[P23] Virtual screening of small molecule libraries based on antiviral chemical space analysis

Anastasia Nikitina[1,2], Alexey Orlov[1,2], Evgeny Khvatov[2], Dragos Horvath[3], Vladimir Palyulin[1], Alexandre Varnek[3], Dmitry Osolodkin[1,2]

[1] Department of Chemistry, Lomonosov Moscow State University, 1 bd. 3 Leninskie Gory, Moscow 119991, Russia.
[2] Chumakov Institute of Poliomyelitis and Viral Encephalitides, Chumakov FSC R&D IBP RAS, 8 bd. 1, Poselok Instituta Poliomielita, Poselenie Moskovsky, Moscow 108819, Russia.
[3] Laboratory of Chemoinformatics, University of Strasbourg, UMR 7140, Strasbourg 67000, France.

More than 200 different viral species are known to cause human diseases. However, only a few viral infections are manageable by vaccination or specific treatment. Despite the accumulation of a huge amount of antiviral activity data in public databases such as ChEMBL and PubChem BioAssay, antiviral drug discovery projects are usually focused on a few “iconic” viruses capturing the attention of humans. As the number of viral pathogenic species is much larger, the concept ‘one bug-one drug’ is not economically feasible. Thus, the large-scale analysis of antiviral chemical space can be beneficial for new broad-spectrum antiviral compound design as well as for the guided navigation in antiviral chemical space.

Data extraction procedures used for big data treatment inevitably lead to inconsistencies affecting any subsequent analysis. Therefore, careful curation of data extracted from databases is needed. In this study we developed dictionary-based procedures for mining and systematization of the virus-related information in ChEMBL. These allowed to retrieve 3 times more activity data points, covering many more compounds tested in antiviral assays as compared to default Web Taxonomy Browser provided on ChEMBL website. Specific taxonomy annotation procedures based on analysis of ChEMBL text fields values were designed. Data were mapped to related viral species according to ICTV taxonomy of viruses. Structures and activities were standardized, that allowed to create an antiviral activity profile for each compound.

Two approaches based on self-organizing maps (SOM) and generative topographic mapping (GTM) algorithms were used for the screening of DrugBank and Drug Repurposing Hub libraries, as well as the catalogue of a local vendor InterBioScreen. These methods have been extensively used for ligand-based similarity searching. Applicability of developed procedures in compound repurposing was demonstrated in several examples. More than 400 compounds were selected for further experimental evaluation of their antiviral activity against flavivirus and enterovirus species.