Chemoinformatic Tools, Compound Databases and Data Management Platforms

9 October 2023

14:00-17:00 CEST

online

Workshop of the COST Action CA21111

One Health drugs against parasitic vector borne diseases in Europe and beyond

OneHealthdrugs

BOOK OF ABSTRACTS

Description

High quality data on compound properties and activities form the backbone of the daily work for many members of the COST Action OneHealthdrugs. Chemoinformatics tools, compound repositories and data management platforms play an essential role in providing access to vital compound information and facilitating the analysis of compound properties for modelling of drug targets and developing compounds as potential drug candidates against parasitic diseases. The aim of this workshop is to introduce knowledge in using various chemoinformatics tools, accessing and navigating compound databases, and efficiently managing chemical data through data management platforms.

Open data resources for drug discovery at EMBL-EBI

Andrew Leach

European Bioinformatics Institute, Hinxton, UK

arl@ebi.ac.uk

ChEMBL (www.ebi.ac.uk/chembl) is a manually curated database of bioactive molecules with drug-like properties. First launched in 2009, it brings together a wealth of chemical, bioactivity, and genomic data relevant to drug discovery. The ChEMBL database has grown significantly since its first launch, and it provides high quality curated data, access-free, to the scientific community. Compound data in ChEMBL are taken from a range of sources, including publications, patents, and deposited data from collaborators. Data on approved drugs, and on drug candidates in clinical trials cover disease indications, mechanisms of action, drug warning information such withdrawn status or black box warnings, drug metabolism and prodrug status as well as other drug properties such as route of administration, first approval etc. All data are curated and standardized, both with manual and automatic checks; there are 2.4 million compounds, 1.6 million assays and 15,000 targets in the most recent ChEMBL release (ChEMBL 33).

ChEMBL is very widely used across academia, industry and not-for-profit institutes to address numerous practical problems and questions in drug discovery and chemical biology. I will describe some of the key features of ChEMBL, give some illustrative examples of its use in drug discovery, and indicate some of our future plans.

The European Lead Factory: 10 years of collaborative drug discovery

Vera Nies

Foundation Lygature, The Netherlands

vera.nies@lygature.org

The European Lead Factory (ELF) is a collaborative initiative funded by the IMI/IHI, established to bridge the innovation gap in early drug discovery and to tackle unmet medical needs. The ELF combines the best that the worlds of industry and academia have to offer: the expert knowledge and assets from major pharmaceutical companies, with the innovative power and agility of academia and SMEs.

ELF builds on a unique shared library of 550.000 compounds that was made accessible for screening by Europe-based academics and SMEs, as well as the pharmaceutical companies in the partnership. Part of the collection was brought in by the pharmaceutical partners, another part was de novo synthesized based on innovative ideas from the med-chem community [1].

Supported by a pharma-like HTS and HCS infrastructure and custom build software that allows for comprehensive triaging while protecting IP, more than 300 screens have been performed on a variety of targets and across a wide range of disease areas. Follow-up work resulted in over 100 peer reviewed publications, several patents, and 3 IND filings. Successes across a range of indications have led to subsequent development programs, further investments and start-up companies [2].

This presentation will explain the unique concept of the ELF and highlight some of its successes, explaining that the ELF is an unprecedented effective model for large scale drug discovery collaboration between academic groups and industry.

References

Van Vlijmen H, Ortholand J-Y, Li V M-J, et al., The European Lead Factory: An updated HTS compound library for innovative drug discovery, Drug Discovery Today d Volume 26, Number 10, 2021.

Jones P, Boucharens S, Elroy S, et al., IMI European Lead Factory — democratizing access to high-throughput screening. Nature Reviews Drug Discovery. 2022 Apr;21(4):245-246.

Disclaimer:

This project has received funding from the Innovative Medicines Initiative 2 Joint Undertaking (JU) under Grant Agreement No 806948: ESCulab. The JU receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA and Medicines for Malaria Venture. This communication reflects the views of the author(s). Hence, the IMI JU is not liable for any use that may be made of the information contained herein.

Chemography applications to drug design: from (ultra)large libraries analysis to *de novo* design of molecules and reactions

Alexandre Varnek

University of Strasbourg, France

varnek@unistra.fr

Chemography considers projections of compounds on 2-dimesional maps obtained from multi-dimensional space of molecular descriptors with the help of dimensionality reduction method. One of such methods - Generative Topographic Mapping (GTM) - allows to obtain both positioning of the compounds on a map and the data probability distribution function, which in turn, makes possible to describe (ultra)large compounds collections. GTM can efficiently be used for various chemoinformatics tasks [1-3]: (*i*) chemical data visualization and analysis, (*i*) prediction of properties or biological activities, (*iii*) comparison of large chemical databases, (v*i*) drugs repurposing and (v) virtual screening. In combination with SMILES-based autoencoder, cartography can be used for automatized generation of chemical structures with desired biological activities. A GTM constructed on the autoencoder latent variables provides a direct and intuitive access to the autoencoder chemical space, which sampling can be "driven" by the map toward the highly relevant zones of a drug discovery project [4].

References

1. N. Kireeva, I. Baskin, H. Gaspar, D. Horvath, G. Marcou, A. Varnek, *Mol. Inf.* 2012, **31**, 301-312
2. I. Casciuc, Y. Zabolotna, D. Horvath, G. Marcou, J. Bajorath, A. Varnek *J. Chem. Inf. Model.,* 2019, **59**(1), 564-572
3. Y. Zabolotna, A. Lin, D. Horvath, G. Marcou, D. M. Volochnyuk, A. Varnek *J. Chem. Inf. Model.* 2021, **61**, 1, 179–188
4. B. Sattarov, I. Baskin, D. Horvath, G. Marcou, E.J. Bjerrum, A. Varnek *J. Chem. Inf. Model.*, 2019, **59**(3), 1182-1196

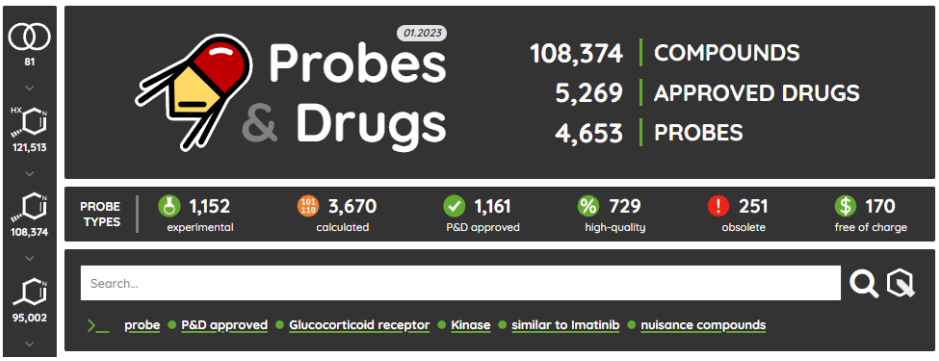
Probes & Drugs portal: a hub for the integration of high-quality bioactive compound sets

Ctibor Škuta[a], Petr Bartůněk[a]

[a]*CZ-OPENSCREEN: National Infrastructure for Chemical Biology, Institute of Molecular Genetics of the Czech Academy of Sciences, Czech Republic*

ctibor.skuta@img.cas.cz

Probes & Drugs (P&D) portal (probes-drugs.org)[1] was released in 2017 to address the problems in the field of high-quality chemical tools, concerning mainly the fragmentation and obsoleteness of the available data.[2] Since then, it has become one of the major and most comprehensive resources in the field that serves as a hub for the integration of high-quality bioactive compound sets, with focus on chemical probes and drugs. Apart from the data integration, P&D established its own approach to evaluating the probe-likeness of compounds labelled as chemical probes (so-called probe-likeness score) and a live set of high-quality chemical probes (utilizing the score in combination with other key criteria), updated with each new version of P&D.[3] As of ver. 01.2023, this set contains 729 compounds covering 558 primary targets. The synergy between integrated data sources and tools available for their analysis makes P&D a unique discovery platform designated not only for experts in the field of chemical biology but for a wider scientific community with limited knowledge in this area.



References

[1] Skuta C et al. Probes & Drugs portal: an interactive, open data resource for chemical biology. Nat Methods. 2017 Jul 28;14(8):759-760. doi: 10.1038/nmeth.4365. PMID: 28753599.

[2] Arrowsmith CH et al. The promise and peril of chemical probes. Nat Chem Biol. 2015 Aug;11(8):536-41. doi: 10.1038/nchembio.1867. PMID: 26196764; PMCID: PMC4706458.

[3] Skuta C, Southan C, Bartunek P. Will the chemical probes please stand up? RSC Med Chem. 2021 Jul 16;12(8):1428-1441. doi: 10.1039/d1md00138h. PMID: 34447939; PMCID: PMC8372204.

Acknowledgment

The Authors acknowledge The Ministry of Education, Youth and Sports (project LM2023052) for the financial support.

Integrating the ATC Database and Structural Analysis for Computational Repurposing of Drugs with Antiparasitic Potential

Marcus M.K. Nielsen[a,c], Bala Krishna Prabhala[b], Karen Angeliki[c]

[a]*Technical University of Denmark, Denmark;* [b] *University of Southern Denmark, Denmark;* [c]*Roskilde University, Denmark*

mmkni@dtu.dk

The repurposing of existing drugs for novel therapeutic indications has gained significant attention as an efficient approach for drug discovery. In this study, we propose a methodology for identifying potential antiparasitic drug candidates through the integration of the Anatomical Therapeutic Chemical (ATC) database, and structural analysis techniques in the KNIME platform.

Our approach focuses on the utilization of the "P" group within the ATC classification, which encompasses antiparasitic products, insecticides and repellents. Leveraging the KNIME platform, we retrieve scientific articles referencing drugs within this group and investigate their co-occurrence with drugs from diverse ATC groups. Additionally, we apply a structural analysis using the Tanimoto index, inspired by the pioneering work of Douglas Kell and his team [1].

Through the comparison of the co-occurrence patterns and structural analysis results, we aim to identify potential repurposable drugs that exhibit antiparasitic effects, despite not originally being approved for such indications. Our approach offers a comprehensive framework for drug repurposing and potential cross-functional therapeutic benefits.

To execute this methodology efficiently, we propose employing high-performance computing (HPC), ensuring expedited processing times for the extensive data analysis involved. By presenting an innovative strategy for identifying repurposable drugs with antiparasitic potential, this study contributes to the advancement of drug discovery paradigms, underscoring the significance of computational approaches in expanding therapeutic options.

References

[1] O'Hagan, S., Kell, D.B. Analysis of drug–endogenous human metabolite similarities in terms of their maximum common substructures. J Cheminform 9, 18 (2017). https://doi.org/10.1186/s13321-017-0198-y

High resolution chemical imaging of surfaces using infrared spectroscopic photo-induced force microscopy (PiF-IR) and its requirements for data analysis

Maryam Ali[a], Robin Schneider[b], Sebastian Unger[a,b], Ute Neugebauer[a,b,c] Rainer Heintzmann[a,b], Daniela Täuber[a,b]

[a] *Friedrich Schiller University Jena;* [b] *Leibniz Institute of Photonic Technology, Jena;* [c] *University Hospital Jena, Jena, Germany*

Maryam.ali@uni-jena.de

Tip-enhanced infrared spectroscopic imaging provides access to nanoscale chemical characterization of biological cells and tissue ex vivo [1]. Infrared spectroscopic photo-induced force microscopy (PiF-IR) is an innovative approach for chemical imaging of surfaces with a resolution of ≈ 5 nm. Figure 1 shows chemical contrasts of Bacillus s*ubtilis* incubated with Vancomycin for an individual bacteria cell [2].

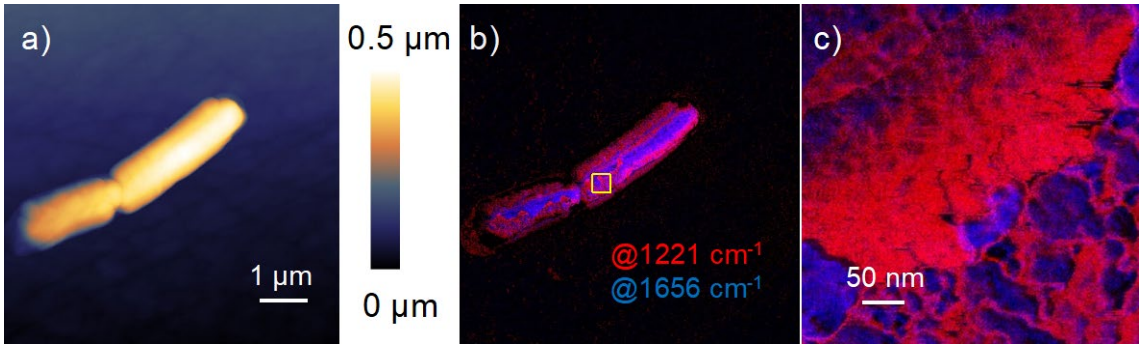


Figure 1: PiF-IR applied to Bacillus *subtilis* incubated with Vancomycin. a) AFM topography, b) RGB merge of PiF-IR contrasts from two succeeding scans with varied excitation frequencies; c) RGB merge of two succeeding high-resolution PiF-IR scans of the yellow area in b) with excitation at @1656 cm^{-1} (blue) and @1221 cm^{-1} (red) [2].

Localized antibiotic interactions can be further visualized via hyperspectral imaging ex vivo. From the investigation of pristine fibrillar Actin samples using principal component analysis of hyperspectral PiF-IR scans, we observed a local variation in the secondary protein structure [3]. The unique spatial and high spectral resolution of PiF-IR requires novel approaches for data analysis and the development of new reference tools. In particular, we are working on evaluating the application of topological data analysis (TDA) [4] for the investigation of hyperspectral scans obtained from our model system Bacillus *subtilis* treated by Vancomycin.

References

[1] H. Wang, Q. Xie, and X. G. Xu, "Super-resolution mid-infrared spectro-microscopy of biological applications through tapping mode and peak force tapping mode atomic force microscope," *Advanced Drug Delivery Reviews*, vol. 180, p. 114080, Jan. 2022, doi: 10.1016/j.addr.2021.114080.

[2] N. Krishnakumar *et al.*, "Nanoscale chemical imaging of cell surface and core-shell nanoparticles using infrared excitation and detection by force microscopy," in *70. Mosbacher Kolloquium - "High-resolution imaging of cells and molecules,"* Mosbach, 2019, doi: 10.13140/RG.2.2.13191.06564.

[3] J. Joseph *et al.*, "Nanoscale chemical characterization of secondary protein structure of F-Actin using mid-infrared photoinduced force microscopy (PiF-IR)," *Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy*, vol. ECSBM2022, in revision.

[4] L. Duponchel, "Exploring hyperspectral imaging data sets with topological data analysis - ScienceDirect," *Analytica Chimica Acta*, vol. 100, pp. 123–131, Feb. 2018.