell growth assay

ASSAY STAGE 3

- confirmatory assay 2
- primary assay 3
- alternate assay conditions 1

BIDASSAY TYPE 2

BIDASSAY 4

- cell growth assay 2
- protease activity assay 3
- signal transduction assay 1
- Fluorescence imaging cell growth inhibition assay 2

ASSAY ORGANISM 4

- Viruses 1

Assay **EOS300008**

Name [Single dose inhibition of SARS-CoV-2 cytopathic effect in VeroE6](#)

Stage primary assay

Type functional

Description To measure inhibition of the SARS-CoV-2 cytopathic effect, 384-well imaging plates (Greiner 781092) were spotted with test compounds and controls (16 ... [read more](#))

Compounds 2464

Activity 126 0 2337 2

Target

Name [Organism Severe acute respiratory syndrome-related coronavirus](#)

Type Organism

Activities 2686

Endpoint percent inhibition

ECBD guide ×

Welcome to the ECBD guide! This guided tour will introduce you to the data and main features of the ECBD browse views.

[NEXT](#)

Assay **EOS300033**

Name [Dose-dependent inhibition of SARS-CoV-2 cytopathic effect in VeroE6](#)

Stage confirmatory assay

Type functional

Description To measure inhibition of the SARS-CoV-2 cytopathic effect, 384-well imaging plates (Greiner 781092) were spotted with test compounds and controls (16 ... [read more](#))

Compounds 134

Activity 19 0 115 0

Target

Name [Organism Severe acute respiratory syndrome-related coronavirus](#)

Type Organism

Activities 50

Endpoint IC50 relative

Assay **EOS300038**

Name [GPR35 antagonistism](#)

Stage primary assay

Type functional

Description Although it has been described as a G protein-coupled receptor (GPCR) activated by either kynurenic acid, or by the chemokine CXCL17, GPR35 remains an... [read more](#)

Compounds 92313

Activity 0 0 92312 2

Target

Name [G-protein coupled receptor 35](#)

Type Protein

Activities 98600

Endpoint percent inhibition

Assay **EOS300040**

Description The activity of 3-CLpro (GenBank : MN908947.3 Residues 3264-3569) was measured by a Förster resonance energy transfer

Compounds 2465

Activity 64 0 2400 2

Target

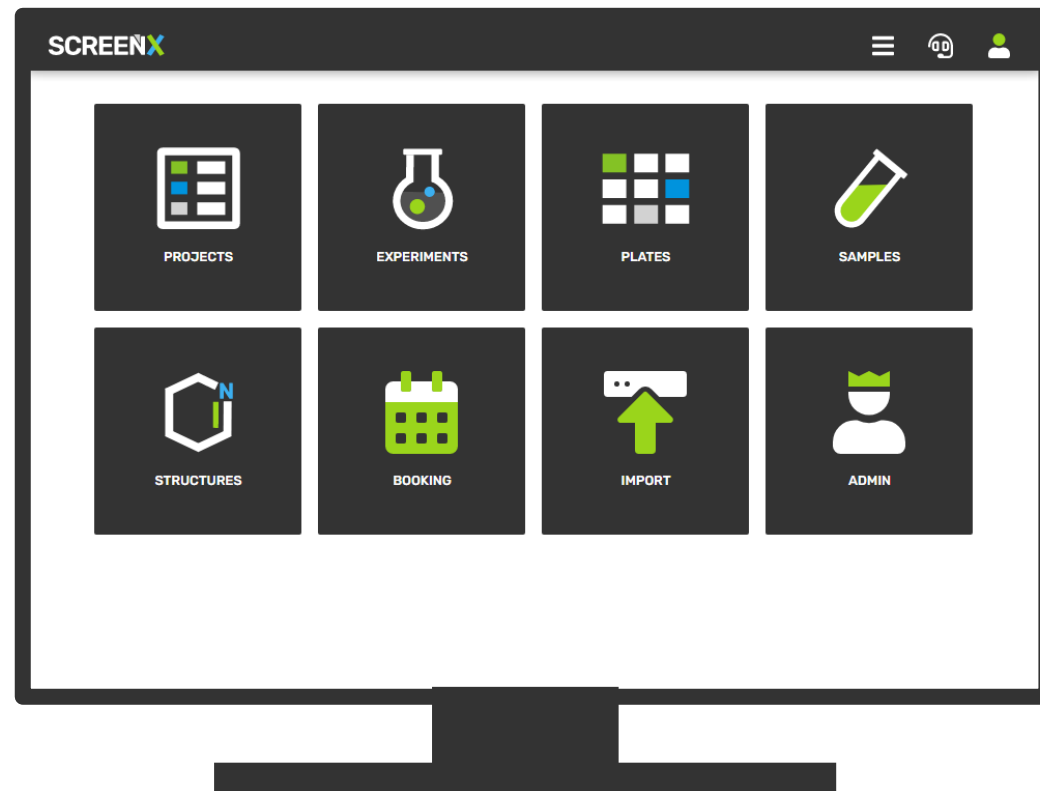
Activities 2688

Endpoint percent inhibition





- Laboratory information management system (LIMS)
- In an ideal case (in a distant future), it should be able to track everything that is happening in the lab
- Compound management, plate reformatting, instrument instructions, project/experiment definition, data normalizations/analysis, reporting, imports/exports, booking system, attachments, user roles





# EXPERIMENT SETUP

### Experiment setup

BACK SAVE

ASSAY TARGET DEVICES & ENDPOINTS PLATES PEOPLE

Name \* Target - HTS\_confirmation 384 cmpds Project \* NR Profiling

Short name of the experiment

Rationale  
Dose response of 384 cmpds (cca 10 nM to 0.1 mM). Triplicates. V tomto experimentu byly označeny pouze hity, které byly potvrzeny i na jiném substrátu, 50 µM XXX-XXX.

Why was the experiment performed

Experiment status \* Performed

#### EVENTS

Name	Type	Start	End
performed	Performed	2022-07-04 17:33	2022-07-04 17:33

+

#### ASSAY

Bioassay \* Ligand-binding Method \* Fluorescence intensity

Flags  Dose-response  Multi assay experiment ?

Assay stage \* confirmatory assay

Assay design

Assay format homogeneous phase

Assay organism Homo sapiens

Bioassay setting in vitro

MULTI-TAB DESCRIPTION FORM

### ECHO reformat

D-R COMBINATIONS

Preview

Start concentration 1e-5

Step 6

End concentration 1.29e-9

Dose-response point count 6

Multiplicates 3

Transfer from 2,5 nl Transfer to 200 nl

Total reaction volume 5 µl

Source parameter

Clip conc.

Manual place To all plates

CREATE

#### Transfer preview

	Dose-response concentrations					
	1	2	3	4	5	6
Theoretical conc LOG	-5	-5.78	-6.56	-7.33	-8.11	-8.89
0: Source conc	0.01	0.01	0.01	0.01	0.01	0.01
0: Theoretical volume [nl]	5.01	0.833	0.139	0.0231	3.86e-3	6.43e-4
0: Real volume [nl]	5	2.5	2.5	2.5	2.5	2.5
0: Real conc	1e-5	5e-6	5e-6	5e-6	5e-6	5e-6
0: Real conc LOG	-5	-5.3	-5.3	-5.3	-5.3	-5.3

1e-5

Source plates

CP-006132 PLATE TABLE

ADD

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z AA AB AC AD AE AF

SUBMIT

ASSAYPLATES REFORMATTING



# EXPERIMENT ANALYSIS

**SCREENX** PROJECTS EXPERIMENTS PLATES SAMPLES STRUCTURES ADMIN

ID 11: Target antagonist Endpoints Luminescence (Luminescence) Normalizations B\_score SET DEFAULT

ASSAY PLATES SCATTERPLOT SAMPLES

On page 20 Zoom -2.67 -0.0706 2.53 Heatmap scale [MAD]

1 2

AP-00113 AP-00114 AP-00115

AP-00116 AP-00117 AP-00118

ASSAYPLATE HEATMAPS

**SCREENX** PROJECTS EXPERIMENTS PLATES SAMPLES STRUCTURES ADMIN

FILTER RANGES HITS On page 20 Sort AUC / Category ratio

SCIPY THUNOR

Parameter boundaries

Apply smoothing

Min. y value for AUC calculation

**Fit curve only if:**

1. Curve slope is ASC DESC BOTH

2. Min / max difference is at least

3. Inflection is in conc. range

4. Min / max is close to controls

5. Min. response is at least

SID | 107559 T2078 CUDC-907 pXC50 6.64

SID | 107105 2288 Sigma Beta-penta-O-g... pXC50 6.60

SID | 115903 2228 Sigma Somatostatin ac... pXC50 5.73

SID | 113025 HY-10398 MedChemExpress CTS-1027 Structure ID: 100473 pXC50 6.39

SID | 113295 17246 Cayman MMP-3 inhibitor ... Structure ID: 100718 pXC50 6.18

SID | 107933 T3807 TargetMol (-)-gallocatechin... Structure ID: 99461 pXC50 5.99

SID | 105081 1789 UNCO224

SID | 106948 HY-10930 UNCO321

SID | 106897 HY-15663 IPA-3

DOSE-RESPONSE CURVE FITTING



● ● ● THE TEAM



● ● ● CONTACT

- [skutac@img.cas.cz](mailto:skutac@img.cas.cz)
- [cz-openscreen@img.cas.cz](mailto:cz-openscreen@img.cas.cz)
- [openscreen.cz/en/contact](https://openscreen.cz/en/contact)

[openscreen.cz](https://openscreen.cz)

