

John Overington gave this presentation at the International Symposium on Compound Design Technologies held in Tokyo and Osaka, Japan on 19 and 20 March 2014.

Abstract

The link between biological and chemical worlds is of critical importance in many fields, not least that of healthcare and chemical safety assessment. A major focus in the integrative understanding of biology are genes/proteins and the networks and pathways describing their interactions and functions; similarly, within chemistry there is much interest in efficiently identifying drug-like, cell-penetrant compounds that specifically interact with and modulate these targets. The number of genes of interest is of the range of 10⁵ to 10⁶, which is modest with respect to plausible drug-like chemical space – 10²⁰ to 10⁶⁰. We have built a public database linking chemical structures (~10⁶) to molecular targets (~10⁴), covering molecular interactions and pharmacological activities and Absorption, Distribution, Metabolism and Excretion (ADME) properties – ChEMBL (<http://www.ebi.ac.uk/chembl>) in an attempt to map the general features of molecular properties and features important for both small molecule and protein targets in drug discovery. We have then used this empirical kernel of data to extend analysis across the human genome, and to large virtual databases of compound structures. Recently we have added large scale text mined chemical structures from patents to our resources (<http://www.surechembl.org>).

You can download this presentation as a [PDF](#) .