P83

Structural connectivity fingerprints – a new method of compound representation

Krzysztof Rataj[1], Wojciech Czarnecki[2], Sabina Podlewska[1,3], Andrzej J. Bojarski[1]

[1]Department of Medicinal Chemistry, Institute of Pharmacology Polish Academy of Sciences,12 Smętna Street, 31-343 Cracow, Poland
[2]Faculty of Mathematics and Computer Science, Jagiellonian University, 30-348 Kraków, Poland
[3]Faculty of Chemistry Jagiellonian University 3 Ingardena Street 30-060 Krakow

The current generation of substructural fingerprints, bit-string methods of representation of chemical compounds' structure, is based on sets of well-designed, pre-defined substructural keys. They are a commonly used method during virtual screening campaigns, since they are able to reject compounds that do not possess the vital chemical moieties. However, the methodology is somewhat flawed, since the fingerprints do not describe the positions of the groups, only their existence. This implies, that two compounds with dissimilar structures (but similar composition) can be described by almost identical fingerprints.

In this research we addressed this issue by designing new substructural fingerprint – the Substructural Connectivity Fingerprint. This new method of compound representation contains additional information about the interconnectivity of chemical groups within compound. To properly analyze this kind of data, currently existing methods of machine learning data analysis were employed, and a brand new method, Extreme Entropy Machines was implemented.

The initial tests conducted on compounds for 9 GPCRs and 5 protein kinases yielded very promising results, outperforming all currently available fingerprints. The fingerprint and the new method of data analysis will be expanded and validated in further research.