ChEMBL: Linking Chemicals, Targets, and Bioactivity Data for Drug Discovery

Challenge

Drug discovery and development are increasingly data-driven but remain hindered by fragmented, inconsistent, and proprietary bioactivity information. Researchers and companies often spend significant resources collecting and standardizing chemical and biological data to identify promising drug candidates. This slows innovation, increases costs, and limits reproducibility in early-stage discovery. With the growing potential that machine learning (ML) offers in drug discovery, there is a rising demand for data that is easily interpretable by computers.

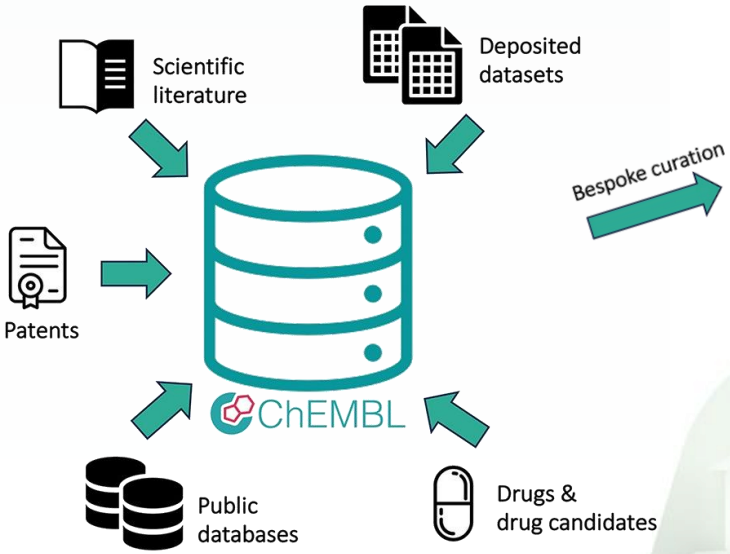
Technology

ChEMBL is an open, curated bioactivity database that integrates chemical structures, drug-like compounds, targets, and assay data either extracted from peer-reviewed journals and other sources or from deposited datasets. Maintained by the European Bioinformatics Institute (EMBL-EBI), ChEMBL provides a unified, high-quality resource linking compounds to their biological effects with a focus on creating ML-ready datasets to support data-driven drug discovery.

Core Expertise

Our core expertise is in chemical biology data services, including:

- ChEMBL database itself
- Patent data extraction
- Bioactivity curation
- Integration of experimental results
- Development of ML-ready datasets



Internal EMBLEM Reference

2026-027

Key Inventors

Noel O'Boyle, PhD
Chemical Biology Resources Team Leader

EMBLEM TECHNOLOGY TRANSFER GmbH

Boxbergring 107
D-69126 Heidelberg
Germany

Tel.: +49 (0) 6221 363 22 10

info@embl-em.de

www.embl-em.de

Dr. Laurène Martins
martins@embl-em.de

ChEMBL: Linking Chemicals, Targets, and Bioactivity Data for Drug Discovery

Applications

Pharmaceuticals

- Accelerating target identification, lead discovery, and optimization.
- Enabling AI-driven prediction of drug activity, toxicity, and pharmacokinetics.
- Supporting drug repurposing and precision medicine approaches.
- Reducing R&D costs through data-driven decision-making.

Agriculture

- Aid design of safer, more selective agrochemicals.
- Supporting discovery of novel modes of action to combat resistance.
- Enabling cross-domain insights between pharmaceutical and agrochemical targets.

Human Health

- Facilitating early toxicity and safety prediction.
- Supporting antimicrobial and antiviral discovery, including neglected diseases.
- Assisting environmental health assessments by linking chemicals to biological targets.

Collaboration Opportunities

We are open to collaborations and partnerships to accelerate drug discovery using curated data and ML-ready resources.

Machine Learning:
We are integrating ChEMBL data with your data and building ML models

Data Extraction:
We can assist with manual data extraction services from patents of interest.

In vivo Data:
We capture appropriate Metadata from assays and build the right Metadata assembly for your assay.

New modalities:
We are engaged in emerging therapeutic approaches, such as peptides, PROTACs, and other non-traditional drugs

Linking bioactivity data from ChEMBL to structural information in PDBe to support drug discovery and development of co folding methods.

Internal EMBLEM Reference

2026-027

Key Inventors

Noel O'Boyle, PhD
Chemical Biology Resources Team Leader, EMBL-EBI, Hinxton

Keywords

Chemical biology services
Bioactivity
Data extraction
Drug discovery
ChEMBL Database
Target Prediction

Further Reading

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

https://github.com/chembl/

https://www.ebi.ac.uk/about/teams/chemical-biology-services/

https://doi.org/10.1093/nar/gkad1004

https://doi.org/10.1093/nar/gkv352

EMBLEM TECHNOLOGY TRANSFER GmbH

Boxbergring 107
D-69126 Heidelberg
Germany

Tel.: +49 (0) 6221 363 22 10

info@embl-em.de

www.embl-em.de

Dr. Laurène Martins
martins@embl-em.de