AI in Medicinal Chemistry

Date: Thursday, November 5th
Time: 03:00 pm – 04:50 pm (CET)
Location: Kaltura Video Conferencing
Registration: Required
Organizing partners: Leiden Academic Centre for Drug Research (LACDR)

Program:

03:00 pm – 03:25 pm
Predicting drug-induced liver injury combining structural and bioactivity-based fingerprints
Olivier Bequignon, MSc

03:25 pm – 04:00 pm
Learning meaningful molecular representations for drug discovery
Dr. Djork-Arne Clevert

04:00 pm – 04:05 pm
Break

04.05 pm – 04:50 pm
Do androids dream of therapeutic molecules?
Prof. Alan Aspuru-Guzik

In this talk, I will review the use of AI methods for the inverse design of molecular systems: if a given target property is desired, how to search chemical space to find such molecule. I will also discuss automation approaches to close the loop between synthesis and testing in the area of materials.

04:50 pm
Closing

Organizing Partners:

LACDR Universiteit Leiden The Netherlands

Abstract: *Olivier Bequignon, MSc*

O. J. M. Béquignon[1], B. van de Water[1], G. J. P. van Westen[1]
[1]*Div. Drug Discovery and Safety, Leiden Academic Centre for Drug Research, Leiden, The Netherlands*

Drug-induced liver injury (DILI) is one of the main reasons of drug attrition during clinical trials and of withdrawal from the market[1] making the early identification of hepatotoxic of compounds a critical challenge. Recent in silico efforts have been made to create more accurate quantitative structure-property relationships (QSPR) models relating hepatotoxicity to chemical structure features[2,3]. One shortcoming of these models is that their design overlooked orthologous experimental in vitro results. Additionally, the underlying mechanisms of DILI are not completely understood yet

Aims
Designing a two-step approach, relying, firstly, on a proteochemometric model to predict translational target inhibition resulting in bioactivity fingerprints, and secondly, on a QSPR model predicting hepatotoxicity from combined structural and bioactivity-based fingerprints, we aim at deciphering the importance of proteins, hence helping in the elucidation of DILI mechanisms.

Methods
A dataset of experimental molecular bioactivity was developed by mining the ChEMBL25 and ExCAPE databases. Compound structures were standardized and described using Mold2[4] two-dimensional chemical features. When applicable, only the most expressed protein isoform was considered. Protein sequence descriptors using Hellberg et *al*'s z-scale[5] and averaged over 50 domains along with global average were derived. Extreme Gradient Boosting (XGBoost) models for inhibition at 10µM, 5µM, 1µM, 500nM and 100nM were designed. Stacked XGBoost models were then developed to predict binary DILI from concatenated extended connectivity fingerprints[6] of at most radius 3 folded over 1024 bits and bioactivity-based fingerprints.

Results / Conclusions
Preliminary results demonstrate that the integration of bioactivity-based fingerprints increases the prediction performance of DILI over QSPR models. Additionally, feature importance can be derived from the model, estimating the importance of each protein of the panel in the final prediction. Out of the 11 most important features, all but 4 were involved in cell-stress pathways associated with DILI, 3 were highly expressed in the liver, 8 were human proteins and 3 were rat. This work highlights the use of machine learning in deciphering complex biological mechanisms and in developing predictive models for DILI.

References:
1. Chen, M. et *al.*, W. FDA-Approved Drug Labeling for the Study of Drug-Induced Liver Injury. *Drug Discov. Today* **2011**, *16*(15–16), 697–703.
2. Ekins, S. Progress in Computational Toxicology. *J. Pharmacol. Toxicol. Methods* **2014**, *69*(2), 115–140.
3. Béquignon, O. J. M. et *al*. Computational Approaches for Drug-Induced Liver Injury (DILI) Prediction: State of the Art and Challenges. In *Systems Medicine*; Elsevier, **2021**; pp 308–329.
4. Hong, H. et *al*. Mold2, Molecular Descriptors from 2D Structures for Chemoinformatics and Toxicoinformatics. *J. Chem. Inf. Model*. **2008**, *48*(7), 1337–1344.
5. Hellberg, S. et *al*. Peptide Quantitative Structure-Activity Relationships, a Multivariate Approach. *J. Med. Chem.* **1987**, *30* (7), 1126–1135.
6. Rogers, D.; Hahn, M. Extended-Connectivity Fingerprints. *J Chem Inf Model*. **2010**, *50*(5), 742–754.