OPEN ACCESS Freely available online

ISSN: 2090-4568
Journal of Advanced Chemical Engineering

Perspective

Cheminformatics: A Drug Discovery Accessment Method

Mengyang Xu*

Department of Chemistry, University of Cambridge, London, UK

DESCRIPTION

Cheminformatics is the application of computer and informational techniques to a variety of chemistry problems. These techniques, also known as Cheminformatics and chemical informatics, are used in pharmaceutical companies during the drug discovery process. Cheminformatics combines the scientific fields of chemistry and computer science, particularly in the areas of chemical graph theory and chemical space mining. The chemical space is expected to contain at least 10-62 molecules. Cheminformatics is a broad term that refers to the creation, organization, management, retrieval, analysis, dissemination, visualization, and application of chemical information. The transformation of data into information and information into knowledge is a necessary endeavour in any branch of chemistry, not just drug design. Chemistry has generated a massive amount of data, which is rapidly increasing. More than 45 million chemical compounds are known, with the number growing by millions each year. Massive amounts of data are generated by novel techniques such as combinatorial chemistry and high-throughput screening. All of this data and information can be managed and made accessible only by storing it in appropriate databases.

Today, the vast majority of the hundreds of compound suppliers around the world provide free information on the molecular structures (and other properties) of their compounds for the purposes of virtual screening and other applications. Synthetic compounds dominate the majority of commercial compound collections. We discovered that only about 10% of the known NPs (approximately 25k) are readily available for experimental testing by overlapping a comprehensive collection of more than 250k NPs (which we compiled by curating and merging all of the Natural Products (NP) datasets available to us31) with the 7.3 million in-stock compounds listed in the ZINC database46, 47 (a comprehensive database of compounds that are available from various commercial sources and research institutes).

This confirms that the availability of experimental materials is the bottleneck in NP-based drug discovery. Allowing minor structural differences between NPs and purchasable compounds, i.e. including primarily NP derivatives and analogues, increases the number of readily available compounds by roughly 10k to 30k. It is also worth noting that the majority of readily available NPs have physicochemical properties that are advantageous in the context of drug discovery. Indeed, more than half of them are fragment-sized (molecular weight less than 300 Da), providing numerous opportunities for optimization.

Cheminformatics has been instrumental in characterising NPs based on their physicochemical and structural properties, as well as comparing NPs to small-molecule drugs, drug-like compounds, and other types of (organic) molecules. NPs cover a much larger chemical space than synthetic compounds, and they populate areas of chemical space that are generally inaccessible (or only with great difficulty) synthetically. Some NPs' structural uniqueness (and complexity) may allow them to target macromolecules that are otherwise unreachable.

NPs are heavier and more hydrophobic on average than synthetic drugs and synthetic, drug-like compounds. Their structural complexity is often higher as well, especially in terms of stereochemistry (commonly quantified by the number of chiral centers,57, 59-66, the fraction of Csp3 atoms,6, 8 and/or the number of bridgehead atoms in ring systems67) and 3D molecular shape.

CONCLUSION

NPs present extraordinary challenges to both experimentalists and theorists, but statistics on recently approved small-molecule medicines show that NP research is worthwhile and can yield valuable, innovative drugs. Modern *in silico* methods can contribute significantly to the acceleration and de-risking of NP-based drug discovery. However, model applicability must be carefully monitored, especially when working with NPs, because computational approaches are typically designed for and trained on data for synthetic compounds. Unfortunately, even recently developed models frequently lack robust definitions of the applicability domain and do not adequately warn users about compounds with unreliable predictions.

Correspondence to: Mengyang Xu, Department of Chemistry, University of Cambridge, London, UK, E-mail: xumen@gmail.com

Received: 02-Nov-2022, Manuscript No. JACE-22-19150; **Editor assigned:** 07-Nov-2022, Pre QC No. JACE-22-19150 (PQ); **Reviewed:** 22-Nov-2022, QC No. JACE-22-19150; **Revised:** 28-Nov-2022, Manuscript No. JACE-22-19150 (R); **Published:** 06-Dec-2022, DOI: 10. 35248/2090-4568.22.12.256.

Citation: Xu M (2022) Cheminformatics: A Drug Discovery Accessment Method. Adv Chem Eng. 12:256.