ChEMBL and SureChEMBL - Open Resources for Chemoinformatics and Drug Discovery

John P. Overington

European Molecular Biology Laboratory - European Bioinformatics Institute (EMBL-EBI), Hinxton, Cambridge CB10 1SD, United Kingdom

Every year (academic) research produces a large amount of research describing the biological effects of chemical substances. This valuable information, while public, is usually in a form not accessible for systematic data extraction (data mining) and lacks consistent standardisation. ChEMBL is an Open Data database that contains this information manually extracted from the literature. The database contains binding, functional and ADMET information for a large number of drug-like bioactive compounds. Data is further curated and standardised (assay read out and chemical structures) to maximise their quality and utility across a wide range of chemical biology and drug-discovery research problems. The current release (18) contains : 53,298 documents, 12,419,715 bioactivity measurements, 1,359,519 compounds, and 9,414 targets. Access is freely available through a web-based interface, data downloads (including a local copy of the database) and web services at https://www.ebi.ac.uk/chembldb. Recently ChEMBL has been joined by the SureChEMBL resource of patent-dervied chemical structures, with integration between the ChEMBL and SureChEMBL sets being achieved via cross-references sharing identical standard InChIs. The session will focus on the application of ChEMBL to drug discovery problems, including target prediction, library design and other analyses.