ChEMBL

ChEMBL or **ChEMBLdb** is a manually curated chemical database of bioactive molecules with drug-like properties.[1] It is maintained by the European Bioinformatics Institute (EBI), of the European Molecular Biology Laboratory (EMBL), based at the Wellcome Trust Genome Campus, Hinxton, UK.

The database, originally known as StARlite, was developed by a biotechnology company called Inpharmatica Ltd. later acquired by Galapagos NV. The data was acquired for EMBL in 2008 with an award from The Wellcome Trust,[2] resulting in the creation of the ChEMBL chemogenomics group at EMBL-EBI, led by John Overington.[3][4]

Contents

Scope and access

The ChEMBL database contains compound bioactivity data against drug targets. Bioactivity is reported in Ki, Kd, IC50, and EC50.[5] Data can be filtered and analyzed to develop compound screening libraries for lead identification during drug discovery.[6]

ChEMBL version 2 (ChEMBL_02) was launched in January 2010, including 2.4 million bioassay measurements covering 622,824 compounds, including 24,000 natural products. This was obtained from curating over 34,000 publications across twelve medicinal chemistry journals. ChEMBL's coverage of available bioactivity data has grown to become "the most comprehensive ever seen in a public database.".[3] In October 2010 ChEMBL version 8 (ChEMBL_08) was launched, with over 2.97 million bioassay measurements covering 636,269 compounds.[7]

ChEMBL_10 saw the addition of the PubChem confirmatory assays, in order to integrate data that is comparable to the type and class of data contained within ChEMBL.[8]

ChEMBL	
Content	
Description	Biological database
Data types captured	Molecules with drug-like properties and biological activity
Contact	
Research center	European Molecular Biology Laboratory
Laboratory	🇬🇧 European Bioinformatics Institute
Authors	Andrew Leach, Team Leader 2016-Present; John Overington, Team Leader 2008-2015
Primary citation	PMID 21948594 (https://pubmed.ncbi.nlm.nih.gov/21948594)
Release date	2009
Access	
Website	ChEMBL (https://www.ebi.ac.uk/chembl/)
Download URL	Downloads (http

ChEMBLdb can be accessed via a web interface or downloaded by File Transfer Protocol. It is formatted in a manner amenable to computerized data mining, and attempts to standardize activities between different publications, to enable comparative analysis.[1] ChEMBL is also integrated into other large-scale chemistry resources, including PubChem and the ChemSpider system of the Royal Society of Chemistry.

Associated resources

In addition to the database, the ChEMBL group have developed tools and resources for data mining.[9] These include Kinase SARfari, an integrated chemogenomics workbench focussed on kinases. The system incorporates and links sequence, structure, compounds and screening data.

GPCR SARfari is a similar workbench focused on GPCRs, and ChEMBL-Neglected Tropical Diseases (ChEMBL-NTD) is a repository for Open Access primary screening and medicinal chemistry data directed at endemic tropical diseases of the developing regions of the Africa, Asia, and the Americas. The primary purpose of ChEMBL-NTD is to provide a freely accessible and permanent archive and distribution centre for deposited data.[3]

July 2012 saw the release of a new malaria data service (https://www.ebi.ac.uk/chembl/malaria/), sponsored by the Medicines for Malaria Venture (MMV (http://www.mmv.org/)), aimed at researchers around the globe. The data in this service includes compounds from the Malaria Box screening set, as well as the other donated malaria data found in ChEMBL-NTD.

myChEMBL (http://chembl.blogspot.co.uk/2013/10/chembl-virtual-machine-aka-mychembl.html), the ChEMBL virtual machine, was released in October 2013 to allow users to access a complete and free, easy-to-install cheminformatics infrastructure.

In December 2013, the operations of the SureChem patent informatics database were transferred to EMBL-EBI. In a portmanteau, SureChem was renamed SureChEMBL.

2014 saw the introduction of the new resource ADME SARfari (https://web.archive.org/web/20151106070004/https://www.ebi.ac.uk/chembl/admesarfari) - a tool for predicting and comparing cross-species ADME targets.[10]

	s://www.ebi.ac.uk/chembl/downloads/)
Web service URL	ChEMBL Webservices (https://www.ebi.ac.uk/chembl/ws)
Sparql endpoint	ChEMBL EBI-RDF Platform (http://www.ebi.ac.uk/rdf/services/chembl/sparql)
Miscellaneous	
License	The ChEMBL data is made available on a Creative Commons Attribution-Share Alike 3.0 Unported Licence
Versioning	ChEMBL_25

See also

- ChEMBL: Quick Tour on EBI Train OnLine (http://www.ebi.ac.uk/training/online/course/chembl-quick-tour)
- ChEBI
- DrugBank

References

1. Gaulton, A; et al. (2011). "ChEMBL: a large-scale bioactivity database for drug discovery" (http

s://www.ncbi.nlm.nih.gov/pmc/articles/PMC3245175). *Nucleic Acids Research*. **40** (Database issue): D1100-7. doi:10.1093/nar/gkr777 (https://doi.org/10.1093%2Fnar%2Fgkr777). PMC 3245175 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3245175). PMID 21948594 (https://pubmed.ncbi.nlm.nih.gov/21948594).

2. "Open access drug discovery database launches with half a million compounds | Wellcome" (https://wellcome.ac.uk/press-release/open-access-drug-discovery-database-launches-half-million-compounds). *wellcome.ac.uk*. 18 January 2010. Retrieved 31 August 2019.

3. Bender, A (2010). "Databases: Compound bioactivities go public". *Nature Chemical Biology*. **6** (5): 309. doi:10.1038/nchembio.354 (https://doi.org/10.1038%2Fnchembio.354).

4. Overington J (April 2009). "ChEMBL. An interview with John Overington, team leader, chemogenomics at the European Bioinformatics Institute Outstation of the European Molecular Biology Laboratory (EMBL-EBI). Interview by Wendy A. Warr". *J. Comput.-Aided Mol. Des*. **23** (4): 195–8. Bibcode:2009JCAMD..23..195W (https://ui.adsabs.harvard.edu/abs/2009JCAMD..23..195W). doi:10.1007/s10822-009-9260-9 (https://doi.org/10.1007%2Fs10822-009-9260-9). PMID 19194660 (https://pubmed.ncbi.nlm.nih.gov/19194660). S2CID 2946417 (https://api.semanticscholar.org/CorpusID:2946417).

5. Mok, N. Yi; Brenk, Ruth (Oct 24, 2011). "Mining the ChEMBL Database: An Efficient Chemoinformatics Workflow for Assembling an Ion Channel-Focused Screening Library" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3200031). *J. Chem. Inf. Model.* **51** (10): 2449–2454. doi:10.1021/ci200260t (https://doi.org/10.1021%2Fci200260t). PMC 3200031 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3200031). PMID 21978256 (https://pubmed.ncbi.nlm.nih.gov/21978256).

6. Brenk, R; Schinpani, A; James, D; Krasowski, A (Mar 2008). "Lessons learnt from assembling screening libraries for drug discovery for neglected diseases" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2628535). *ChemMedChem*. **3** (3): 435–44. doi:10.1002/cmdc.200700139 (https://doi.org/10.1002%2Fcmdc.200700139). PMC 2628535 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2628535). PMID 18064617 (https://pubmed.ncbi.nlm.nih.gov/18064617).

7. ChEMBL-og (15 November 2010), *ChEMBL_08 Released* (http://chembl.blogspot.com/2010/11/chembl08-released.html), retrieved 2010-11-15

8. ChEMBL-og (6 June 2011), *ChEMBL_10 Released* (http://chembl.blogspot.com/2011/06/chembl-10-released.html), retrieved 2011-06-09

9. Bellis, L J; et al. (2011). "Collation and data-mining of literature bioactivity data for drug discovery". *Biochemical Society Transactions*. **39** (5): 1365–1370. doi:10.1042/BST0391365 (https://doi.org/10.1042%2FBST0391365). PMID 21936816 (https://pubmed.ncbi.nlm.nih.gov/21936816).

10. Davies, M; et al. (2015). "ADME SARfari: Comparative Genomics of Drug Metabolising Systems" (http://bioinformatics.oxfordjournals.org/content/early/2015/01/08/bioinformatics.btv010.full.pdf+html). *Bioinformatics*. **31** (10): 1695–1697. doi:10.1093/bioinformatics/btv010 (https://doi.org/10.1093%2Fbioinformatics%2Fbtv010). PMC 4426839 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4426839). PMID 25964657 (https://pubmed.ncbi.nlm.nih.gov/25964657). Retrieved 2015-01-08.

External links

- ChEMBLdb (https://www.ebi.ac.uk/chembldb)
- Kinase SARfari (https://www.ebi.ac.uk/chembl/sarfari/kinasesarfari)
- ChEMBL-Neglected Tropical Disease Archive (https://web.archive.org/web/20170707155137/https://www.ebi.ac.uk/chemblntd)
- GPCR SARfari (https://web.archive.org/web/20180126094043/https://www.ebi.ac.uk/chembl/sarfari/gpcrsarfari)
- The ChEMBL-og (http://chembl.blogspot.com/search/label/Peptide%20Drugs) Open data and drug discovery blog run by the ChEMBL team.