[L16] Challenges and Opportunities for Big Data Analysis in Chemistry

Igor V. Tetko,[1,2] Engkvist Ola,[3] Uwe Koch,[4] Jean-Louis Reymond,[5] Hongming Chen[3]

[1]Helmholtz Zentrum München - German Research Center for Environmental Health (GmbH), Institute of Structural Biology, Ingolstädter Landstraße 1, b. 60w, D-85764 Neuherberg, Germany;
[2]BIGCHEM GmbH, Ingolstädter Landstraße 1, b. 60w, D-85764 Neuherberg, Germany;
[3]Chemistry Innovation Centre, Discovery Sciences, AstraZeneca R&D Gothenburg, Pepparedsleden 1, Mölndal, SE-43183, Sweden;
[4]Lead Discovery Center GmbH, Otto-Hahn Strasse 15, Dortmund 44227, Germany;
[5]Department of Chemistry and Biochemistry, University of Bern, Freiestrasse 3, 3012 Bern, Switzerland

The increasing volume of biomedical data in chemistry and life sciences requires the development of new methods and approaches for their handling. In this presentation we discuss some challenges and opportunities of this fast growing area of research with a focus on those to be addressed within the BIGCHEM project. The presentation starts with a brief description of some available resources for “Big Data” in chemistry and a discussion of the importance of data quality. We then discuss challenges with visualization of millions of compounds by combining chemical and biological data, the expectations from mining the “Big Data” using advanced machine-learning methods, and their applications in polypharmacology prediction and target de-convolution in phenotypic screening. We show that the efficient exploration of billions of molecules requires the development of smart strategies. We also address the issue of secure activity data sharing without disclosing chemical structures, which is critical to enable bi-party or multi-party data sharing. Data sharing is important in the context of the recent trend of “open innovation” in pharmaceutical industry, which has led to not only more information sharing among academics and pharma industries but also the so-called “precompetitive” collaboration between pharma companies. At the end we highlight the importance of education in “Big Data” for further progress of this area.

The project leading to this presentation has received funding from the European Union’s Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 676434, “Big Data in Chemistry” (“BIGCHEM”, http://bigchem.eu). The abstract reflects only the authors’ view and neither the European Commission nor the Research Executive Agency are responsible for any use that may be made of the information it contains.