27O-ISMS24 Similarity Implicated Exploration of the Fragment Galaxy

Gábor IMRE[1], Ákos TARCSAY[1], ○Fumiaki ARUGA[2]
[1]ChemAxon Ltd., [2]Patcore, Inc.

Fragment space is a marginal subgroup of the druggable universe of chemical entities considering its population. However, it spans astonishingly high number of structures. GDB-13, the largest publicly available virtually enumerated collection counts nearly 1 billion structures of fragment size compounds. Chemically intelligent navigation in this vast dataset demands special purpose solutions. Our study focuses on making the very large chemical datasets live by ultra fast similarity search method. As a use-case, we search GDB-13 to find similar structures to that of the FDA approved drugs not exemplified in the space of patented structures available within SureChEMBL. This framework represents a scaffold hopping approach exploiting the GDB-13 under the hood and demonstrates the benefit of using MadFast SimilaritySearch technique.

About MadFast:
MadFast is a high-end toolkit for ultra fast chemical similarity search using in-memory data storage and optimized multi threaded-implementation. The outstanding search performance extends the chemical space available for live search to hundreds of millions of compounds. Rapid fingerprint generation and short initialization time, along with the large set of comparison methods, provide you with the possibility to optimize the similarity space. MadFast is a Java application that is available via versatile interfaces: command line, REST API and Web UI. Extensive documentation and usage examples are provided.