Cheminformatics' applications in drug discovery: still a long way to go?

Hai Pham-The[a]

[a] *Hanoi University of Pharmacy, 13-15 Le Thanh Tong, Hoan Kiem, Hanoi, Vietnam*

Email: haipham@hup.edu.vn or hpham.phd@gmail.com

Tel: +84-966-923-117.

ABSTRACT

Cheminformatics involves the use of computer technologies to process chemical data. It plays many roles in modern drug discovery high-throughput screening (HTS) research, including identifying drug targets and active compounds against those targets, HTS data mining, and design/prediction of lead compound biological activities and absorption, distribution, metabolism, excretion, and toxicology (ADMET) properties. In general, cheminformatics applications rely on three components: (i) Chemical structure featuring, (ii) data mining techniques, and (iii) suitable validation and application of cheminformatics tools. In the new drug discovery strategy, challenges of these integrating components have been posed in the following aspects: development of alternative methods to codify relevant and orthogonal chemical information, exploration of learning algorithm able to extract knowledge from large-scale raw HTS databases in a shorter time periods, development of efficient tools to predict ADMET properties, multi-objective optimization procedures for better hit-to-lead transformation, and so on. This discussion will revisit the theoretical aspects of the main components and outline some achievements of cheminformatics approaches applied in drug discovery during the past decades. Based on the advances and limitations of current cheminformatics tools, potential opportunities of this field in future directions are also suggested.