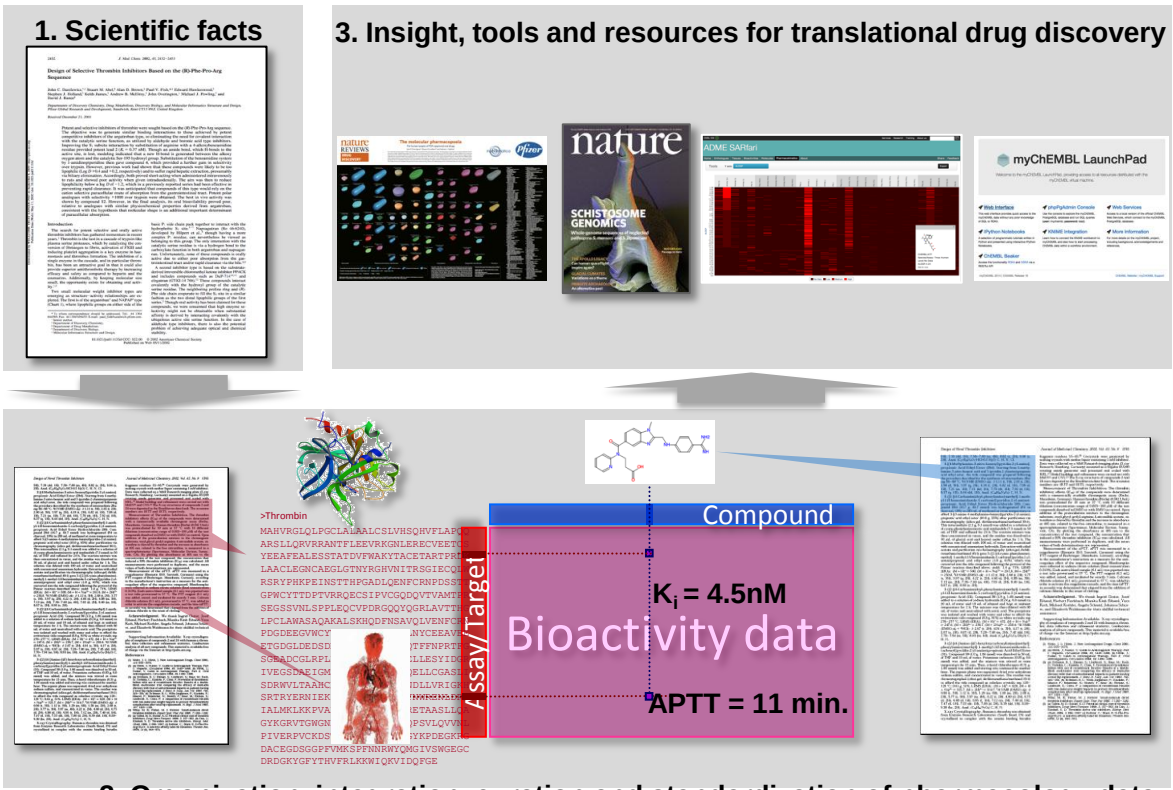
ChEMBL: The Organization of Drug Discovery Data

Anne Hersey & George Papadatos

Workshop: January 28th 2015 1-4pm @ MolSoft, San Diego.



Every year, research produces a vast amount of data 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 standardization.

ChEMBL is an Open Data database that contains this information manually extracted from the primary scientific literature. The database contains binding, functional and ADMET information for a large number of drug-like bioactive compounds. Data is further curated and standardized (assay read outs and chemical structures) to maximize their quality and utility across a wide range of chemical biology and drug-discovery research problems.

We have recently complemented ChEMBL with SureChEMBL, a web-based daily updated database of full patent text and automatically extracted compound structures. This allows advanced chemistry based querying of the patent literature. We will present an overview of both resources, and then show specific use common usage and data analysis scenarios.

ChEMBL Group
EMBL-EBI
Wellcome Trust Genome Campus, Hinxton, CB10 1SD, UK
chembl-help@ebi.ac.uk
www.ebi.ac.uk/chembl