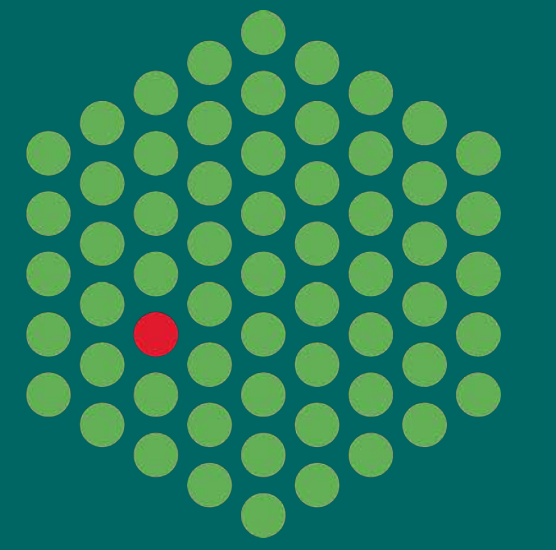


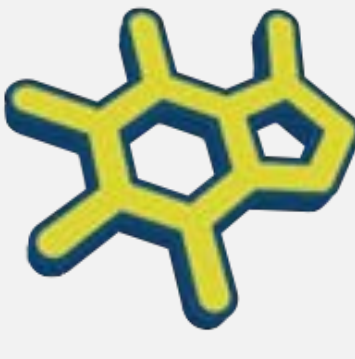
Chemical Biology Resources @EMBL-EBI

EMBL-EBI




Barbara Zdrzil,* Eloy Felix, Fiona Hunter, Melissa Adasme, Ricardo Arcila, Muhammad Arsalan, James Blackshaw, Nicolas Bosc, Sybilla Corbett, Harris Ioannidis, Maria Paula Magarinos, Adnan Malik, Emma Manners, David Mendez Lopez, Carlos Moreno, Juan F. Mosquera, Tevfik Kizilören, Ines Smith, Marleen de Veij, Andrew R. Leach

European Molecular Biology Laboratory-European Bioinformatics Institute (EMBL-EBI), Hinxton, Cambridge CB10 1SD, UK, *bzdrzil@ebi.ac.uk




ChEBI
~60K structures

2004




ChEMBL
>2m structures

2009



UniChem Links
~40 resources
>175m structures

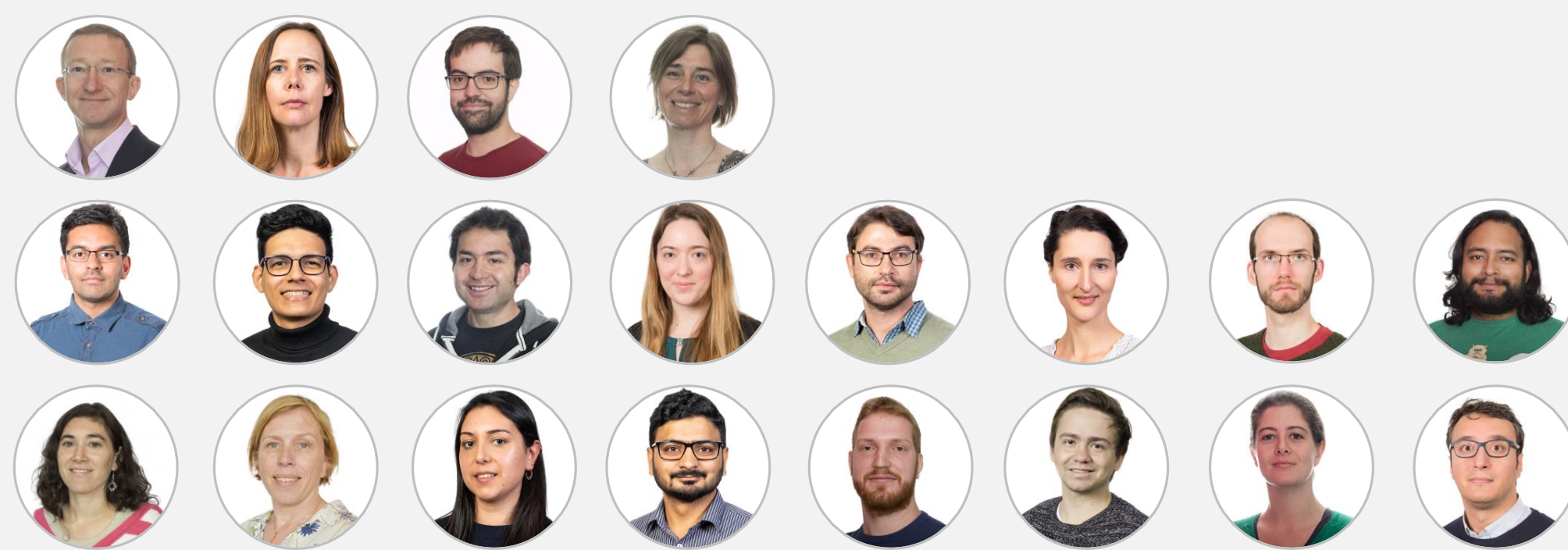
2013




SureChEMBL
>20m structures

2014

Chemical Biology Resources Team





SureChEMBL Facts

145M Patents

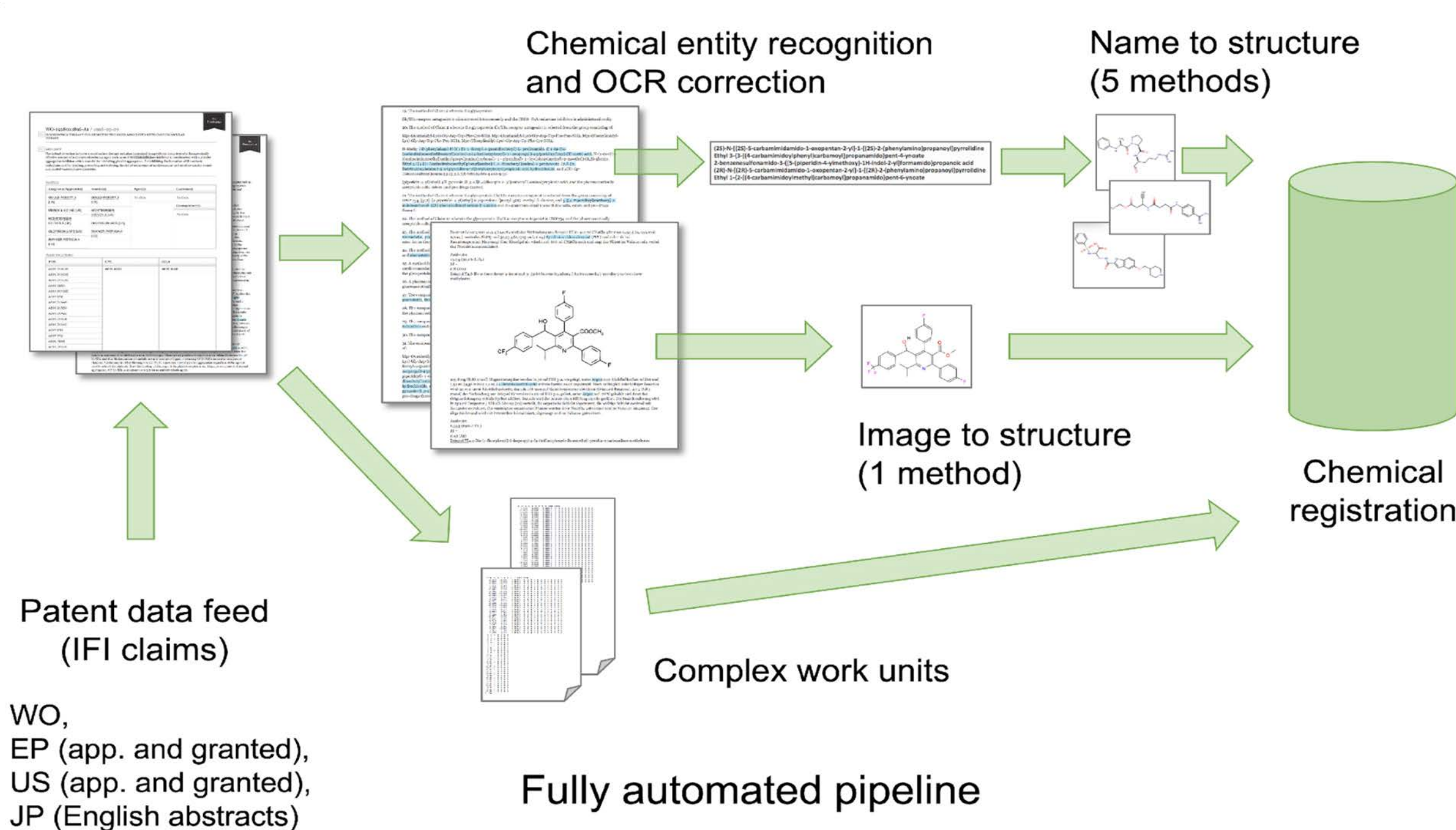
26M Documents

24.5M Chemicals

105K BioAnnotations

SureChEMBL Pipeline

- SureChEMBL is a publicly available large-scale resource that contains compounds extracted on a daily basis from the full text, images and attachments of patent documents, through an automated text and image-mining pipeline




New SureChEMBL UI

The interface shows a search page with filters for 'All chemically annotated authorities' and 'All authorities (inc. DocDB)'. It also displays 'Compound Details' for a specific entry (SCHEMBL1210), including its SMILES, InChI, and UniChem Cross References.

SureChEMBL New Architecture Features

- Scalable Kubernetes Microservice Architecture
- Pipeline Moved from AWS SQS to Kafka
- No Lucene Dependency
- Modern Mobile Friendly UI (Aligned with EBI Standards)
- Biological Annotations in Data Exports
- Public API



ChEMBL facts

15,398 Targets

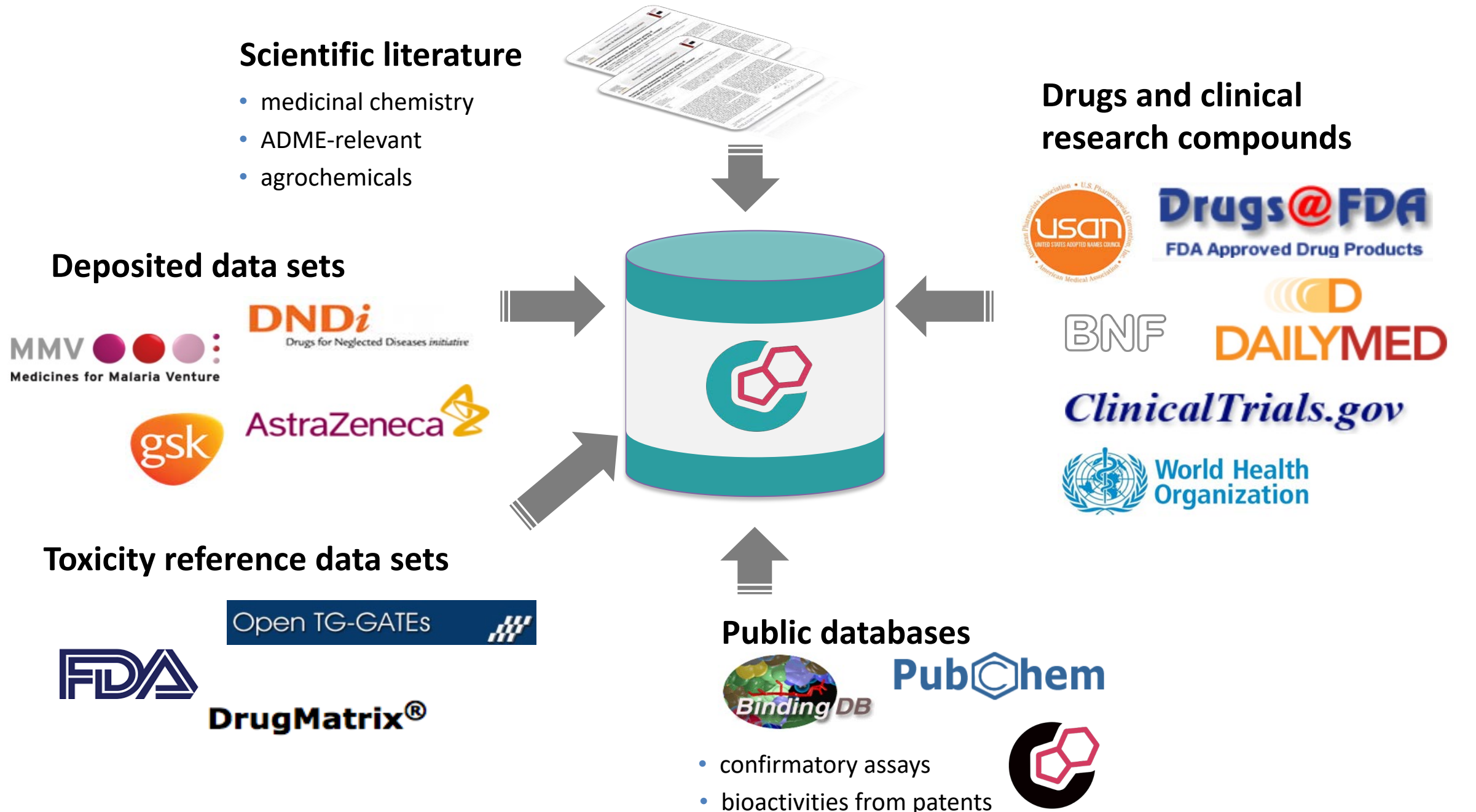
2,399,743 Distinct compounds

20,334,684 Activities

88,630 Publications

215 Deposited Datasets

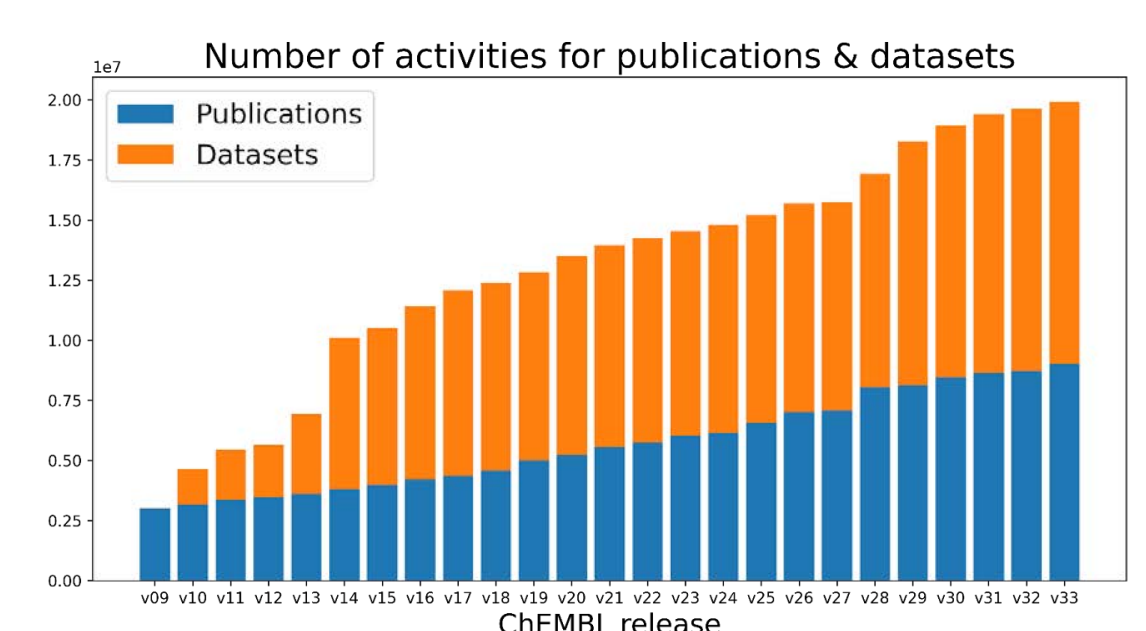
Current Release: ChEMBL 33
Last Update on 2023-05-31



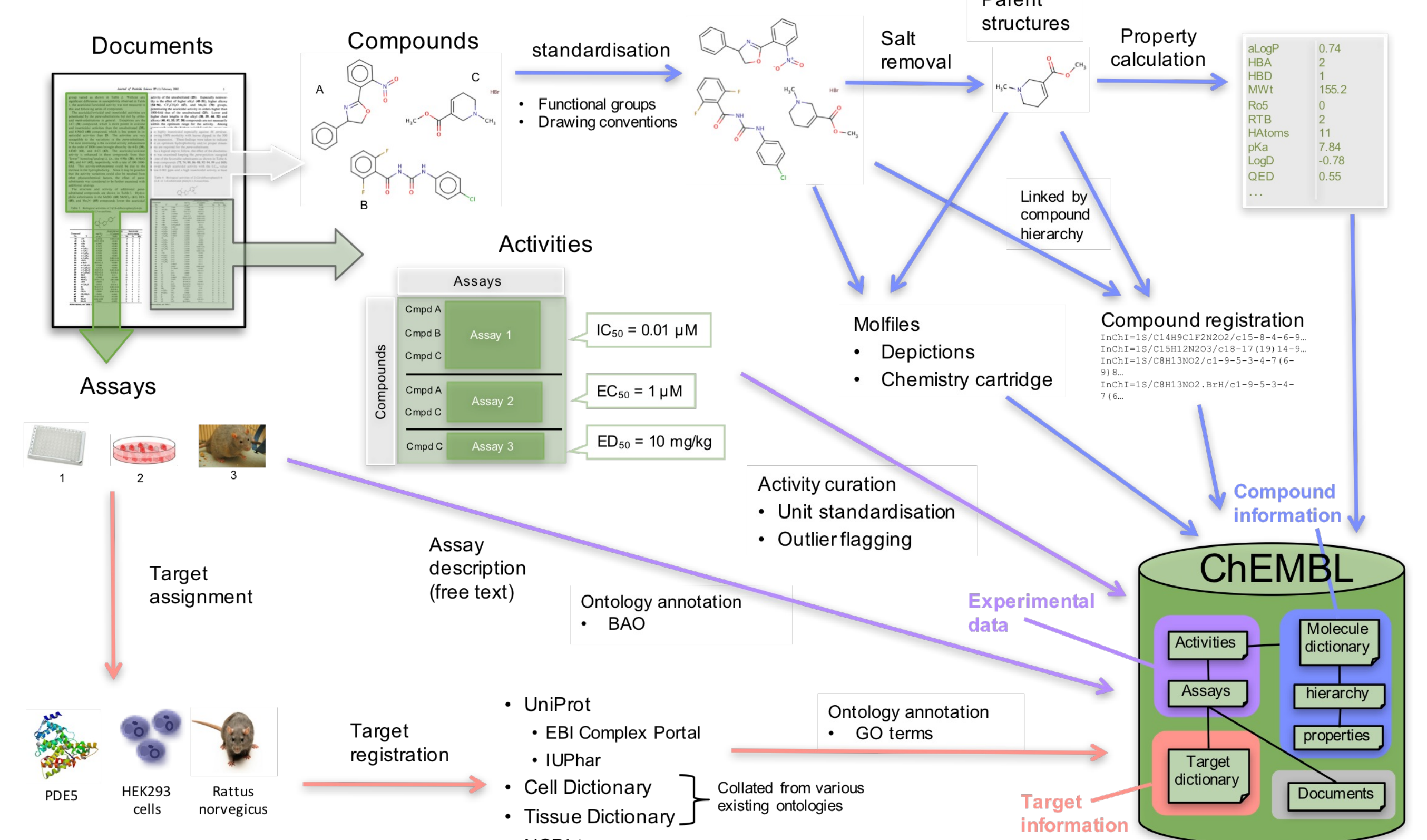
- ChEMBL integrates bioactivity data from diverse data sources
- Open, FAIR, extensively curated and high-quality drug-discovery data
- Bioactivities from Deposited Data Sets are now equal in number compared to those from published MedChem literature

New data & new features in ChEMBL 33:

- Bioactivity data on understudied SLC transporters from IMI-RESOLUTE
- New Chemical Probe data from IMI-EUROPEN
- Flag for Natural Products: annotations from COCONUT
- Flag for Chemical Probes: from chemicalprobes.org
- Action_Type (e.g., inhibitor, substrate): now available for ca. 270K bioactivities!!



Curation & Standardization in ChEMBL



Patent data in ChEMBL

Patent prioritisation
Patents from 2012–2018 (life science related): 3.7 million documents

Search patents for T_{dark}/T_{bio} targets in specific contexts (title, abstract, description, claims)

Presence of tables with bioactivity keywords (IC50, EC50,...)

Using this protocol, the following results were generated:

Example	JAKs: IC50 (nM)	JAKs: IC50 (nM)	JAKs: IC50 (nM)
1	463	18.1	5.84
2	453	43.8	11
3	317	4.09	1.2
4	846	31.7	44.6
5	339	21.6	7.07
6	604	54.3	7.05
7	172	18.5	8.46

4-pyrimidinesulfamide derivative

ABSTRACT
The invention relates to the compound of structural formula (I) and the salts thereof. Said compound is useful as **endothelin receptor antagonist**. The invention further relates to a process for preparing said compound.

The compound of formula I is an **endothelin receptor inhibitor** and useful as **endothelin receptor antagonist**.

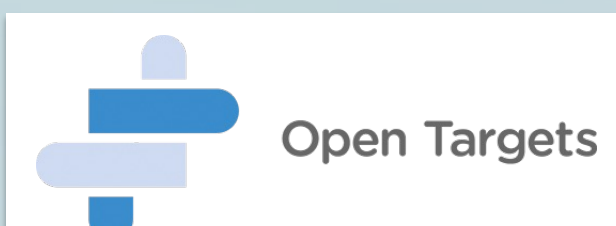


ChEMBL contains bioactivity data for selected patents from **SureChEMBL** (www.surechembl.org) & **BindingDB**

Recently, 3.7 Mio life science relevant patents were prioritised in order to extract bioactivity data for T_{dark} & T_{bio} targets from **SureChEMBL**

190 patents had data for 152 T_{dark} & T_{bio} targets

76 T_{dark} & T_{bio} targets are annotated to at least 1 active compound (according to IDG activity cut-off's)



European Bioinformatics Institute
Wellcome Genome Campus
Hinxton, Cambridge, CB10 1SD, UK

Tel. +44 (0) 1223 494 444
comms@ebi.ac.uk
www.ebi.ac.uk