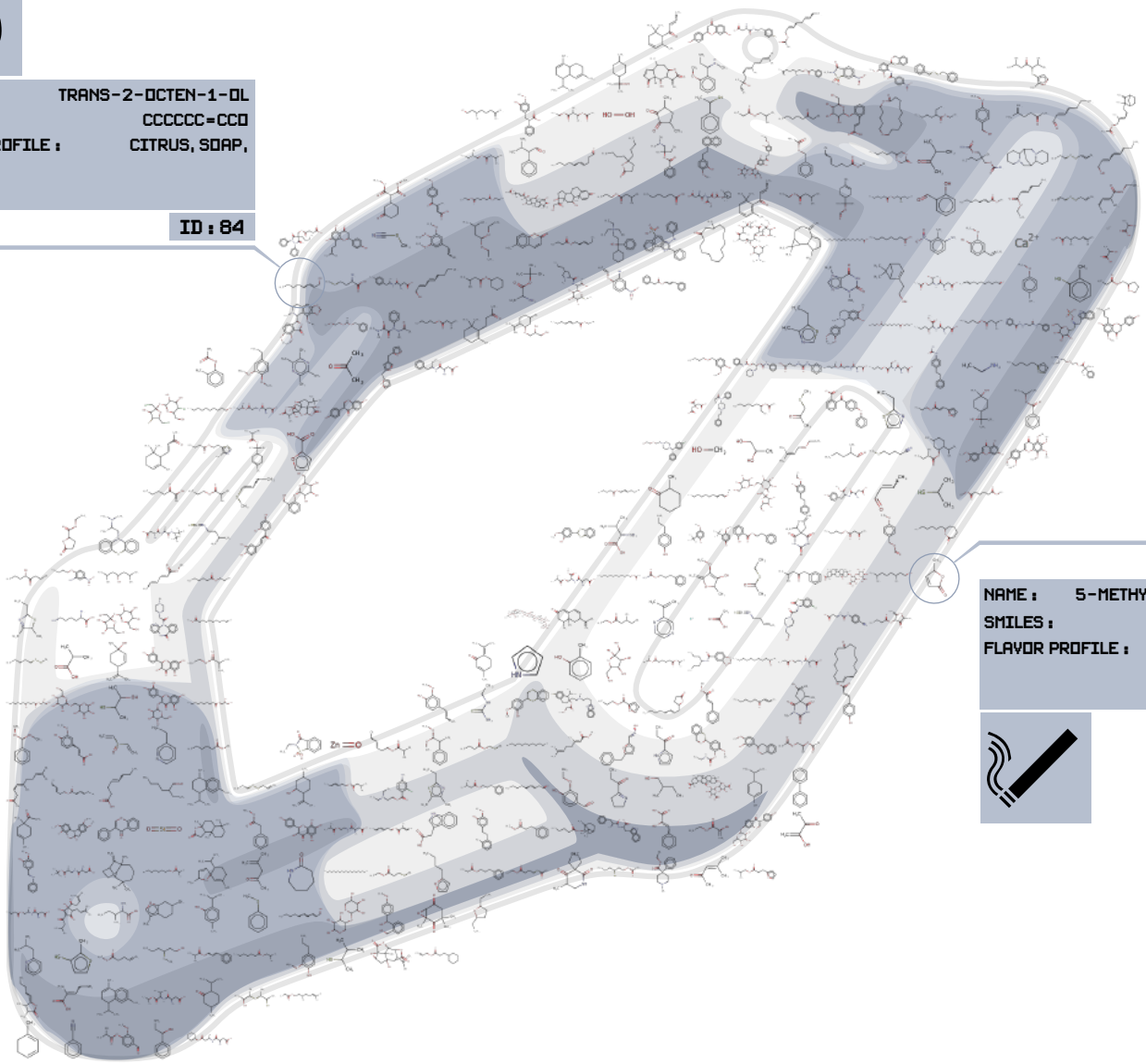
8TH CHEMOINFORMATICS STRASBOURG SUMMER SCHOOL

NAME : TRANS-2-OCTEN-1-OL
SMILES : CCCCCC=CCO
FLAVOR PROFILE : CITRUS, SOAP,

ID : 84



NAME : 5-METHYL-2(5H)-FURANONE
SMILES : CC1C=CC(=O)O1
FLAVOR PROFILE : SOLVENT, TOBACCO, COUMARIN, OILY, SWEET, NUTTY, TONKA

COLLEGE DOCTORAL EUROPEEN

46 BOULEVARD DE LA VICTOIRE

67000 STRASBOURG

27 JUNE - 1 JULY 2022

CONTENT

Program

Monday 27 June

14:00 - 16:40	***Registration***	
16:45 - 17:00	***Opening***	
17:00 - 18:00	**Jürgen BAJORATH** University of Bonn, Germany	Rationalizing Molecular Promiscuity through Data Analysis and Explainable Machine Learning
18:00 - 19:45	***Welcome Party***	

Tuesday 28 June

9:00 - 9:40	**Gisbert SCHNEIDER** ETH Zurich, Switzerland	*De Novo* Molecular Design with Machine Intelligence
9:40-10:20	**Emilio BENFENATI** Ist. Ricerche Farm. Milan, Italy	*In silico* models for the REACH and the food regulation: perspectives for the near future
10:20-10:40	***coffee break***	
10:40-11:20	**Connor COLEY** MIT, Cambridge, USA	Learning patterns of chemical reactivity from experimental data
11:20-12:00	**Matthias REREY** University of Hamburg, Germany	Chemoinformatics operating in Chemical Space
12:00-14:00	***Lunch***	
14:00-14:20	**Marcus GASTREICH**, BioSolveIT Germany	Why Everybody Talks About Chemical Space Exploration
14:20-14:40	**David RINALDO** Schrödinger GmbH	Finding Hits in Large Chemical Spaces by Combining Docking with Deep Learning
14:40-15:00	**Brice HOFFMANN**, IKTOS, France	Representing, predicting, and generating simple and complex peptides
15:00-17:00	**Tutorial 1 –Gilles MARCOU** University of Strasbourg, France	Tutorial on Generative Topographic Mapping Landscapes
17:00 - 19:00	***Poster session***	*Beer & Bretzel*

Wednesday 29 June

9:00 - 9:40	**Alex TROPSHA** University of North Carolina, USA	Development of Biomedical Knowledge Graphs and their application to drug discovery
9:40-10:20	**Hanoch SENDEROWITZ**, Bar Ilan University, Israel	Materials Informatics: The marriage of materials and data sciences
10:20-10:40	***coffee break***	
10:40-11:20	**Thierry LANGER** University of Vienna, Austria	Approaches to Next Generation Pharmacophore Modelling
11:20-12:00	**Pavel POLISHCHUK** Palacky University, Czech Republic	Explainable artificial intelligence: evolution, achievements and perspectives
12:00-14:00	***Lunch***	
14:00-14:20	**Francois BERENGER** University of Tokyo, Japan	Molecular Generation by Fast Assembly of (Deep)SMILES Fragments
14:20-14:40	**Joao AIRES-DE-SOUSA** New University of Lisbon, Portugal	ML prediction of C-H bond energies: calibration of DFT-based models with experimental data
14:40-15:00	**Matthieu MONTES** HESAM University, Paris, France	VTX: High-performance molecular structure and dynamics visualization software
15:00-17:00	**Tutorial 2 –Célien JACQUEMARD** University of Strasbourg, France	Comparison of binding sites for fragment based-drug discovery
17:00 - 19:00	***Poster session***	*Beer & Bretzel*

Thursday 30 June

9:00 - 9:40	**Ola ENGKVIST** AstraZeneca, Gothenburg, Sweden	AI for drug design an industrial perspective
9:40-10:20	**Johannes KIRCHMAIR** University of Vienna, Austria	Cheminformatics in Natural Product-based Drug Discovery
10:20-10:40	***coffee break***	
10:40-11:00	**Yuliana ZABOLOTNA**, University of Strasbourg, France	ChemSpace Atlas: Multiscale Chemography of Ultra-Large Libraries For Drug Discovery
11:00-11:20	**Črtomir PODLIPNIK** University of Ljubljana, Slovenia	COVID.SI - A Crowdsourced Drug Discovery Project
11:20-11:40	**Johanna GIOVANNINI** University of Caen, France	Towards DAG-based interactive pharmacophore exploration: application to the BCR-ABL ligand set
11:40-12:00	**Moritz WALTER** University of Sheffield, United Kingdom	Interpreting Neural Network Models for Toxicity Prediction by Extracting Learned Chemical Features
12:00-14:00	***Lunch***	
15:00-19:00	***Cultural Program***	
19:00	***Conference Dinner***	

Friday 1 July

9:00 - 9:40	**José L. MEDINA-FRANCO** National Autonomous University, Mexico	Advancing epigenetic drug discovery with epi-informatics
9:40-10:20	**Artem CHERKASOV** University of British Columbia, Canada	Deep Docking – the AI-enabled platform for advanced virtual screening
10:20-10:40	***coffee break***	
10:40-11:20	**Olivier TABOUREAU** Paris Cité University, France	Cheminformatics and Network science applied with high content screening data
11:20 – 12:00	**Nikolaus STIEFL** Novartis, Switzerland	Augmenting Drug Hunters with Generative Chemistry Models
12:00	***Closure***	

Poster session

1	AKHMETSHIN Tagir	HyFactor: Hydrogen-count labelled graph-based defactorization autoencoder
2	ASGARKHANOVA Farah	Computer-aided design of selective chemical probes of angiotensin-converting enzyme 2
3	BAYBEKOV Shamkhal	Prediction of DMSO solubility for fragment-based screening
4	BORT William	Harnessing the “creativity” of AI to generate novel chemical reactions
5	CHEN Ya	Cheminformatic Analysis of Ring Systems in Natural Products
6	CHIESA Luca	A new machine learning based method for ADRB2 agonist detection using single-ligand dynamic interaction data
7	BENKAIDALI Lydia	Visualization and analysis of metabolomic space of Alzheimer’s disease using Generative topographic mapping
8	LEGEHAR Ashenafi	Drugmapper: a Web Resource to Explore Active Pharmaceutical Ingredients (Apis)
9	MERVEILLE KOSSIWA Eguida	Protein subpocket cloud comparison revealed similarity between HIV-1 reverse transcriptase and tumor necrosis factor binding sites
10	GAMBACORTA Nicola	PLATO: a user-friendly web platform for target fishing and bioactivity prediction
11	GAWALSKA Alicja	Application of automated machine learning in search for multi-target-directed ligands blocking PDE4B, PDE8A and TRPA1 ion channel with potential use in the treatment of asthma and COPD
12	GESLIN Damien	Deciphering a pharmacophore network generated from BCR-ABL data
13	LEHEMBRE Etienne	Towards DAG-based interactive pharmacophore exploration
14	IWATA Michio	Dynamic sensitivity analysis to predict time-course drug transcriptomic responses of the cellular system
15	JAMROZIK Marek	Computer-aided search for new anthracycline antibiotic reductases inhibitors with a potential to support anticancer therapy
16	JONCZYK Jakub	Molecular modelling and machine learning techniques in search for novel SARS-CoV-2 main protease inhibitors
17	KAITOH Kazuma	Scaffold-Retained Structure Generator to Extensively Produce Molecules with Unique Chemical Substructures
18	LAMANNA Giuseppe	DeLA-Drug: A Deep Learning Algorithm for Automated Design of Drug-like Analogues
19	HANDA Koichi	Prediction of Compound Plasma Concentration Time Profiles after Oral Administration in Mice Using Random Forests
20	OUDAHMANE Mehdi	Flexible Protein-Ligand Docking Protocol: a Case Study on the Androgen Receptor
21	PEREBYINIS Mariana	Assessment of the overlap between the ‘on-the-shelf’ drug-like space and ultra-large ‘on-demand’ combinatorial libraries

22	PEREZ-PENA Helena	Structure-guided design of novel tubulin binders – towards site-specific cysteine targeting
23	PIKALYOVA Karina	HIV-1 drug resistance profiling using amino acid sequence space cartography
24	PIKALYOVA Regina	Exploration of the Chemical Space of DNA-Encoded Libraries
25	PLYER Luis	Implementation of a Soft Grading System for Chemistry in a Moodle Plugin
26	REVILLO IMBERNON Julia	Comprehensive analysis of commercial fragment libraries
27	RUGARD Marylène	Chemo-biological analysis applied to the olfaction field
28	SANTHAPURI Sai Prashanth	Visualization and analysis of metabolomic space of Alzheimer's disease using Generative topographic mapping
29	SELLAMI Asma	Predicting ligand binding to nuclear receptors using a pipeline combining docking and pharmacophore models
30	SHERMUKHAMEDOV Shokirbek	Machine-learning for predicting material properties with atomistic potentials
31	SINDT François	Protein-applied computer vision and deep generative linking to generate potent kinase inhibitors: Influence of fragments definition
32	TROMELIN Anne	Study of odorants sharing the odor notes of an aroma blending mixture by a pharmacophore approach
33	VOLKOV Mikhail	Applicability of graph neural networks to binding affinities prediction from protein-ligand structures
34	WALTER Moritz	Interpreting neural network models for toxicity prediction by extracting learned chemical features
35	OLENEVA Polina	GTM-based analysis of the chemical space of the Chimiothèque Nationale
36	ZABOLOTNA Yuliana	Synt-On: A New Open-Source Tool for Synthon-Based Library Design and Building Blocks Analysis
37	ZABOLOTNA Yuliana	ChemSpace Atlas: Multiscale Chemography of Ultra-Large Libraries for Drug Discovery
38	ZANKOV Dmitry	Multi-Instance Learning Approach to Predictive Modeling of Catalyst Enantioselectivity
39	ZHANG Chonghuan	Exploration of bioinformatic domain based on data mining, reaction predictions and enzyme promiscuous predictions
40	PETER Sonja	Computational elucidation of GPCR allosteric modulators
41	VILLACAMPA Marina	Machine Learning to Discover Antibiotics Against *Klebsiella Pneumoniae*
42	REHIOUI Hajar	Pharmacophores *vs* circular fingerprints with learned feature transformation before clustering. Comparative studies on Bcr-Abl data.
43	PINEL Philippe	Large-step scaffold hopping benchmark

PLENARY LECTURES

[L1] Rationalizing Molecular Promiscuity through Data Analysis and Explainable Machine Learning

Jürgen BAJORATH

Department of Life Science Informatics, B-IT, LIMES Program Chemical Biology and Medicinal Chemistry, University of Bonn, Endenicher Allee 19c, D-53115 Bonn, Germany

Multi-target activity of small molecules, also termed promiscuity, leads to desired and undesired effects in drug discovery. Exploring the ability of small molecules to form pseudospecific interactions with different targets is of interest to better understand molecular recognition phenomena and devise multi-target drug design strategies. In addition to proteomics or target profiling, compound promiscuity can also be investigated computationally, for example, through systematic analysis of structural and activity data and diagnostic machine learning for hypothesis testing. These studies confirm the presence of structural features that distinguish multi- and single-target compounds. Explaining machine learning predictions reveals structural characteristics of promiscuous compounds.

[L2] *De Novo* Molecular Design with Machine Intelligence

Gisbert SCHNEIDER

ETH Zurich, Department of Chemistry and Applied Biosciences, Zurich, Switzerland; ETH Singapore SEC Ltd, Singapore.

Molecular design may be regarded as a pattern recognition process. Chemists are skilled in visual chemical structure recognition and their association with (retro)synthesis routes and molecular properties. In this context, various “artificial intelligence” (AI) methods have emerged as potentially enabling technology for drug discovery and automation, because these systems aim to mimic the chemist’s pattern recognition process and take it to the next level by considering the available domain–specific data and associations during model development. Part of the appeal of applying AI methods in drug design lies in the potential to develop data-driven, implicit model building processes to navigate vast datasets and to prioritize alternatives. This concept represents at least a partial transfer of decision power to a machine intelligence, and could be viewed as synergistic with human intelligence; that is, a domain-specific implicit AI that would augment the capabilities of chemists in molecular design and selection. More ambitiously, the ultimate challenge for drug design with AI is to autonomously generate new chemical entities with the desired properties from scratch (de novo), without the need for the often prohibitively costly experimental compound screening.

We will review the principles of AI methods for de novo drug design, emphasizing ligand-based approaches that have proven useful and reliable in “little-data” scenarios. Selected prospective case studies will be presented, ranging from targeted molecular design to fully automated design-make-test-analyze cycles. We provide a critical assessment of the possibilities and limitations of the individual approaches and dare forecasting the future of drug design with machine intelligence.

References:

Grisoni, F., Huisman, B., Button, A., Moret, M., Atz, K., Merk, D., Schneider, G. (2021) Combining generative artificial intelligence and on-chip synthesis for de novo drug design. Science Advances 7, eabg3338.

Friedrich, L., Cingolani, G., Ko, Y.-H., Iaselli, M., Miciaccia, M., Perrone, M. G., Neukirch, K., Bobinger, V., Merk, D., Hofstetter, R. K., Werz, O., Koeberle, A., Scilimati, A., Schneider, G. (2021) Learning from nature: From a marine natural product to synthetic cyclooxygenase-1 inhibitors by automated de novo design. Advanced Science 8, 2100832.

Schneider, P., Walters, W. P. Plowright, A. T., Sieroka, N., Listgarten, J., Goodnow Jr., R. A., Fisher, J., Jansen, J. M., Duca, J. S., Rush, T. S., Zentgraf, M., Hill, J. E., Krutoholow, E., Kohler, M., Blaney, J., Funatsu, K., Luebkemann, C., Schneider, G. (2020) Rethinking drug design in the artificial intelligence era. Nature Reviews Drug Discovery 19, 353–364.

Schneider, G. (2018) Automating drug discovery. Nature Reviews Drug Discovery 17, 97–113.

[L3] *In silico* models for the REACH and the food regulation: perspectives for the near future

Emilio BENFENATI

Laboratory of Chemistry and Environmental Toxicology, Mario Negri Institute for Pharmacological Research, Milano, Italy

In silico models can and should provide tools to reduce the impact of chemical substances in our life, minimizing the risk for the human health and the environment. These tools should be cheap and interconnected. The policy and society require documentation and transparency. The in silico models should cope with multiple features, since they should explore toxicodynamic and toxicokinetic properties, related to human and environment, within the One Health strategy. At the same time, these tools should be linked with tools used by industry, with approaches suitable to address functional use, to investigate in the same system both adverse and beneficial properties.

We will provide examples of tools able to deal with tens of models simultaneously, addressing hazard and exposure at the same time, merging numerical models and predictive ones. The JANUS software for instance integrates 48 models for prioritization, the VERMEER tools predict risks for specific scenarios, and the ToxEraser tools suggest safer substances, to replace the riskiest ones.

The specific regulatory thresholds are incorporated into the tools, which apply batteries of the VEGA models, depending on the needs. The VEGAHUB system offers these solutions (ww.vegahub.eu). However, the challenges are many more, and networking solutions have to be planned, facilitating links between different platforms. These are explored within the CONCERT REACH and the OptiTox projects, for instance, addressing the REACH and the food regulations.

[L4] Learning patterns of chemical reactivity from experimental data

Connor COLEY

Department of Chemical Engineering, Department of Electrical Engineering and Computer Science Massachusetts Institute of Technology, Cambridge, USA

The abundance of chemical reaction data in tabulated databases has enabled new data-driven approaches in reaction informatics. In particular, data-driven programs for Computer-Aided Synthesis Planning have rapidly matured and can now propose plausible synthetic pathways for many druglike compounds. We will discuss cheminformatic and machine learning-based approaches for learning patterns of chemical reactivity to perform the core tasks of synthesis planning: retrosynthesis, reaction condition recommendation, and reaction outcome prediction.

[L5] Chemoinformatics operating in Chemical Space

Matthias RAREY

University of Hamburg, ZBH – Center for Bioinformatics, Bundesstraße 43, 20146 Hamburg, Germany

With the rise of large make-on-demand chemical fragment spaces the need to directly operate in chemical space rather than in chemical libraries emerged. Due to the shear size, enumerating the spaces to libraries is highly inefficient and energy wasting, in many cases even impossible. Many methods working with chemical spaces are heuristic, i.e. they do not give any optimality guarantee. But this hasn't to be like this. Chemoinformatics standard operations like similarity and substructure search can be done directly in chemical fragment space with little to no approximation loss on standard desktop computers. In this talk, chemical space algorithms, their performance data and some applications will be presented.

[L6] Development of Biomedical Knowledge Graphs and their Application to Drug Discovery

Alexander TROPSHA

Laboratory for Molecular Modeling, UNC Eshelman School of Pharmacy, University of North Carolina, Chapel Hill, NC, 27599, USA.

The volume of biomedical research data stored in various databases has grown immensely in recent years due to the proliferation of high-throughput biomedical '-omics' technologies. Nearly all of respective databases, or 'knowledge sources' (KSs), address a particular area of biomedical research, leading to natural diversity but also growing disintegration between individual KSs, which generates downstream inefficiencies when mining diverse databases for knowledge discovery. Expanding efforts, both in academia and industry, are focused on the development of methods and tools to enable semantic integration and concurrent exploration of disparate biomedical KSs, using specially constructed biomedical 'graph knowledgebases' (GKBs) that support the generation of new knowledge through the application of reasoning tools and algorithms. Our group has contributed to these efforts by initiating the development of a GKB-based question-answering system termed Reasoning Over Biomedical Objects linked in Knowledge-Oriented Pathways (ROBOKOP) [1], [2]. ROBOKOP's publicly accessible user interface (UI) [3] allows users to address both relatively simple questions such as "what genes are associated with drug-induced liver injury?" and more complex ones such as "what drugs could be used to treat airborne pollutant–induced asthma exacerbations in patients who are non-responsive to traditional medications?" I will discuss the development of ROBOKOP and provide examples of applications including the elucidation of Clinical Outcome Pathways of drug action and drug repurposing including methodologies relying on knowledge graph embedding [4].

Bibliography:

[1] K. Morton et al., "ROBOKOP: An abstraction layer and user interface for knowledge graphs to support question answering," Bioinformatics, vol. 35, no. 24, pp. 5382–5384, Dec. (2019), doi: 10.1093/bioinformatics/btz604.
[2] C. Bizon et al., "ROBOKOP KG and KGB: Integrated Knowledge Graphs from Federated Sources.," J. Chem. Inf. Model., vol. 59, no. 12, pp. 4968–4973, Dec. (2019), doi: 10.1021/acs.jcim.9b00683.
[3] "ROBOKOP." https://robokop.renci.org/.
[4] C. Moon et al., "Learning Drug-Disease-Target Embedding (DDTE) from knowledge graphs to inform drug repurposing hypotheses," J. Biomed. Inform., vol. 119, p. 103838, Jul. (2021), doi: 10.1016/J.JBI.2021.103838.

[L7] Materials Informatics: The marriage of materials and data sciences

Omer KASPI,[a] Hadar BINYAMIN,[a] Abraham YOSIPOF,[b] Hanoch SENDEROWITZ[a]

[a] *Department of Chemistry, Bar Ilan University, Ramat Gan, 5290002, Israel*
[b] *Department of Information Systems, College of Law & Business, Ramat-Gan, P.O.Box 852 Bnei Brak 5110801, Israel*

Materials informatics is rapidly developing as evidenced by the large increase in the number of publications in this field. This development is fueled by the continuous growth of available data, both experimental and computational, on one hand and by the ability to draw from the rich repertoire of methods in its sister field, chemoinformatics, on the other. Similar to chemoinformatics, materials informatics can help bridge the so-called data-knowledge gap by offering smart ways to navigate the enormous materials space in search of new materials with favorable properties.

A key computational technique in materials informatics is Machine Learning (ML). In this talk, we will therefore comparatively analyze the similarities and differences between materials informatics and cheminformatics, within the framework of what is required to derive reliable ML-based models.

Next we will focus on the application of ML methods to the study of solar cells. Such cells hold the potential to meet the growing worldwide demand for clean energy. Today most solar cells are based on silicon yet new alternatives are continuously emerging including organic photovoltaic cells, dye sensitized solar cells (DSSCs), and metal-oxide (MO)-based solar cells. However despite significant progress, all these cells could benefit from improvements in key components, e.g., the dye in DSSCs and the MO composition of MO-based solar cells. Thus, we will present how the concept of pharmacophore can help identify new dyes with favorable (predicted) electronic properties for DSSCs and how combining combinatorial materials sciences with ML can lead to predictive models for key solar cells parameters such as current, voltage and quantum efficiency.

Finally, we will shift our attention to the field of forensic informatics and in particular to the usage of ML in order to analyze physical evidence found in crime scenes. This work will highlight the usefulness of experimentally determined spectra, in particular such that reflect the elemental composition of such evidence. In particular, we will describe the development of a reliable of ML-based classifier from glass fragments retrieved from different types of car windshields using ion beam analysis.

As a final take home message, we emphasize the need to conduct research in the field of materials informatics in close collaboration with experimentalists in order to provide insight into the observed trends and to capitalize on the results.

[L8] Approaches to Next Generation Pharmacophore Modelling

Thierry LANGER

Department of Pharmaceutical Sciences, Pharmaceutical Chemistry Division, University of Vienna, Althanstrasse 14, 1090 Vienna, Austria

Chemical feature based 3D pharmacophore models have been used for several decades supporting medicinal chemists in their early drug discovery programs. (1) In this presentation, an overview on recent approaches in further developing the field of pharmacophore modeling is given.

At Inte:Ligand GmbH, we developed the program LigandScout (2) as an integrated software solution containing rapid and efficient tools for automatic interpretation of ligand-protein interactions and subsequent transformation of this information into 3D chemical feature-based pharmacophore models. In addition, pattern recognition-based algorithms were developed for ligand-based pharmacophore modeling in the absence of a target 3D structure, as well as for establishing a novel and accurate virtual technique.

Since recently, we study the possibility to transfer the pharmacophore concept from a static approach to a dynamic one, by analyzing molecular dynamics simulation trajectories, in order to develop pharmacophore ensembles representing the dynamic event of binding (3) and to analyze them using both grid-based probability functions (4) as well as hierarchical graphs. (5) First results obtained from frequency information are indicating that MD simulations can add significantly to the refinement such models, by guiding the user to add or remove pharmacophore features, depending on their stability during the simulation and use them to increase prediction power in virtual screening. (6) Machine learning has been utilized to transform qualitative pharmacophore models to quantitative ones. (7)

Finally, as an extension of this approach, parallel pharmacophore-based screening has been introduced as an innovative in silico method to predict the potential biological activities of compounds by screening them with a multitude of pharmacophore models, e.g. in order to predict molecular initiating events finally leading to neurotoxic outcomes. We recently have made available this approach as a LigandScout Extension Workflow Node within the NeuroDeRisk KNIME platform. (8,9)

Bibliography:

[1] Langer, T., Pharmacophores in Drug Research, Mol. Inf. (2010), 29, 470.
[2] Wolber, G., Langer, T.; LigandScout: 3D Pharmacophores Derived from Protein-Bound Ligands and their Use as Virtual Screening Filters, J. Chem. Inf. Model. (2005), 45, 160.
[3] Wieder, M., Perricone, U., Boresch, S., Seidel, T., Langer, T.: Evaluating the stability of pharmacophore features using molecular dynamics simulations, Biochem. Biophys. Res. Comm. (2016), 470, 685.
[4] Schütz, D. A., Seidel, T., Garon, A., Martini, R., Körbel, M., Ecker, G.F., Langer, T. GRAIL: GRids of phArmacophore Interaction fieLds, J. Chem. Theory Comput. (2018), 14, 4958.
[5] Garon, A., Wieder, O., Bareis, K., Seidel, T., Ibis, G., Bryant, S. D., Theret, I., Ducrot, P., Langer, T. Hierarchical Graph Representation of Pharmacophore Models. Front Mol Biosici (2020), 7, 599059.
[6v Wieder, M., Garon, A., Perricone, U., Boresch, S., Seidel, T., Almerico, A.M., Langer, T. Common Hits Approach: Combining Pharmacophore Modeling and Molecular Dynamics Simulations. J. Chem. Inf. Model. (2017), 57, 365.
[7] Kohlbacher, S. M., Langer, T., Seidel. T. QPHAR: quantitative pharmacophore activity relationship: method and validation. J Cheminform (2021), 13, 57
[8v KoNstanz Information MinEr, available from KNIME.COM AG, Zurich, Switzerland (https://knime.org), LigandScout Extensions available from Inte:Ligand GmbH, Vienna, Austria (https://www.inteligand.com).
[9] The NeuroDeRisk in silico Toolbox (https://neuroderisk.eu/in-silico-toolbox/)

[L9] Explainable artificial intelligence: evolution, achievements and perspectives

Pavel POLISHCHUK

Institute of Molecular and Translational Medicine, Faculty of Medicine and Dentistry, Palacky University, Hnevotinska 5, 77900, Olomouc, Czech Republic

From the very beginning QSAR models followed human intuition. They used clearly interpretable descriptors and simple methods to establish structure-activity relationships. This made these models transparent and easy to understand for researchers. Information retrieved from such models could help to understand the underlying mechanisms or could be used to guide further design steps of compounds with improved properties. Further development of QSAR models was focused on improvement of their predictive ability that was successfully achieved by introduction of novel descriptors and machine learning methods. That made the resulting models more obscure due to their very complex internal structure or using non-interpretable descriptors. These models started to be considered as “black boxes” and a popular belief appeared that a trade-off between predictivity and explainability of models exists. This was true to some extent until the development of interpretation approaches which could estimate contributions of atoms or fragments directly from models. In theory, these approaches are applicable to any kind of predictive models that makes all such models interpretable.

More recently, neural networks gained much attention and wide applicability in many fields including chemoinformatic. These models are used to predict properties of molecules, reaction outcomes and optimal conditions, generate new chemical entities with desired properties, etc. Due to the high complexity of neural networks they could capture hidden biases in datasets or spurious correlations that lower the credibility in them. Therefore, the interest in interpretation of neural networks arose greatly last years. Multiple approaches were suggested and many of them were adapted in chemoinformatics. Some of them are model-specific, others have a wider applicability. The large number of these approaches makes it difficult to choose a proper method for interpretation of a particular model. Therefore, certain steps were performed to create specific data sets to benchmark existing and developing approaches. It was demonstrated that not all interpretation approaches were able to retrieve proper structure-activity relationships and their careful investigation is required. However, approaches which were developed previously are also applicable to explain decisions of neural network models.

Despite of all successes in explainability of complex machine learning models the interpretation is still in infancy and not widely used in research work. However, it may provide great advantages and feedback for specialist beyond the chemoinformatic field. Therefore, integration of these approaches in research pipelines and their active use for making decisions can be considered as a major challenge for explainable artificial intelligence in the near future.

This work was funded by the INTER-EXCELLENCE LTARF18013 project (MEYS), the European Regional Development Fund - Project ENOCH (No. CZ.02.1.01/0.0/0.0/16_019/0000868), ELIXIR CZ research infrastructure project (MEYS Grant No: LM2018131).

[L10] AI for Drug Design an Industrial Perspective

r
Ola ENGKVIST

Discovery Sciences, R&D, AstraZeneca Gothenburg, Sweden

Artificial Intelligence has become impactful during the last few years in chemistry and the life sciences, pushing the scientific boundaries forward as exemplified by the recent success of AlphaFold2. In this lecture I will provide an overview of how AI have impacted drug design in the last few years, where we are now and what progress we can reasonably expect in the coming years. There will be an emphasis on how AI for drug design is applied in the industry. The presentation will have a focus on deep learning based molecular de novo design, however, also aspects of synthesis prediction, molecular property predictions and chemistry automation will be covered.

[L11] Cheminformatics in Natural Product-based Drug Discovery

Johannes KIRCHMAIR

Computational Drug Discovery and Design Group (COMP3D), Department of Pharmaceutical Sciences, Division of Pharmaceutical Chemistry of the University of Vienna, Austria

Natural products (NPs) remain the most prolific resource of inspiration for small-molecule drug discovery. Computational methods can make a substantial contribution to NP research and the design of NP-inspired drugs. This lecture aims to provide an overview of the scope and limitations of modern cheminformatics methods for NP research, such as approaches for virtual screening, target prediction, ADME/T prediction and many other applications. Further to that, we will present our latest works on the (i) assessment of the NP chemical space and its relevance to drug discovery, (ii) analysis of NP ring systems and their representation by synthetic compounds, and (iii) target prediction for structurally complex NPs.

[L12] Advancing Epigenetic Drug Discovery with Epi-Informatics

José L. MEDINA-FRANCO

Research group DIFACQUIM, Faculty of Pharmacy, National Autonomous University of Mexico (UNAM), Mexico

A broad range of computational approaches collectively called “epi-informatics” are increasingly used to advance epigenetic drug and probe discovery. Herein, we discuss the recent advances in epi-informatics to chart the epigenetic relevant space and guide the development of targeted libraries. We also discuss the applications of computational approaches to guide the identification of small molecules active against one or more epigenetic targets. In particular, we will cover current trends of machine learning models generated based on sizeable public compound databases annotated with biological activity and implemented in a free webserver. As a case study, we will present the identification of potent and dual inhibitors of DNA and histone methyltransferases. In addition to showing low micromolar enzymatic inhibition, the small molecules are also active in various cell lines. The hit compounds were identified from synthetic screening libraries focused on epigenetic targets after an exhaustive analysis of the diversity and coverage of the chemical space. Computational approaches helped to rationalize the activity at the molecular level.

[L13] Deep Docking – the AI-enabled Platform for Advanced Virtual Screening

Artem CHERKASOV

Vancouver Prostate Centre, Faculty of Medicine, University of British Columbia

With the recent explosion of chemical libraries beyond billion molecules, more efficient virtual screening approaches are needed. The Deep Docking (DD) platform enables up to hundred-fold acceleration of structure-based virtual screening by docking only a subset of a chemical library, iteratively synchronized with a ligand-based prediction of the remaining docking scores. This method results in hundreds-to-thousands fold virtual hit enrichment (without significant loss of potential drug candidates) and hence, enables screening billion-sized chemical libraries without using extraordinary computational resources. Herein we present the generalized DD protocol that has been proven successful in a variety of computer-aided drug discovery (CADD) campaigns and can be applied in conjunction with any conventional docking program.

[L14] Cheminformatics and Network Science Applied with High Content Screening Data

Olivier TABOUREAU

CMPLI team (Inserm U1133), BFA unit (CNRS -UMR 8251), Université de Paris, Paris, France

For over a decade, computational chemical biology has contributed to a wide array of scientific tasks from analytical chemistry and biochemistry to pharmacology and toxicology. With the increasing availability of data from the"omics" technologies, we start to be able to profile chemical effect, not only at the molecular level, but also at more complex layers (cells, tissues, organs) allowing a better understanding of the mechanism of action underlying complex diseases.

Recently, phenotypic drug discovery has re-emerged as promising approaches in the identification and development of novel and safe drugs. Although, phenotypic screening does not rely on knowledge of specific drug targets, combination of chemical biology data with network sciences and cheminformatics give the opportunity to suggest therapeutic targets and mechanisms of actions induced by drugs and associated with an observable phenotype.

In our laboratory, we have developed a system pharmacology network integrating chemicals-proteins/genes-pathway-disease relationships and high content imaging-based high-throughput phenotypic profiling assays.

First, I will present the implementation of this system pharmacology network and then I will go through different examples showing how such analysis might be of interest in the study of chemical action across multiple scales of complexity from molecular and cellular to phenotypes and diseases.

[L15] Augmenting Drug Hunters with Generative Chemistry Models

Nikolaus STIEFL

Novartis, Basel, Switzerland

Small-Molecule Drug discovery is a multi-objective optimization problem in which finding the next drug candidate depends on various characteristics of compounds including efficacy, pharmacokinetics and safety. In the design process of small molecule drugs, medicinal chemistry project teams routinely face this complex multidimensional optimization challenge. Given the massive size of the relevant "chemical space" (estimated to be in the range of up to 10^{60} drug-like molecules), the key question for medicinal chemists is *"What is the best compound to make and test next"*. While humans are extremely good in understanding the bigger picture, computers/algorithms are potentially much better in coming up with and evaluating a large body of complementary solutions – such as the described multidimensional optimization problem.

In this session, the concept of Generative Chemistry and how it approaches the above-mentioned optimization problem will be introduced based on a Novartis in-house initiative. Examples from medicinal chemistry project applications will be provided to highlight how such an *in silico* decision-support system can assists medicinal chemists in multi-objective compound design, selection and prioritization.

SHORT COMMUNICATIONS

[SC1] Why Everybody Talks About Chemical Space Exploration

Marcus GASTREICH

Sr. Director Application Science, BioSolveIT Germany

Until a few years ago, and besides corporate compound repositories of merely a few million compounds, companies screened the same idea pool for IP novelty: eMolecules, ZINC, the Enamine catalog and the like. This, for obvious reasons, not only limits the IP space from which to mine but it hampers knowledge gaining due to the duplication character of the “novelty” sources.
Since very few years,[1] both companies and even academic institutions therefore mine from new, incredibly large spaces, sometimes with enormous computational effort.[2] The talk shall briefly shed light on the strategic reasons behind this paradigmatic change in the field.

Dating back to collaborations with Roche,[3] Boehringer Ingelheim,[4] Pfizer,[5] and several other Big Pharma organizations, BioSolveIT has created infrastructures that allow searching these vast spaces — the counts of which sometimes even exceeds zetta ranges [10^{21}]. Such numbers cannot be accesses with traditional, enumeration-based methods and needs an exploitation of combinatorics.

With a focus on didactics, the talk shall provide a “feeling” for the sizes and today’s breadth of approaches,[6] with an emphasis on the application side, rather than of going into details of the technical implementations ideas of which largely root in academic collaborations by the Rarey Lab in Hamburg.
Please note that Matthias Rarey will also provide a lecture at this school.

Bibliography :

[1] Hoffmann, T.; Gastreich, M. The next Level in Chemical Space Navigation: Going Far beyond Enumerable Compound Libraries. *Drug Discov. Today* 2019, *24* [5], 1148–1156. https://doi.org/https://doi.org/10.1016/j.drudis.2019.02.013.
[2] Warr, W. A. NIH Meeting “Ultra-Large Databases in Chemistry” https://chemrxiv.org/articles/preprint/Report_on_an_NIH_Workshop_on_Ultralarge_Chemistry_Databases/14554803 [accessed Aug 12, 2021].
[3] Rarey, M.; Stahl, M. Similarity Searching in Large Combinatorial Chemistry Spaces. *J. Comput. Aided. Mol. Des.* 2001, *15* [6], 497–520. https://doi.org/10.1023/A:1011144622059.
[4] Lessel, U.; Wellenzohn, B.; Lilienthal, M.; Claussen, H. Searching Fragment Spaces with Feature Trees. *J. Chem. Inf. Model.* 2009, *49*, 270–279. https://doi.org/10.1021/ci800272a.
[5] Boehm, M.; Wu, T. Y.; Haussen, H.; Lemmen, C. Similarity Searching and Scaffold Hopping in Synthetically Accessible Combinatorial Chemistry Spaces. *J. Med. Chem.* 2008, *51*, 2468–2480. https://doi.org/10.1021/jm0707727.
[6] Warr, W. A.; Nicklaus, M. C.; Nicolaou, C. A.; Rarey, M. Exploration of Ultralarge Compound Collections for Drug Discovery. *J. Chem. Inf. Model.* 2022, *ASAP*. https://doi.org/10.1021/acs.jcim.2c00224.

[SC2] Finding Hits in Large Chemical Spaces by Combining Docking with Deep Learning

David RINALDO[1], Kun YAO[2], Matthew P. REPASKY[3], Karl LESWING[2], Robert ABEL[2], and Steven V. JEROME[4]

[1] Schrödinger GmbH, Glücksteinallee 25, 68163 Mannheim, Germany
[2] Schrödinger, Inc., New York, New York 10036, United States
[3] Schrödinger, Inc., Portland, Oregon 97239, United States
[4] Schrödinger, Inc., San Diego, California 92121, United States

On-demand synthesizable screening libraries have been growing very rapidly in the recent years to reach several tens billions of compounds. Exploring such large and diverse chemicals spaces in screenings would enable the discovery of more-potent hits and new scaffolds. But applying physics-based virtual screening methods in an exhaustive manner on such big libraries would be cost-prohibitive.

Here, we introduce a protocol[1,2] for machine learning-enhanced molecular docking based on active learning to dramatically increase throughput over traditional docking. We will see how such approach enables the identification of the best scoring compounds and the exploration of a large region of chemical space. Together with automated redocking of the top compounds, this method captures almost all the high scoring scaffolds in the library found by exhaustive docking.

The performances of this protocol were assessed on virtual screening campaigns, and we observed it can produce several highly potent, novel inhibitors at a reduced computational cost but preserving the diversity of the experimentally confirmed hit compounds.

Bibliography :

[1] Y. Yang *et al*., J. Chem. Theory Comput. 17, (2021), 7106-7119
[2] K. Konze *et al*., J. Chem. Inf. Model. 59, (2019) 3782-3793

[SC3] Representing, predicting, and generating simple and complex peptides

Massina ABDERRAHMANE[1], Brice HOFFMANN[1], Nicolas DEVAUX[1], Maud JUSOT[1]

[1]*Iktos, 65 rue de Prony, 75017 Paris, France.*

Peptides are a class of drugs in a growing therapeutic space between small molecules and biologics. More than 80 peptide drugs have reached the market for a wide range of diseases, including diabetes, cancer, osteoporosis, multiple sclerosis, HIV infection and chronic pain[1].
Predicting peptide properties using machine learning methods has gained interest in recent years, with a particular focus on anticancer[2], antimicrobial[3] or improving permeability[4] properties, using publicly available datasets.

Since 2018, multiple generative AI approaches for peptides have also been proposed. Methods include three popular deep generative model frameworks: neural language models (NLMs), variational autoencoders (VAEs), and generative adversarial networks (GANs)[5].
These predictive and generative approaches have shown interesting performances, but also limitations, especially on the type of amino acids considered, most often restricted to natural amino acids.

In this work, we have developed new representations of peptides, including graph representations at the amino-acid, backbone - side chain and pharmacophore level. These representations are suitable for peptides constituted of natural residues, but also for complex peptides which contain modified amino acids and cross-links. We applied a circular algorithm to transform these graphs into a vectorial representation and evaluated them for predictive tasks, including permeability. Finally, we proposed generative models based on the new representations proposed here, able to design new peptides with optimized properties.

Bibliography:

[1] Muttenthaler, Markus, et al. "Trends in peptide drug discovery." Nature Reviews Drug Discovery 20.4 (2021): 309-325.
[2] Basith, Shaherin, et al. "Evolution of machine learning algorithms in the prediction and design of anticancer peptides." Current Protein and Peptide Science 21.12 (2020): 1242-1250.
[3] Rondon-Villarreal, Paola, Daniel A Sierra, and Rodrigo Torres. "Machine learning in the rational design of antimicrobial peptides." Current Computer-Aided Drug Design 10.3 (2014): 183-190.
[4] Wei, Huan-Huan, et al. "The development of machine learning methods in cell-penetrating peptides identification: a brief review." Current Drug Metabolism 20.3 (2019): 217-223.
[5] Wan, Fangping, Daphne Kontogiorgos-Heintz, and Cesar de la Fuente-Nunez. "Deep generative models for peptide design." Digital Discovery (2022).

[SC4] Molecular Generation by Fast Assembly of (Deep)SMILES Fragments

Francois BERENGER[1], Koji TSUDA[1]

[1]*Graduate School of Frontier Sciences, The University of Tokyo, 5-1-5 Kashiwa-no-ha, Kashiwa, Chiba 277-8561, Japan.*

In recent years, in silico molecular design is regaining interest. To generate on a computer molecules with optimized properties, scoring functions can be coupled with a molecular generator to design novel molecules with a desired property profile. In this poster, a simple method is described to generate only valid molecules at high frequency (>300000 molecule/s using a single CPU core; Figure 1), given a molecular training set. The proposed method generates diverse SMILES (or DeepSMILES) encoded molecules while also showing some propensity at training set distribution matching. When working with DeepSMILES, the method reaches peak performance (>340000 molecule/s) because it relies almost exclusively on string operations. The "Fast Assembly of SMILES Fragments" software is released as open-source at https://github.com/UnixJunkie/FASMIFRA.

Experiments regarding speed, training set distribution matching, molecular diversity and benchmark against several other methods are also shown.

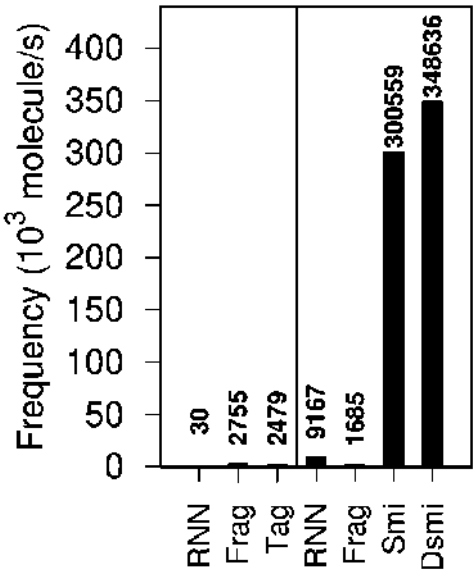


Figure 1: Model training (left) and sampling speed (right). RNN: Recurrent Neural Network (numbers cited from literature [2]); Frag (train): molecular fragmentation (Python/RDKit); Tag: tagging cleaved bond (proposed method, Python/RDKit); Frag (sampling): assembly of molecular fragments using molecular graph operations (Python/RDKit); Smi: fast assembly of SMILES fragments (proposed method, OCaml). Dsmi: fast assembly of DeepSMILES fragments (proposed method, OCaml); Tag is the model training prerequisite of Smi and Dsmi sampling. All methods use a single CPU, except RNN which uses four CPUs and one GPU.

Bibliography:

[1] Berenger, F., & Tsuda, K. (2021). Molecular generation by Fast Assembly of (Deep) SMILES fragments. Journal of Cheminformatics, 13(1), 1-10.

[2] Arús-Pous, J., Blaschke, T., Ulander, S., Reymond, J. L., Chen, H., & Engkvist, O. (2019). Exploring the GDB-13 chemical space using deep generative models. Journal of cheminformatics, 11(1), 1-14.

[SC5] ML prediction of C-H bond energies: calibration of DFT-based models with experimental data

Wanli LI, Yue LUAN, Qingyou ZHANG[1*], Joao AIRES-DE-SOUSA[2*]

[1] Henan Engineering Research Center of Industrial Circulating Water Treatment, Henan Joint International Research Laboratory of environmental pollution control materials, Henan University, Kaifeng, 475004, PR China; qingyou@vip.henu.edu.cn
[2] LAQV and REQUIMTE, Departamento de Química, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, 2829-516 Caparica, Portugal; jas@fct.unl.pt

Random Forest QSPR models were developed with a data set of homolytic bond dissociation energies previously calculated by B3LYP/6-311++G(d,p)//DFTB for 2263 C-H covalent bonds between hydrogen and sp3 carbon atoms [1]. The bonds were represented by atomic descriptors of the carbon atom – counts of 34 atom types (specified by the element and number of H neighbors) in spheres around the atom, and sizes of rings incorporating the atom. The best set of atomic attributes consisted in 114 ring descriptors and counts of atom types in 5 spheres around the kernel atom. The optimized model predicted the bond energies of an independent test set of 224 bonds with R^2=0.85, MAE=2.86 kcal/mol and RMS=4.57 kcal/mol. A new data set of 409 bonds from the ibond database [2] was predicted by the RF and compared with the experimental energies [3]. A modest MAE (5.36 kcal/mol) but a relatively high R^2 (0.74) suggested a systematic deviation. A prediction scheme was thus explored that corrects the RF prediction with the average deviation observed for the k nearest neighbors in an additional memory of experimental data. The corrected predictions achieved R^2=0.87, MAE=2.22 kcal/mol and RMS=2.94 kcal/mol in experiments with an independent test set of 145 bonds and the corresponding experimental bond energies. Such a protocol relied on RF similarities for the definition of the KNN distance between objects, and performed better than a KNN-only prediction.

Acknowledgments:
This work was supported with financial support from Fundação para a Ciência e Tecnologia (FCT, MCTES) Portugal, under grant UIDB/50006/2020 and UIDP/50006/2020 (provided to the Associate Laboratory for Green Chemistry LAQV). JAS thanks David Ponting and co-workers at Lhasa Limited for useful suggestions and discussions. This work was also supported by the National Natural Science Foundation of China [Grant number 21875061, 21975066] and the program for Science & Technology Innovation Team in Universities of Henan Province [Grant number 19IRTSTHN029].

Bibliography:

[1]. X. Qu; D.A.R.S. Latino; J. Aires-de-Sousa J. Cheminform. 5 (2013) 34.
[2]. http://ibond.nankai.edu.cn
[3]. P.C. St. John;Y. Guan;Y. Kim; et al. Nat. Commun. 11 (2020) 2328.

[SC6] VTX: High-performance molecular structure and dynamics visualization software

Maxime MARIA[1,2], Simon GUIONNIERE[1], Nicolas DACQUAY[1], Yassine NAIMI[3], Jean-Philip PIQUEMAL[4], Guillaume LEVIEUX[5], and Matthieu MONTES[1]

[1] *Laboratoire GBCM, EA758, Hesam Universite, CNAM ;*
[2] *Laboratoire XLIM, Universite de Limoges ;*
[3] *Qubit Pharmaceuticals SAS ;*
[4] *LCT, UMR7616, Sorbonne Universite ;*
[5] *Laboratoire CEDRIC, EA 4626, CNAM*

Molecular visualization is a critical task usually performed by structural biologists and bioinformaticians to aid three processes that are essential in science and fundamental to understand structural molecular biology: synthesis, analysis and communication [1].
Here we present VTX, a new molecular visualization software that includes a real-time high-performance molecular graphics engine dedicated to the visualization of large structure and dynamics of molecular systems. It is capable to process most molecular structures and trajectories file formats. VTX disposes of an interactive camera system controllable via the keyboard and/or mouse that includes different modes: 1. a classical trackball mode where the cam-era revolves around a fixed focus point and 2. a first-person free-fly navigation mode where the user fully controls the movement of the camera. VTX includes an intuitive and highly usable graphical user interface and tools designed for expert and non-expert users. It is free for non-commercial use at http://vtx.drugdesign.fr

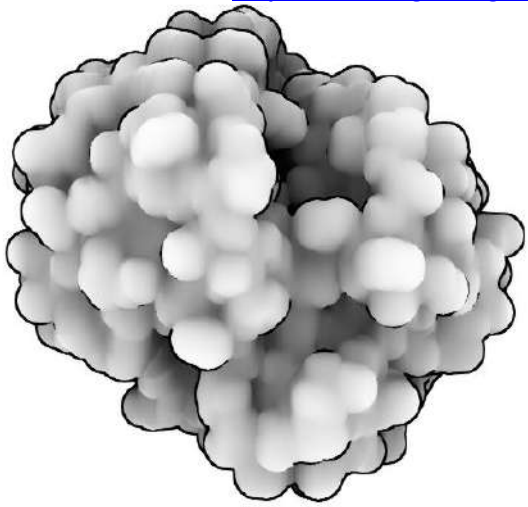


Figure 1. Illustration of VTX Image export of a black and white flat color SAS of human haemoglobin (PDB ID:1A3N).

Bibliography:

[1] Olson, AJ. Perspectives on structural molecular biology visualization: from past to present. J Mol Biol (2018) 430(21): 3997–4012.

.

[SC7] ChemSpace Atlas: Multiscale Chemography of Ultra-Large Libraries for Drug Discovery

Yuliana ZABOLOTNA, Fanny BONACHERA, Dragos HORVATH, Olga KLIMCHUK, Arkadiy LIN, Gilles MARCOU, Alexandre VARNEK

Laboratoire de Chémoinformatique, UMR7140 CNRS, Institut Le Bel, University of Strasbourg, 4 rue Blaise Pascal, 67081, Strasbourg, France

Nowadays, drug discovery is inevitably intertwined with the usage of large compound collections. Understanding of their chemotype composition and physicochemical property profiles is of the highest importance for successful hit identification. However, the development of efficient polyfunctional tools allowing multi-faceted analysis of the constantly growing chemical libraries is complicated by the Big Data challenges.

Thus, we present ChemSpace Atlas (https://infochm.chimie.unistra.fr/) - an intuitive polyvalent tool for the ultra-large chemical space exploration and analysis with respect to medicinal chemistry problems. Being based on the hierarchical ensemble of tens of thousands of Generative Topographic Maps(GTM)[1], it provides access to ChEMBL, ZINC and COCONUT collections allowing easy navigation through the hundreds of millions of compounds from a global bird's eye view to structural pattern detection[2]. ChemSpace Atlas functionality is not limited to a simple visualization of the similarity relationships in the chemical space but it also allows users to analyze physicochemical properties and biological activities, perform polypharmacological profiling (around 750 biological activities), analogs search, and detailed structural analysis with the help of MCSs and scaffolds. In Figure 1 one can see the example of the activity visualization page for CDK4 (CHEMBL331) ligand series from ChEMBL.

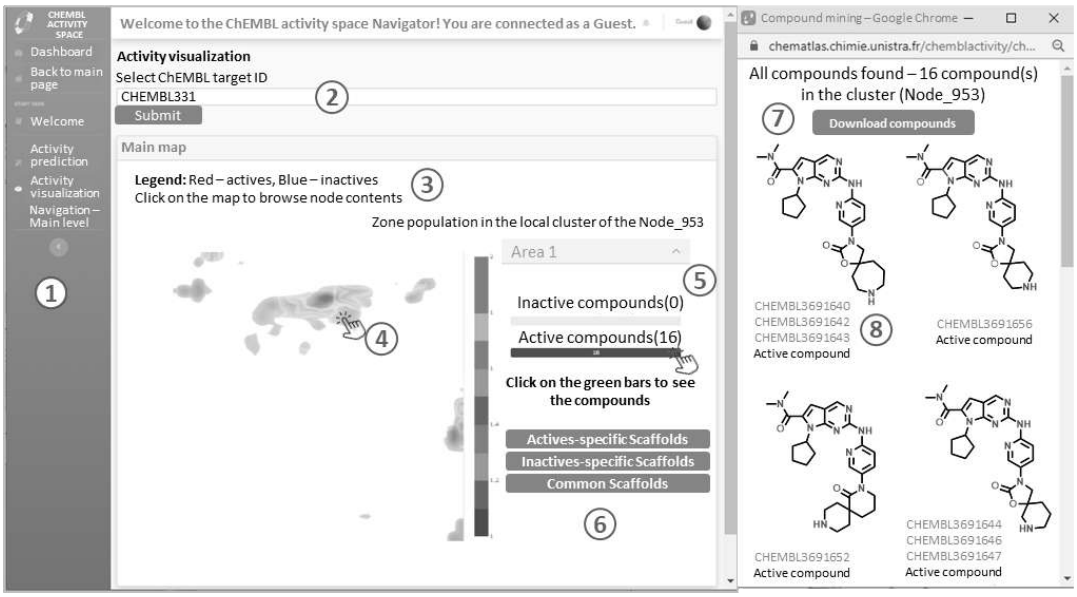


Figure 1. Activity visualization page of the ChEMBL activity space Navigator on the example of CDK4 (CHEMBL331) ligands. 1) navigation bar; 2) target selection menu; 3) legend of the map; 4) interactive activity landscape; 5) zone population information (if green, bars become clickable and corresponding compounds can be displayed); 7) button for downloading compounds from the selected area of the chemical space; 8) direct links to the source database (here ChEMBL).

Bibliography:

[1] C. M. Bishop, M. Svensén, C. K. I. Williams, Neural Comput. (1998), 10, 215-234.
[2] A. Lin, B. Beck, D. Horvath, G. Marcou, A. Varnek, J. Comput. Aided Mol. Des. (2019).

[SC8] COVID.SI - A Crowdsourced Drug Discovery Project

Črtomir PODLIPNIK[1], Sebastian PLEŠKO[1], Gašper TOMŠIČ[2], Boštjan LABA[3], Jukić MARKO[4]

[1]Faculty of Chemistry and Chemical Technology, University of Ljubljana, Večna pot 113, 1000 Ljubljana, Slovenia.
[2]Independent researcher, Cesta Cirila Kosmača 66, 1000 Ljubljana, Slovenia
[3]Independent researcher, Neolink d.o.o., Ponoviška cesta 11b, 1270 Litija, Slovenia
[4]Faculty of Chemistry and Chemical Technology, University of Maribor, Smetanova 17, 2000 Maribor, Slovenia.

The project COVID.SI has been running since March 2020. As the name suggests, the main goal of the project is to find compounds that are complementary to receptor targets associated with COVID -19 disease pathology. The core of our project is the virtual screening (docking) of ten million purchasable compounds against drug targets using distributed computing. We use an in-house developed server-client solution (https://github.com/ COVID -si). The server programme, written in Javascript, is incredibly robust and extensible, and node cloning should not cause any problems. Clients are available for Linux, macOS and Windows. The docking software used is CmDock (https://gitlab.com/Jukic/cmdock), a fork of RxDock, a fast, versatile, and open-source programme for docking ligands to proteins and nucleic acids.

To date, we have tested 10 million compounds against various targets, mainly related to COVID. Some of the results from our virtual screening campaign are already being used to prioritise compounds for 3CLpro inhibitor development [1]. The COVID.SI project also has a Boinc-based sister project, SIDOCK@HOME, developed with colleagues in Petropavlovsk, Russia. It is designed to screen billions of compounds for pharmacologically important biomolecular targets, which today are primarily associated with COVID -19. [2]

Our goal is to expand the project to other important targets associated with other dangerous virus-borne diseases such as Ebola, Zika, Nipah, and the parasitic diseases malaria and trypanosomiasis.

Bibliography :

[1] Jukič M, Škrlj B, Tomšič G, Pleško S, Podlipnik Č, Bren U. Prioritisation of Compounds for 3CLpro Inhibitor Development on SARS-CoV-2 Variants. Molecules. 2021 May 18;26(10):3003.
[2] Nikitina N., Manzyuk M., Jukić M., Podlipnik Č., Kurochkin I., Albertian A. (2021) Toward Crowdsourced Drug Discovery: Start-Up of the Volunteer Computing Project SiDock@home. In: Voevodin V., Sobolev S. (eds) Supercomputing. RuSCDays 2021. Communications in Computer and Information Science, vol 1510. Springer, Cham.

[SC9] Towards DAG-based interactive pharmacophore exploration: application to the BCR-ABL ligand set

Johanna GIOVANNINI[2#], Etienne LEHEMBRE[1#], Jean-Luc LAMOTTE[2], Abdelkader OUALI[1], Alban LEPAILLEUR[2], Albrecht ZIMMERMANN[1], Bertrand CUISSART[1], Ronan BUREAU[2], Bruno CREMILLLEUX[1]

[1]GREYC, Normandie Univ., UNICAEN, CNRS – UMR 6072, 14000 Caen, France
[2]CERMN, Normandie Univ, UNICAEN, CERMN, 14000 Caen, France

In medicinal chemistry, a pharmacophore denominates a spatial arrangement of chemical features which is responsible for a favorable interaction with the binding site of the target. We have recently designed a method that automatically computes pharmacophores from a large data set of molecules without any prior supervised selection of a small subset of molecules [1, 2]. The connections between the computed pharmacophores provide a hierarchical organization: the pharmacophore network. The latter is structured by the pharmacophores' size and contains a large number of them. The current work aims at developing a structure which a medicinal chemist can use to support his analysis without having to repeatedly mine pharmacophores. For this purpose, we enhance the knowledge on the pharmacophore network organization by taking into account parent-children relations and/or grouping the pharmacophores into equivalence classes, i.e. sets of pharmacophores occurring in the exact same molecule group. Additionally, each pharmacophore is annotated with one or several quality measures (e.g. confidence or growth rate measurements), information that will be exploited later-on.

The directed acyclic graph (DAG) is built and organized in layers which contain pharmacophores with the same number of pharmacophoric features, as well as the molecules that respond to them. Layers are linked by establishing parent-children relationships based on a sub-graph relationship. The DAG can then be used to apply different filtering and selection algorithms, e.g. based on user-specified quality measures, the visualization of the effect on each layer. In order not to overwhelm the expert, the visualization is condensed by clustering nodes in each layer using the above-mentioned equivalence classes. We will also use the DAG to identify Pharmacophore Activity Delta (PAD). PADs are syntactic pharmacophores families, i.e. pharmacophores linked via the afore-mentioned parent-child relationships exhibiting a minimum amount of syntactic similarity, yet members of which show a significant difference between the values of their quality measures.
Interactive Pattern Mining [3] is one of the objective of this work related to ANR project InvolvD (ANR-20-CE23-0023, https://involvd.greyc.fr/). This DAG was used to simulate the effects of users or functions interacting with the search space during the search itself.
The DAG process was applied to BCR-ABL chemogenomic dataset (ChEMBL1862, ChEMBL_24). The BCR-ABL tyrosine kinase is an oncogene associated to chronic myeloid leukemia. We aim to expose our method and how far the PAD analysis highlighted original chemical structural features linked to active or inactive ligands.

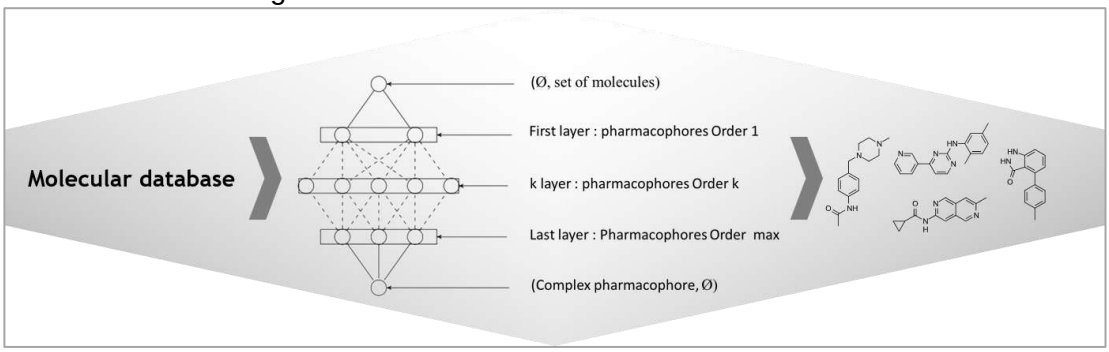


Figure: From a chemical database to the highlight of structural features of interest, we aim to present how our pharmacophore DAG can be used for fast 2D structural discrimination.

Bibliography:
[1] Métivier *et al*., J. Med. Chem. 2018, 61.
[2] Geslin *et al*., J. Chem. Inf. Model. 2022, 62.
[3] Van Leeuwen, Interactive knowledge discovery and data mining in biomedical informatics. Springer, 2014.

[SC10] Interpreting Neural Network Models for Toxicity Prediction by Extracting Learned Chemical Features

Moritz WALTER[1], Samuel J. WEBB[2], Valerie J. GILLET [1]

[1]Information School, University of Sheffield, Regent Court, 211 Portobello, S1 4DP, Sheffield, UK.
[2]Lhasa Limited, Granary Wharf House, 2 Canal Wharf, LS11 5PS, Leeds, UK

Deep Neural Network (DNN) models have become a popular machine learning technique for toxicity prediction of chemicals. Due to their complex structure it is difficult to understand predictions made by these models which limits confidence. Current approaches to tackle this problem such as SHAP or integrated gradients provide insights by attributing importance to input features of individual compounds.[1,2] While these methods have produced promising results, they do not shed light on representations of compounds in hidden layers. In the realm of image classification models, feature visualization has been developed as a tool to depict features learned in specific hidden layer neurons.[3] Detecting complex chemical features learned in hidden layers might complement current approaches for interpreting DNN models used for toxicity prediction.

The present study focuses on feedforward neural networks with RDKit's Morgan fingerprints as input. A novel method was developed to automatically extract chemical features responsible for activation of hidden neurons. This method leverages both information about training compounds strongly activating hidden neurons and learned model parameters. Using Ames mutagenicity as a well-understood toxicity endpoint, the method was able to extract known toxicophores. Moreover, extracted substructures can be mapped onto test compounds to obtain model explanations incorporating hidden layer representations of compounds. Using toxicophores from the Derek expert system[4] as ground truth, the explanatory capability of the approach was evaluated using attribution AUCs as metric.[5] For the majority of compounds provided explanations match the ground truth very well with AUC values above 0.8.

The proposed method may be used to extract novel toxicophores by leveraging chemical features encoded in DNN models. Furthermore, understanding of model predictions is increased by providing explanations complementary to those obtained with established attribution methods. While not explored in the present study, the proposed method could be adapted to other DNN architectures such as graph-convolutional neural networks.

Bibliography:

[1] Rodriguez-Perez, R., et al.; Interpretation of Compound Activity Predictions from Complex Machine Learning Models Using Local Approximations and Shapley Values. J. Med. Chem. 63 (2020) 8761-8777.
[2] Preuer, K., et al.; Interpretable Deep Learning in Drug Discovery. arXiv :190302788v2 (2019).
[3] Olah, C., et al.; Feature Visualization. Distill (2017).
[4] Marchant, C., et al.; In Silico Tools for Sharing Data and Knowledge on Toxicity and Metabolism: Derek for Windows, Meteor and Vitic. Toxicol. Mech. Methods. 18 (2008) 177-187.
[5] McCloskey, K., et al.; Using Attribution to Decode Binding Mechanisms in Neural Network Models for Chemistry. PNAS. 116 (2019) 11624-11629.

POSTERS

[P1] Hyfactor: Hydrogen-Count Labelled Graph-Based Defactorization Autoencoder

Tagir AKHMETSHIN[1,2], Arkadii LIN[1], Timur MADZHIDOV[2] and Alexandre VARNEK[1]

[1] *Laboratory of Chemoinformatics, UMR 7140, CNRS University of Strasbourg, 4 rue Blaise Pascal, 67000 Strasbourg, France*
[2] *Chemoinformatics and Molecular Modelling Laboratory, Kazan Federal University, 18 Kremlyovskaya str., 420008 Kazan, Russia*

One of the new trends in the generation of chemical structures is the use of graph-based neural networks[1]. Here we present a novel open-source autoencoder architecture, HyFactor[2]. This network operates with a hydrogen-count labelled graph (HLG), where the number of hydrogens attached to heavy atoms is used instead of bond types (

Figure *2*). HyFactor was benchmarked on the ZINC 250K, MOSES and ChEMBL data sets, against the conventional graph-based architecture ReFactor. The latter represents our implementation of the DEFactor architecture[3]. In average, HyFactor models contain some 20% less training parameters than required by ReFactor. Both ReFactor and HyFactor showed high (>90%) reconstruction rates both in ZINC250K and ChEMBL datasets, which is similar (or even better) than earlier reported graph-based or SMILES-based approaches. The two architectures display similar validity and uniqueness in the molecular generation task. Compared to the training set compounds, HyFactor generates more similar structures than ReFactor. This could be explained by the fact the latter generates many open-chain analogues of cyclic structures in the training set. It has been demonstrated that the reconstruction error of heavy molecules can be significantly reduced using the data augmentation technique. The code of HyFactor and ReFactor as well as all models obtained in this study are publicly available from our GitHub repository: https://github.com/Laboratoire-de-Chemoinformatique/HyFactor

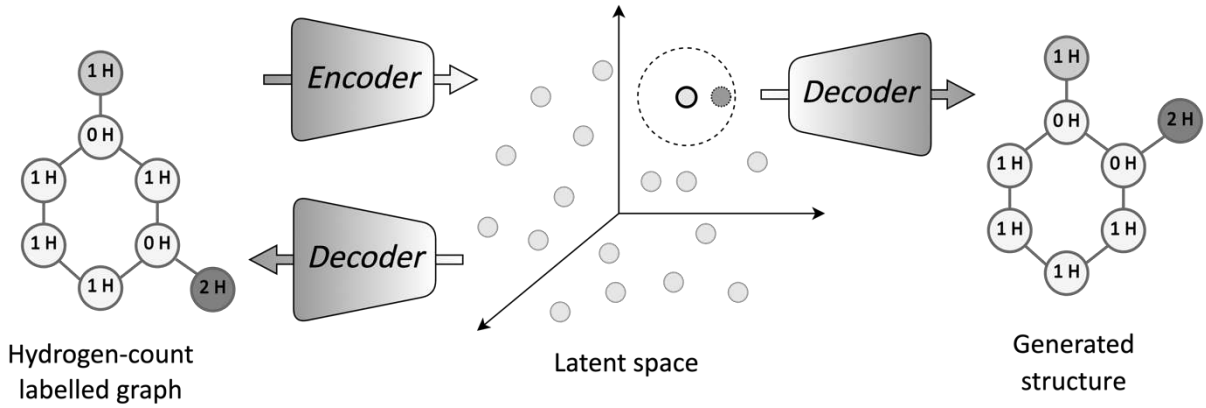


Figure 2. The concept of HyFactor graph-based autoencoder. Here, the hydrogen-labelled graph is translated into a vector representation and back using HyFactor. The new molecular structure is generated by sampling latent vector in the vicinity of the molecule.

Bibliography:

[1] I. I. Baskin, Expert Opin. Drug Discov. (2020), 15, 755–764.
[2] T. Akhmetshin, A. Lin, D. Mazitov, E. Ziaikin, T. Madzhidov, A. Varnek, ChemRxiv (2021), 1–24.
[3] R. Assouel, M. Ahmed, M. H. Segler, A. Saffari, Y. Bengio, arXiv (2018), 1–14.

[P2] Computer-aided design of selective chemical probes of angiotensin-converting enzyme 2

Farah ASGARKHANOVA, Shamkhal BAYBEKOV, Dragos HORVATH, Yuliana Zabolotna, Gilles MARCOU, Alexandre VARNEK*
Laboratoire de Chémoinformatique UMR 7140 CNRS, Institut Le Bel, University of Strasbourg, 4 Rue Blaise Pascal, 67081 Strasbourg, France

Angiotensin-converting enzyme 2 (ACE2) is a key element of the blood pressure regulation mechanism but also turns out to mediate SARS-CoV-2 cellular entry and infection.

The aim of this work is to predict *in silico* selective ACE2 binders. Experimentally confirmed hits may act as chemical probes for SARS-CoV-2 side effects investigation, to distinguish which COVID-19 symptoms may be directly related to perturbed ACE2 activity.

We built a QSAR classification model and performed pharmacophore modelling for ACE2 receptor binding. Models for related targets angiotensin-converting enzyme (ACE) and neprilysin (NEP) were also developed and used to assess ACE2 selectivity of virtual hits. Experimental data extracted from ChEMBL and PubChem, and provided by Enamine (for ACE-2) were used as a training set. In addition, structure-based pharmacophore model has been developed with the LigandScout program. The developed models were used to screen both Enamine in-stock collection (2.6M compounds) and virtual library generated with the Synt-On[1] tool (4080 compounds). Virtual hits predicted by QSAR and pharmacophore models were docked with PLANTES into the ACE2 active site and ranked according to the docking score. A set of 39 potential selective ACE2 binders has been submitted for the experimental validation.

References

[1] Y. Zabolotna *et al.*, ' *J. Chem. Inf. Model.*, 2021,doi: 10.1021/acs.jcim.1c00754.

.

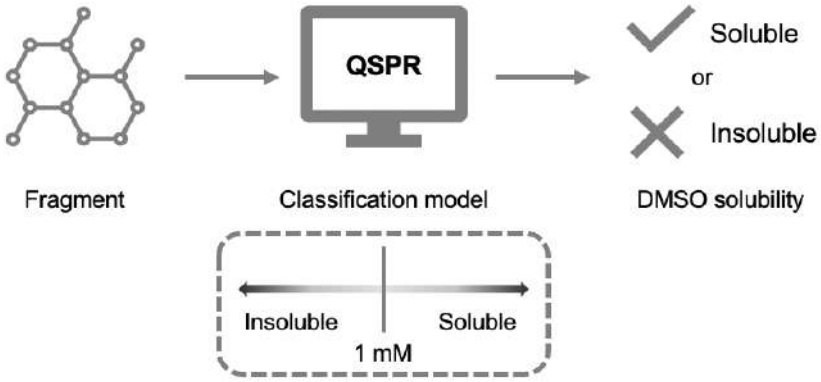
[P3] Prediction of DMSO Solubility for Fragment-Based Screening

Shamkhal BAYBEKOV[1], Gilles MARCOU[1], Pascal RAMOS[2], Olivier SAUREL[2], Jean-Luc GALZI[3,4], Alexandre VARNEK[1]

[1]Laboratoire de Chémoinformatique UMR 7140 CNRS, Institut Le Bel, University of Strasbourg, 4 Rue Blaise Pascal, 67081 Strasbourg, France
[2]Institut de Pharmacologie et de Biologie Structurale, Université de Toulouse CNRS, UPS, 205 Route de Narbonne, 31077 Toulouse, France
[3]Biotechnologie et Signalisation Cellulaire UMR 7242 CNRS, École Supérieure de Biotechnologie de Strasbourg, University of Strasbourg, 300 Boulevard Sébastien Brant, 67412 Illkirch, France
[4]ChemBioFrance – Chimiothèque Nationale UAR3035, 8 Rue de L'école Normale, CEDEX 05, 34296 Montpellier, France

Over the past decades, fragment-based screening (FBS) has been recognized as an efficient alternative to the conventional high-throughput screening (HTS).[1] FBS methods are focused on a relatively small chemical collections composed of small yet diverse organic compounds, called fragment-like compounds.

In this study, we have developed a classification quantitative structure-property relationship (QSPR) model that predicts fragment solubility in dimethyl sulfoxide (DMSO).[2] This solvent is commonly used in screening methods. Therefore, preliminary assessment of solubility of fragments in DMSO is beneficial, saving material and human resources. The categorical threshold for solubility classification was set to 1 mM, which is a common FBS sample concentration. The predictive model is freely accessible on the Predictor web-server of the Laboratory of Chemoinformatics.[3]



Bibliography:

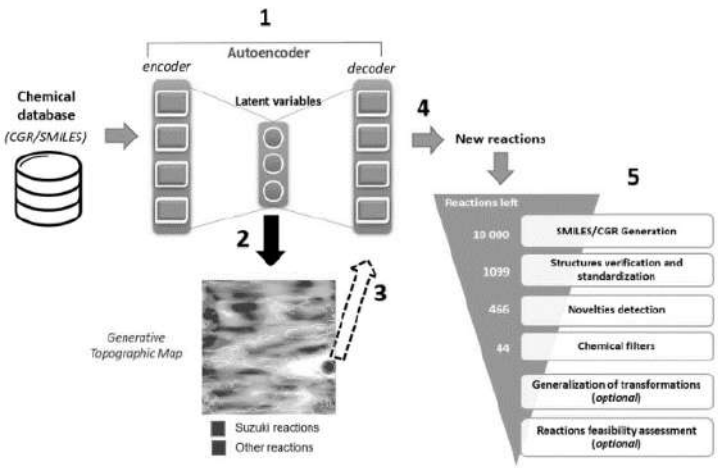
[1] Kirsch, P.; Hartman, A. M.; Hirsch, A. K. H.; Empting, M. Concepts and Core Principles of Fragment-Based Drug Design. Molecules (2019), 24 (23), 4309.
[2] Baybekov, S.; Marcou, G.; Ramos, P.; Saurel, O.; Galzi, J.-L.; Varnek, A. DMSO Solubility Assessment for Fragment-Based Screening. Molecules (2021), 26 (13).
[3] Laboratory of Chemoinformatics - Online tools http://infochim.u-strasbg.fr/cgi-bin/predictor2.cgi (accessed 2022 -04 -21).

[P4] Harnessing the “creativity” of AI to generate novel chemical reactions

William BORT[1], Igor I. BASKIN[1,2,4], Timur GIMADIEV[3], Artem MUKANOV[2], Ramil NUGMANOV[2], Pavel SIDOROV[3], Gilles MARCOU[1], Dragos HORVATH[1], Olga KLIMCHUK[1], Timur MADZHIDOV[2], Alexandre VARNEK[1,3]

[1] *Laboratory of Chemoinformatics, UMR 7140 CNRS, University of Strasbourg, 1, rue Blaise Pascal, 67000 Strasbourg, France.*
[2] *Laboratory of Chemoinformatics and Molecular Modeling, Butlerov Institute of Chemistry, Kazan Federal University, Kremlyovskaya str. 18, 420008 Kazan, Russia.*
[3] *Institute for Chemical Reaction Design and Discovery (WPI-ICReDD), Hokkaido University, Kita 21 Nishi 10, Kita-ku, Sapporo 001-0021, Japan.*
[4] *Department of Materials Science and Engineering, Technion – Israel Institute of Technology, 3200003 Haifa, Israel*

In this study, we investigate the creativity of AI to discover novel types of chemical transformations. For this purpose, a combination of autoencoder with chemical space cartography was used. A classical sequence-to-sequence autoencoder with bidirectional Long Short-Term Memory layers was trained on specially developed “SMILES/CGR” strings, encoding chemical transformations from the USPTO database. Generative Topographic Mapping (GTM) was used to visualize the latent space of chemical reactions. New latent vectors were sampled from the GTM area populated by Suzuki coupling reactions followed by decoding to corresponding reaction equations. An automatic detection of novel transformations was set up by analyzing reaction centers specifically. Reaction feasibility analysis was performed on the basis of reaction heats calculated with the DFT method. Among generated reactions, we identified 12 transformations absent in the training set, 5 of which were then found in recent publications.



Bibliography:

Bort, W., Baskin, I.I., Gimadiev, T. *et al.* Discovery of novel chemical reactions by deep generative recurrent neural network. *Nat. Sci Rep* **11,** 3178 (2021).

[P5] Cheminformatic Analysis of Ring Systems in Natural Products

Ya CHEN[1], Cara ROSENKRANZ[2], Steffen HIRTE[1,3], Johannes KIRCHMAIR[1]

[1]*Department of Pharmaceutical Sciences, Division of Pharmaceutical Chemistry, Faculty of Life Sciences, University of Vienna, 1090 Vienna, Austria*
[2]*Center for Bioinformatics (ZBH), Universität Hamburg, 20146 Hamburg, Germany*
[3]*Vienna Doctoral School of Pharmaceutical, Nutritional and Sport Sciences (PhaNuSpo), University of Vienna, 1090 Vienna, Austria*

More than half of all modern small-molecule drugs are related, to some extent, to natural products (NPs) [1]. Much of the significance of NPs can be attributed to their ring systems, which form the structural core of many drugs. However, despite the importance of NP ring systems, the understanding of their structural properties and how the full potential of NP ring systems can be harnessed in drug discovery are still limited.

This contribution will present a comprehensive cheminformatic analysis of more than 35,000 NP ring systems with regard to their structural and physicochemical properties, and compare them with those of ring systems found in readily purchasable, synthetic compounds and approved drugs. The data sets were carefully curated to obtain clean collections of NPs and synthetic compounds. In addition to key 2D physicochemical properties such as molecular weight and log*P*, 3D shape and electrostatic properties were explored.

Stereochemistry is important, in particular to NP research, as it contributes substantially to the structural complexity and biological activities of compounds. However, stereochemical information is often incomplete and sometimes even wrong [2]. Therefore, most cheminformatics studies disregard stereochemical information. To maximize the usage of the available chemical information we elaborated an evidence-based approach to utilize the stereochemical information whenever it adds value and the data situation permits.

This study shows that approximately one in two NP ring systems are represented by ring systems found in synthetic compounds that have identical or related 3D shape and electrostatic properties. Meanwhile, only about 2% of the NP ring systems are observed in approved drugs, leaving a large number of ring systems to be explored in small-molecule drug discovery.

Bibliography:

[1] Newman, D. J., Cragg, G. M. J. Nat. Prod. 83 (2020) 770–803.
[2] Kramer, C., Podewitz, M., Ertl, P., Liedl, K. R. Planta Med. 81 (2015) 459–466.

[P6] A New Machine Learning Based Method for ADRB2 Agonist Detection Using Single-Ligand Dynamic Interaction Data

Luca CHIESA, Esther KELLENBERGER

UMR 7200 – Centre National de la Recherche Scientifique et Université de Strasbourg, Institut du Médicament de Strasbourg, Faculté de Pharmacie, 67400 Illkirch, France.

The G-protein coupled receptor family is responsible for signaling transduction in many biological processes. The binding of a ligand regulates the signaling by stimulating it (agonist) or inhibiting it (antagonist, inverse agonist). The β2 adrenergic receptor (ADRB2) is one of the most studied GPCR, with many known ligands with an agonistic or antagonistic action.

The ligand binding information provided by crystallographic structures of ADRB2 is often used to improve virtual screening performance, by allowing better separation not only of active and inactive ligands, but also of agonists and antagonists [1][2]. Here, we propose a method that takes into account the conformational dynamics of the ADRB2/ligand reference complex with the aim of improving the biased search towards ligands with specific pharmacological properties.

An ensemble of binding poses was obtained from the crystal structure of ADRB2-agonist complex using molecular dynamics (MD) simulations [3][4]. Key interaction-patterns for agonist activity were selected by a machine learning algorithm. As a test, the developed model was used, in combination with protein-ligand docking, to screen a small library containing well-characterized agonists and antagonists targeting ADRB2.

The proposed technique could be used to post-process docking poses to determine if they can be considered as agonist-like. It can be applied as a filter to remove non relevant poses, non-active ligands, and ligands with an undesired pharmacological effect.

Bibliography:

[1] J.Desaphy; E.Raimbaud; P.Ducrot; D.Rognan J. Chem. Inf. Model. 53 (2013) 623-637
[2] A.J.Kooistra; R.Leurs ; I.J.P.De Esch ; C.De Graaf J. Chem. Inf. Model. 55 (2015) 1045-1061.
[3] I.Rodríguez-Espigares;M.Torrens-Fontanals ;J.K.S.Tiemann ;et al *Nat Methods*. 17(2020) 777-787.
[4] R.O.Dror; et al. Proc. Natl. Acad. Sci. U. S. A. 108 (2011) 18684– 18689

[P7] A Machine learning model of CYP3A4 ligand selectivity

Lydia BENKAIDALI[1]*, François ANDRE[2], Michel PETITJEAN[3], Anne-Claude CAMPROUX[3].

[1]*ICOA, UMR 7311, CNRS, Orléans University, France.*
[2]*I2BC/LSOD, CNRS, CEA, Université Paris-Saclay, France.*
[3]*CMPLI, INSERM ERL U 133, BFA, CNRS UMR 8251, Université de Paris, France.*
[*]*Corresponding author (lydia.benkaidali@univ-orleans.fr)*

Cytochromes P450 (CYPs) constitute a large family of ubiquitous hemoproteins involved in drug metabolism in mammals and many biosynthesis pathways in all living organisms where monoxygenase activity is required. CYP3A4 isoform is the main human metabolizer in liver, and as a multidrug enzyme, it remarkably recognizes a broad spectrum of molecules sharing no common molecular motif. This multispecificity is not yet understood, although it seems clear that ligand recognition occurs not only in the active site, but also during its route through specific ingress channels. A channel is defined as a ligand accessible pathway leading from the protein surface to the buried heminic cavity. In a previous study we gave evidence that CYP3A4 exhibits 4 major drug access channels with different structural features. In this work, we investigated the relationships between the nature of CYP3A4 ligands and the potential pathway they follow to reach the active site.

CYP3A4 presents to date 77 crystallographic structures released in PDB, apo and complexed with various ligands. This remarkable dataset allows comparing the various conformational states of the ligand-bound enzyme, and the binding mode of 71 drugs. By comprehensive scrutiny of the co-crystal structures, we found that the enzyme exhibits 3 main conformations [2] allowing the opening of 4 main channels [3] identified by geometrical calculation [1]. In our interpretation, channel plasticity allows the geometric (shape and size) and electronic (physico-chemical) selection for different classes of drugs. In this mechanism, the ligand-channel interaction leads to a specific deformation of the CYP3A4 structure associated to the position of key loop F-G, and drives the selection of one of the three conformational states for admission of the substrate. A first diagram predicting these access openings was proposed [2] that proved to be consistent with all known crystal structures in 2017.

In a subsequent study we used these data to perform a machine learning validation of the structural study, in order to propose a predictive model of CYP3A4 multi-drug recognition. We took advantage of about twenty new crystallographic structures of the protein to improve our model enriched with the new set of ligands. We built a Classification and regression Tree model, using cross-validation, to predict the 3A4 channeling selectivity and understand how the 3A4 opens specific channels to given ligands. We aim to propose a relevant 3A4 predictive conformation model which sheds light on the relationship between the ubiquitousity (diversity of substrates) and the plasticity of the channels of an isoform. Preliminary results on new data are promising and should contribute to a better understanding of substrate recognition and anticipation inhibition or metabolism by CYP3A4.

Bibliography:

[1] Benkaidali, L.; André, F.; Maouche, B.; Siregar, P.; Benyettou, M.; Maurel, F.; Petitjean, M. Computing cavities, channels, pores and pockets in proteins from non spherical ligands models. Bioinformatics 2014, 30 [6], 792-800. doi 10.1093/bioinformatics/btt644

[2] Benkaidali L, André F, Moroy G, Tangour B, Maurel F, Petitjean M. The Cytochrome P450 3A4 has three Major Conformations: New Clues to Drug Recognition by this Promiscuous Enzyme. Mol Inform. 2017 Oct;36(10). doi: 10.1002/minf.201700044.

[3] Benkaidali L, André F, Moroy G, Tangour B, Maurel F, Petitjean M. Four Major Channels Detected in the Cytochrome P450 3A4: A Step toward Understanding Its Multispecificity. Int J Mol Sci. 2019;20(4):987. doi: 10.3390/ijms20040987.

[P8] Drugmapper: a Web Resource to Explore Active Pharmaceutical Ingredients (Apis)

Ashenafi LEGEHAR[1], Akseli NIEMELÄ[1], Henri XHAARD[1], and Leo GHEMTIO[1,2]

[1]Drug Research Program, Division of Pharmaceutical Chemistry and Technology, Faculty of Pharmacy, University of Helsinki, P.O. Box 56 (Viikinkaari 5 E), FI-00014 University of Helsinki, Finland.
[2]Department of Biosciences, Division of Biochemistry, University of Helsinki, P.O. Box 56 (Viikinkaari 9), FI-00014 University of Helsinki, Finland.

Approximately 90% of the drugs fail during clinical studies, which makes the drug development process complex, expensive, and time-consuming [1]. The main reason for the majority of drug failures is a lack of efficacy and toxicity[2], [3]. Even though the effort to improve the predicting the performance of safety and toxicity for the drug candidate is increased over time using preclinical and post-clinical data[3], [4], drug approval become more costly and challenging[5]. Therefore, high-quality clinical and approved drug data, which are integrated with a knowledge graph, are required to help the effort in predicting potential off-targets, primary and secondary targets combing with therapeutic information for active pharmaceutical ingredients (APIs) to improve their profiles.

Here, we present DrugMapper, the successor of IDAAPM[6], which is released in 2016. IDAAPM focused on integrating FDA-approved drugs and their available information. Similarly, DrugMapper is a web-based platform and drug discovery database that provides clinical candidates and approved drugs and integrates detailed information such as clinical phase studies, FDA-approved application data, mechanism of action, therapeutic indication, structure, molecular descriptors, biological activities, targets, human metabolites, ADMET, and adverse effects. Further, this resource is embedded with interactive knowledge graphs to navigate chemically similar APIs and their pharmacogenomics and pharmacology data. This provides a great advantage to understanding pharmaceutical attrition on APIs during the clinical phase, a new mechanism of action for approved drugs, new treatment, and identifying off-targets in drug discovery. In addition, it allows users better understand the evolution of approved drugs using the chemical scaffold associated with the chemical space in DrugMapper, predicting potential primary and secondary targets, and further building a predictive model for a new target for APIs, ADMET, and adverse effects. DrugMapper is integrated with KNIME[7], free and open-source data analytics, and an integration platform.

Bibliography:

[1] D. Sun, W. Gao, H. Hu, and S. Zhou, “Why 90% of clinical drug development fails and how to improve it?,” Acta Pharm. Sin. B, Feb. 2022.
[2] R. K. Harrison, “Phase II and phase III failures: 2013-2015,” Nat. Rev. Drug Discov., vol. 15, no. 12, pp. 817–818, Nov. 2016.
[3] S. E. Lazic and D. P. Williams, “Improving drug safety predictions by reducing poor analytical practices:,” https://doi.org/10.1177/2397847320978633, vol. 4, p. 239784732097863, Dec. 2020.
[4] E. Semenova, D. P. Williams, A. M. Afzal, and S. E. Lazic, “A Bayesian neural network for toxicity prediction,” Comput. Toxicol., vol. 16, Nov. 2020.
[5] S. Simoens and I. Huys, “R&D Costs of New Medicines: A Landscape Analysis.,” Front. Med., vol. 8, p. 760762, Oct. 2021.
[6] A. Legehar, H. Xhaard, and L. Ghemtio, “IDAAPM: integrated database of ADMET and adverse effects of predictive modeling based on FDA approved drug data,” J. Cheminform., vol. 8, no. 1, p. 33, 2016.
[7] M. Berthold, N. Cebron, and F. Dill, “KNIME: The Konstanz information miner,” Data Anal. Mach. …, 2008.

[P9] Protein Subpocket Cloud Comparison Revealed Similarity Between HIV-1 Reverse Transcriptase and Tumor Necrosis Factor Binding Sites

Merveille EGUIDA, Didier ROGNAN

Laboratoire d'Innovation Thérapeutique, CNRS-Université de Strasbourg, 74 route du Rhin, 67400, Illkirch-Graffenstaden, France.

The increasing number of druggable pockets in protein structures enables structure-based drug design via pocket similarity assessment. Identifying pocket similarity between unrelated targets across the proteome is valuable to drug design [1] but still is a challenge to binding site comparison methods, notably local similarities arising from cavity microenvironment.

We therefore developed ProCare [2], a novel computational method to compare protein pockets using a 3D point cloud registration algorithm. In computer vision, point cloud registration is a fundamental problem of finding the best transformation (rotation, translation, scaling) to match two clouds of points. A protein pocket is here represented as an ensemble of 3D points annotated by atomic coordinates and microenvironment specific pharmacophoric properties. Following the characterization of each point with a hybrid shape-chemical descriptor (c-FPFH) [2-3], two pockets are aligned by superimposing their corresponding points sharing the most similar patterns. The alignments are evaluated by estimating the proportion of aligned points sharing the same pharmacophoric properties.

Out of a large-scale comparison where subpockets from the sc-PDB database were compared to the Tumor Necrosis Factor alpha (TNF-α) [4], ProCare suggested a similarity with the non-nucleoside binding site of HIV-1 reverse transcriptase (HIVRT) [2, 5]. Extensive binding site comparisons using different structures of TNF-α [4] and HIVRT reinforced this similarity hypothesis, which was later confirmed by microscale thermophoresis assay where two out of three tested HIVRT approved drugs were found to bind to TNF-α (KD in the 20-40 µM range). Remarkably, other state of the art binding site comparison methods as well as ligand 2D fingerprints and 3D shape methods were not able to detect that similarity [5].

ProCare allows local comparison of protein pockets of different sizes while yielding visually interpretable results. The method is ideally suited to identify local and unobvious similarities among totally unrelated targets, and appears as a promising idea generator for fragment-based ligand design, able to pick starting fragments at a proteomic scale, not necessarily influenced by existing ligand or cavity neighborhoods.

Bibliography:

[1] Ehrt, C.; Brinkjost, T.; Koch, O. J. Med. Chem. (2016), 59, 4121–4151.
[2] Eguida, M.; Rognan, D. J. Med. Chem. (2020), 63, 7127–7142.
[3] Rusu, R. B.; Blodow, N.; Beetz, M. (2009), IEEE Int. Conf. Robot. Autom. (2009), 3212–3217.
[4] O'Connell, J. et al. Nat. Commun. (2019), 10, 5795.
[5] Eguida, M.; Rognan, D. J. Cheminform. (2021), 13, 1–13.

[P10] PLATO: a User-Friendly Web Platform for Target Fishing and Bioactivity Prediction

Nicola GAMBACORTA[1], Fulvio CIRIACO[2], Daniela TRISCIUZZI [1] and Orazio NICOLOTTI[1]

[1]Dipartimento di Farmacia Scienze del Farmaco, Università degli Studi di Bari "Aldo Moro", Via E. Orabona, 4, I-70125 Bari, Italy
[2]Dipartimento di Chimica, Università degli Studi di Bari "Aldo Moro", Via E. Orabona, 4, I-70125 Bari, Italy;

The web-platform PLATO (standing for Polypharmacology pLATform for predictiOn), is a ligand-based polypharmacology predictive platform, which has been designed with a two fold objective: to shortlist a number of putative protein drug targets and to compute the bioactivity affinity values. PLATO employs a pool including 632,119 druglike ligands and 6004 targets provided with experimental annotations retrieved from the latest update of ChEMBL (release 30) [1] according to transparent filtering rules elsewhere described and implements two just optimized multifingerprint similarity-based algorithms, which have been recently published, [2, 3] in a python web framework accessible through a graphical user-friendly interface available online at the following link: http://plato.uniba.it/

Users can interrogate PLATO by simply drawing the chemical structure of a given query or, alternatively, by pasting its SMILES notation. Two different screening options are thus given. The first is aimed at searching for putative drug targets based on molecular similarity. The second allows making quantitative predictions of bioactivity based on a statistical approach. In a few seconds, PLATO returns a standard report in a portable document format, which includes the list of the top-scored 30 solutions as well as a wealth of additional information for each single result regarding the ligand chemical structure, the protein drug target and the bioactivity values. The standard report includes hyperlinks to redirect users to ChEMBL for further and deeper investigations.
Please note that all the information gathered by PLATO are stored in a downloadable .json file.

Bibliography:

[1] Mendez,D., Gaulton,A., Bento,A.P., Chambers,J., De Veij,M., Félix,E., Magariños,M.P., Mosquera,J.F., Mutowo,P., Nowotka,M., et al. (2019) ChEMBL: towards direct deposition of bioassay data. Nucleic Acids Res, 47, D930–D940.
[2] Alberga,D., Trisciuzzi,D., Montaruli,M., Leonetti,F., Mangiatordi,G.F. and Nicolotti,O. (2019) A New Approach for Drug Target and Bioactivity Prediction: The Multifingerprint Similarity Search Algorithm (MuSSeL). J Chem Inf Model, 59, 586–596.
[3] Ciriaco,F., Gambacorta,N., Alberga,D. and Nicolotti,O. (2021) Quantitative Polypharmacology Profiling Based on a Multifingerprint Similarity Predictive Approach. J. Chem. Inf. Model., 61, 4868–4876.

[P11] Application of Automated Machine Learning in Search for Multi-Target-Directed Ligands Blocking PDE4B, PDE8A and TRPA1 Ion Channel with Potential Use in the Treatment of Asthma and COPD

Alicja GAWALSKA[1], Aleksander MENDYK [2], Adam BUCKI[1], Marcin KOŁACZKOWSKI[1]

[1]Department of Medicinal Chemistry, Jagiellonian University Medical College, ul. Medyczna 9, 30-688 Kraków, Poland
[2]Department of Pharmaceutical Technology and Biopharmaceutics, Jagiellonian University Medical College, ul. Medyczna 9, 30-688 Kraków, Poland

Asthma and COPD are characterized by complex pathophysiology associated with chronic inflammation, bronchoconstriction, and bronchial hyperresponsiveness resulting in airway remodeling. The currently available therapeutic strategies do not address all of the most important pathological processes in the course of both diseases. Therefore, it is an urgent need to work out comprehensive solutions that fully affect the pathological processes of both diseases. As a possible solution for the enhanced treatment of asthma and COPD, we propose the rationally designed multi-target-directed ligands (MTDLs), combining PDE4B and PDE8A inhibition with TRPA1 ion channel blockade. This approach allows forobtaining synergistic bronchodilatory, anti-inflammatory, and additional anti-remodeling activity.

The aim of the study is to develop AutoML models to search for MTDL blocking PDE4B, PDE8A and TRPA1, which will allow the selection of novel MTDL chemotypes.

Using “mljar-supervised” - Automated Machine Learning Python package [1], there were regression models developed for each of the biological targets (PDE4B, PDE8A, TRPA1). For this purpose, libraries of inhibitors derived from CHEMBL database and collected from scientific articles were used. Each inhibitor was represented by a set of calculated molecular descriptors (PADEL software) and IC_{50} values. Ensemble systems combining various modeling tools like artificial neural networks, decision trees and i.e. XGBoost were found as the best models for each of the biological targets. On their basis, virtual screenings of commercially available compounds derived from the ZINC15 database, were performed. A common group of compounds placed within the top results were selected as potential novel chemotypes of multifunctional ligands.

Further studies, carried out after the purchase of selected compounds, will be focused on in vitro activity tests providing reliable results to aid the discovery of MTDLs blocking PDE4B, PDE8A and TRPA1.

The study was financially supported by the National Science Centre, Poland (grant no. 2020/37/N/NZ7/02365). Calculations were performed partially with use of computers co-financed by the qLIFE Priority Research Area under the program “Excellence Initiative Research University” at Jagiellonian University and Polish Operating Programme for Intelligent Development POIR4.2 project no. POIR.04.02.00-00-D023/20.

Bibliography:

[1] Aleksandra Płońska and Piotr Płoński ; MLJAR: State-of-the-art Automated Machine Learning Framework for Tabular Data. Version 0.10.3 ; 2021; https://github.com/mljar/mljar-supervised

[P12] Deciphering a Pharmacophore Network Generated from BCR-ABL Data

Damien GESLIN[1], Alban LEPAILLEUR[1], Jean-Luc MANGUIN[2], Vinh Vo NHAT[1], Jean-Luc LAMOTTE[1] , Bertrand CUISSART [2], Ronan BUREAU[1]

[1]Centre d'Etudes et de Recherche sur le Médicament de Normandie, Normandie Univ, UNICAEN, CERMN, 14000 Caen, France
[2]Groupe de Recherche en Informatique, Image, Automatique et Instrumentation de Caen, Normandie Univ, UNICAEN, ENSICAEN, CNRS, GREYC, 1400 Caen France

In this poster, we present a method used to create clusters of pharmacophores in order to support their detailed analysis.[1] First, from a BCR-ABL molecular dataset, pharmacophores were automatically extracted and organized into a network. The network was spatialized by computing the graph edit distances between the pharmacophores as a similarity measure. The application of a force-directed layout algorithm allowed us to discriminate pharmacophores associated with active molecules from those associated with inactive molecules. Second, a clustering approach was used to refine the partitioning by grouping pharmacophores according to their structures, activities and binding modes. The analysis of this pharmacophore network provided us key information on structure-activity relationships, including the distinction between activity classes and chemical families.

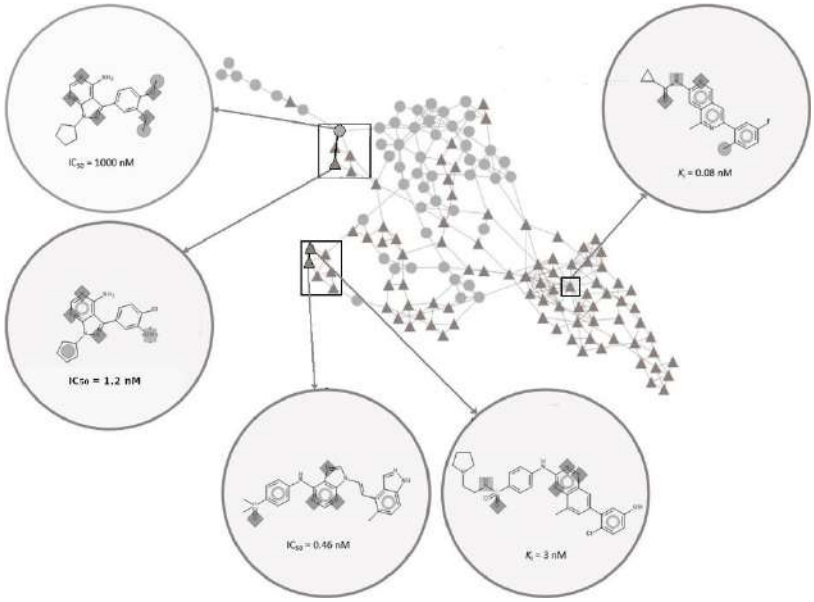


Figure 1: Representation of a BCR-ABL pharmacophore network with some associated molecules (Triangle: active pharmacophore; Circle: inactive pharmacophore).

Bibliography:

[1] Geslin D, Lepailleur A, Manguin JL, et al. Deciphering a Pharmacophore Network: A Case Study Using BCR-ABL Data. J Chem Inf Model. (2022); 62(3):678-691. doi:10.1021/acs.jcim.1c00427

[P13] Towards DAG-Based Interactive Pharmacophore Exploration

Etienne LEHEMBRE[1], Johanna GIOVANNINI[2], Jean-Luc LAMOTTE[2], Abdelkader OUALI[1], Alban LEPAILLEUR[2], Albrecht ZIMMERMANN[1], Bertrand CUISSART[1], Ronan BUREAU[2], Bruno CREMILLLEUX[1]

[1]GREYC, Normandie Univ., UNICAEN, CNRS – UMR 6072, 14000 Caen, France
[2]CERMN, Normandie Univ, UNICAEN, CERMN, 14000 Caen, France

In medicinal chemistry, a pharmacophore denominates a spatial arrangement of chemical features which is responsible for a favorable interaction with the binding site of the target. We have recently designed a method that automatically computes pharmacophores from a large data set of molecules without any prior supervised selection of a small subset of molecules [1, 2]. The connections between the computed pharmacophores provide a hierarchical organization: the pharmacophore network. The latter is structured by the pharmacophores' size and contains a large number of them. The current work aims at developing a structure which a medicinal chemist can use to support his analysis without having to repeatedly mine pharmacophores. For this purpose, we enhance the knowledge on the pharmacophore network organization by taking into account parent-children relations and/or grouping the pharmacophores into equivalence classes, i.e. sets of pharmacophores occurring in the exact same molecule group. Additionally, each pharmacophore is annotated with one or several quality measures (e.g. confidence or growth rate measurements), information that will be exploited later-on.

The directed acyclic graph (DAG) is built and organized in layers which contain pharmacophores with the same number of pharmacophoric features, as well as the molecules that respond to them. Layers are linked by establishing parent-children relationships based on a sub-graph relationship. The DAG can then be used to apply different filtering and selection algorithms, e.g. based on user-specified quality measures, the visualization of the effect on each layer. In order not to overwhelm the expert, the visualization is condensed by clustering nodes in each layer using the above-mentioned equivalence classes. We will also use the DAG to identify Pharmacophore Activity Delta (PAD). PADs are syntactic pharmacophores families, i.e. pharmacophores linked via the afore-mentioned parent-child relationships exhibiting a minimum amount of syntactic similarity, yet members of which show a significant difference between the values of their quality measures.

Interactive Pattern Mining [3] is one of the objective of this work related to ANR project InvolvD (ANR-20-CE23-0023, https://involvd.greyc.fr/). This DAG was used to simulate the effects of users or functions interacting with the search space during the search itself.
The DAG process was applied to BCR-ABL chemogenomic dataset (ChEMBL1862, ChEMBL_24). The BCR-ABL tyrosine kinase is an oncogene associated to chronic myeloid leukemia. We aim to expose our method and how far the PAD analysis highlighted original chemical structural features linked to active or inactive ligands.

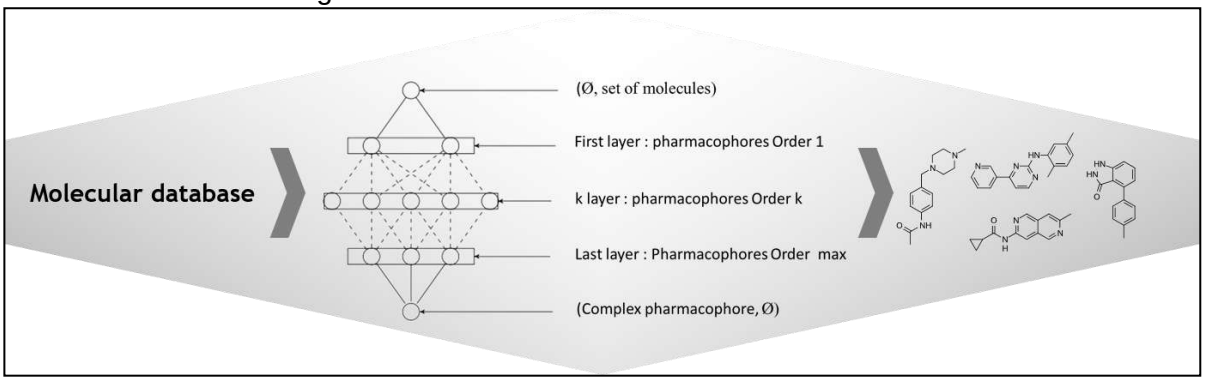


Figure: From a chemical database to the highlight of structural features of interest, we aim to present how our pharmacophore DAG can be used for fast 2D structural discrimination.

Bibliography:
[1] Métivier *et al*., J. Med. Chem. (2018), 61.
[2] Geslin *et al*., J. Chem. Inf. Model. (2022), 62.
[3] Van Leeuwen, Interactive knowledge discovery and data mining in biomedical informatics. Springer, (2014).

[P14] Dynamic Sensitivity Analysis to Predict Time-Course Drug Transcriptomic Responses of the Cellular System

Michio IWATA and Yoshihiro YAMANISHI

Department of Bioscience and Bioinformatics, Faculty of Computer Science and Systems Engineering, Kyushu Institute of Technology, 680-4 Kawazu, Iizuka, Fukuoka 820-8502, Japan.

Introduction:
Identifying the dynamic human cell line response to drug therapies is necessary to determine the time-course mode of drug action in medical and pharmaceutical research. However, the effect of drug perturbations on the cellular system is not well time-dependently investigated; thus, there is an incomplete picture of their mode of action. In this study, we developed a novel computational method to predict time-course drug-induced transcriptomic responses of the cellular system by integrating simulation- and data-driven approaches.

Materials and methods:
We performed dynamic sensitivity analyses with a mathematical model of the tricarboxylic acid cycle that was constructed in a previous study [1] and constructed drug-induced enzymatic sensitivity signatures as follows:

$$\boldsymbol{P}_k(t) = \left(P_{1,k}(t), P_{2,k}(t), \ldots, P_{l,k}(t)\right)^{\mathrm{T}}, (k = n+1, \ldots, n+m),$$

where $P_{j,k}(t)(j = 1, \ldots, l; k = n+1, \ldots, n+m)$ is the sensitivity for each enzyme at time *t*, *n* is the number of metabolites, *m* is the number of enzymes, and *l* is the number of enzyme genes. The signature represents the sensitivity of *l* enzyme genes to the perturbation of enzyme *k* at time *t*. Then, we compared enzymatic sensitivity signatures to drug-induced gene expression signatures [2] and predict the time course of drug responses for all genes [3]. The prediction is performed based on the correlation between the two types of signatures through common time points.

Results and discussion:
Figure 1 shows an example of observed and a time course of predicted drug-induced gene expression signatures in the CJM cell line in response to enasidenib, an anticancer drug for acute myeloid leukemia. The predicted time-course drug-induced gene expression data enabled the detection of various biological pathways affected by a drug treatment in a time-dependent manner. For example, it was identified that the translation process could be inactivated in the early stages of medication, which cannot be achieved through previous methods based on static gene expression data. Therefore, the proposed method can be used to increase our understanding of drug-induced transcriptomic responses over time.

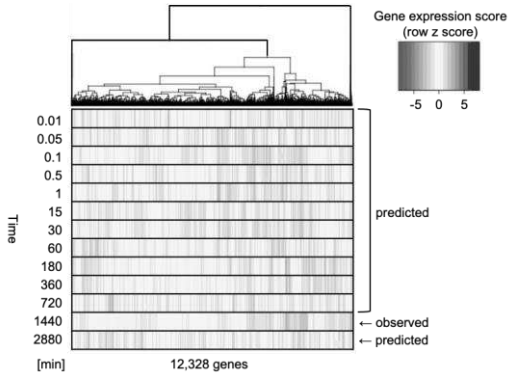


Figure 1 : Observed and predicted drug-induced gene expression signatures at various time points. The dendrogram above the heatmap shows the similarity among genes.

Bibliography:

[1] Shiraishi,F. and Savageau,M.A. J. Biol. Chem. 268 (1993) 16917–16928.
[2] Subramanian,A. et al. Cell 171 (2017) 1437–1452.
[3] Acar,E. et al. Chemometr. Intell. Lab. Syst. 106 (2011) 41–56.

[P15] Computer-Aided Search for New Anthracycline Antibiotic Reductases Inhibitors with a Potential to Support Anticancer Therapy

Marek JAMROZIK[1], Adam BUCKI[1], Kamil PISKA[2], Paulina KOCZURKIEWICZ-ADAMCZYK[2], Benedykt WŁADYKA[3], Elżbieta PĘKALA[2], Marcin KOŁACZKOWSKI[1]

[1]Department of Medicinal Chemistry, Jagiellonian University Medical College, Medyczna 9 st, Kraków, Poland,
[2]Department of Pharmaceutical Biochemistry, Jagiellonian University Medical College, Medyczna 9 st, Kraków, Poland
[3]Department of Analytical Biochemistry, Faculty of Biochemistry, Biophysics and Biotechnology, Jagiellonian University, Gronostajowa 7 st, Kraków, Poland

Anthracycline antibiotics (ANT) are among the most widely used group of anticancer drugs - both in solid tumors and hematological malignancies. Unfortunately, their usage is limited due to the drug resistance and cardiotoxicity, which can manifest not only during the therapy, but even many years after the treatment. It is postulated that those effects are caused by ANT metabolism (the two-electron reduction of a carbonyl moiety), performed mainly by carbonyl reductase 1 (CBR1) and aldo-keto reductase 1C3 (AKR1C3). Thus, CBR1 and/or AKR1C3 inhibitors appear to become a potential support for ANT pharmacotherapy [1, 2].

The aim of the study was to implement computational methods in optimization of CBR1 and AKR1 crystal structures to obtain high-quality models, which then were used to perform virtual screenings, leading to a selection of potential dual CBR1-AKR1C3 inhibitors – a group of substances with a unique inhibitory effect against both crucial ANT reductases.

20 CBR1 ligands and 13 AKR1C3 ligands were used in Induced-Fit docking procedure, leading to the initial models of the enzymes with optimized conformations of amino acid residues of the catalytic sites. The models were then evaluated in retrospective virtual screening and molecular dynamics simulations. Those steps resulted in the selection of one model for both CBR1 (model 11E3: $BEDROC_{\alpha=20}$: 0.698; $EF_{1\%}$: 36) and AKR1C3 (model 74H1: $BEDROC_{\alpha=20}$: 0.643; $EF_{1\%}$: 22), which then were used in prospective virtual screenings. Over 1 million compounds taken from ZINC database have been docked into the prepared models of CBR1 and AKR1C3 to find such structures, that would present the correct binding mode within the catalytic sites of both analysed molecular targets (based on visual examination and assessment of 'glide gscore' scoring function values). The most promising compounds were then evaluated in *in vitro* assessment performed on recombinant CBR1 and AKR1C3 enzymes.

The adopted methods of structure-based drug design led to the selection of initial compounds with moderate inhibitory activity against both crucial ANT reductases: CBR1 and AKR1C3. It is planned to use computational methods for further optimizations and eventually to determine the structure-activity relationships within ANT-reductases inhibitors.

Bibliography:

[1] Minotti G. et al.: Chem Res Toxicol. (2000) Dec;13(12):1336-41,
[2] Olson RD. et al.: Proc Natl Acad Sci U S A. (1988) May;85(10):3585-9,
[3] Wingert BM. et al.: Curr Opin Chem Biol. (2018) Jun;44:87-92.

Acknowledgements: This study was funded by National Science Centre's grant 2020/37/N/NZ7/01097.

[P16] Molecular Modelling and Machine Learning Techniques in Search for Novel SARS-Cov-2 Main Protease Inhibitors.

Jakub JOŃCZYK[1], Magdalena HAWRYLAK[1], Natalia JASKÓLSKA[1], Justyna KOZYRA[1], Agnieszka ZAGÓRSKA[1], Sarah d'ALESSANDRO[2]

[1]Faculty of Pharmacy, Chair of Medicinal Chemistry, Jagiellonian University Medical College, Medyczna Street 9, 30-688 Kraków, Poland
[2]Dipartimento di Scienze Farmacologiche e Biomolecolari, Università degli Studi di Milano, Via Pascal, 36, 20133 Milano

Introduction: Among the many challenges in the field of medicine, viral diseases have always been a great challenge for the world. A sudden emergence of a new stream of coronaviruses and a rapid spread of SARS-CoV-2 around the world shows how we stand in need of new, fast methods for discovering compounds with potential antiviral activity. Use of computational methods in search for patterns that can indicate compounds with desired activity can successfully accelerate future research. In our studies, we wanted to integrate molecular modeling and machine learning techniques into an effective algorithm that can indicate the activity against SARS-CoV-2 main protease during virtual screening.
Aim: The study aimed to develop and apply virtual screening protocol that will allow us to discover novel inhibitors of SARS-CoV-2 Mpro.
Material and methods: Models were built on a group of 8702 ligands with known activity expressed as a percentage of SARS-Cov-2 Mpro inhibition derived from the ChEMBL database [1]. We prepared three separate classification models assigning compounds as active (percent inhibition> 80%) and inactive. The first one was based on molecular descriptors generated by Padel software [2] and ML classification methods. The second model was based on pharmacophores prepared with the LigandScout software. The last model was prepared by molecular docking with the Glide from Maestro - Schrödinger. The best models were integrated into virtual screening protocol.
Results: We used the ROC curves to evaluate the prognostic ability of the obtained models. The best models had high AUC values of 0.81. This indicates a significant advantage of true positive predictions against the false positive ones.
Conclusions: Taking into account the high diversity in the structure of the compounds used to build the models and the promising results of their validation, we believe that the proposed virtual screening protocol can be successfully applied in the search for new SARS-Cov-2 Mpro inhibitors.
Acknowledgments: The project is co-financed by the Polish National Agency for Academic Exchange, grant PPN/BIT/2021/1/00056/U/00001 and the Jagiellonian University Medical College grant N42/DBS/000217

Bibliography:

[1] Maria Kuzikov et al. Identification of inhibitors of SARS-Cov2 M-Pro enzymatic activity using a small molecule repurposing screen. DOI: 10.6019/CHEMBL4495564
[2] Chun Wei Yap. PaDEL-descriptor: An open source software to calculate molecular descriptors and fingerprints. J. Comput. Chem., 32: 1466-1474. https://doi.org/10.1002/jcc.21707

[P17] Scaffold-Retained Structure Generator to Extensively Produce Molecules with Unique Chemical Substructures

Kazuma KAITOH, Yoshihiro YAMANISHI

Department of Bioscience and Bioinformatics, Faculty of Computer Science and Systems Engineering, Kyushu Institute of Technology, 680-4 Kawazu, 820-8502, Iizuka, Fukuoka, Japan.

The construction of a virtual library (VL) consisting of novel compounds based on structure−activity relationships is important for lead optimization in *de novo* drug design. In this study, we develop a novel scaffold-retained structure generator, EMPIRE (Exhaustive Molecular library Production In a scaffold-REtained manner), to produce novel compounds in an arbitrary chemical space.[1] An overview of the proposed method is shown in Figure 1. By combining a deep generative model-based generator and a building block-based generator, our proposed method efficiently provides a VL consisting of compounds that retain the input scaffold and contain unique arbitrary substructures. Our proposed method enables us to construct high quality VLs located in unexplored chemical spaces containing compounds with unique structures such as bicyclo[1.1.1]pentane and cubane or elements such as boron and silicon. We expect EMPIRE to contribute to efficient *de novo* drug design with unique substructures by virtual screening.

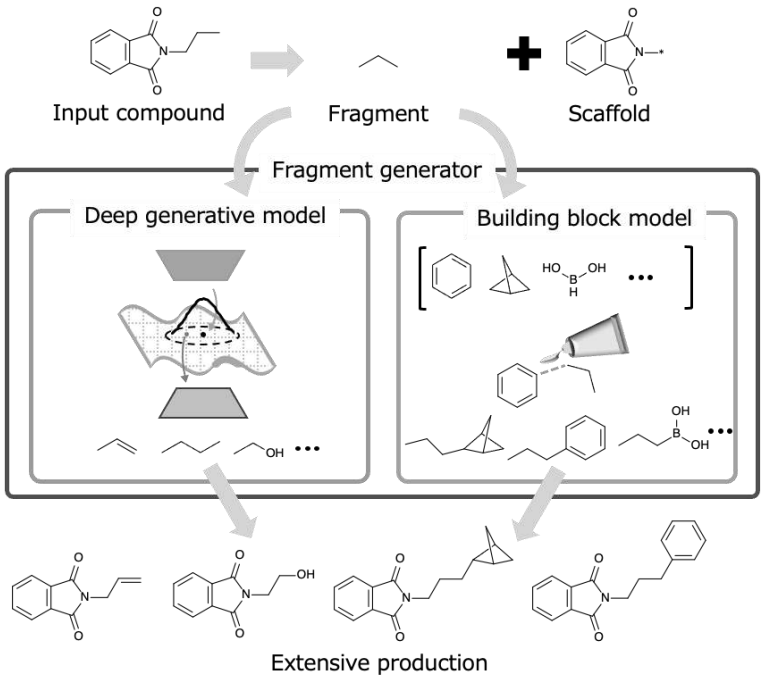


Figure 1. Overview of our proposed structure generator called EMPIRE. The compound is entered as a combination of a scaffold and fragment structures. If the input fragments are none, a methyl group is formally inputted as a fragment. The fragment structures are used in a deep generative and building block models. The fragments produced by each model are inserted into asterisks on the input scaffold, and new compounds are produced.

Bibliography :

[1] K. Kaitoh; Y. Yamanishi. J. Chem. Inf. Model. (2022) https://doi.org/10.1021/acs.jcim.1c01130

[P18] DeLA-Drug: a Deep Learning Algorithm for Automated Design of Drug-like Analogues

Giuseppe LAMANNA[2,3], Teresa Maria CREANZA[1], Pietro DELRE[2,3], Marialessandra CONTINO[4], Nicola CORRIERO[3], Michele SAVIANO[3], Giuseppe Felice MANGIATORDI[3] and Nicola ANCONA[1]

[1] CNR - Institute of Intelligent Industrial Technologies and Systems for Advanced Manufacturing, Via Amendola 122/o, 70126 Bari, Italy
[2] Chemistry Department, University of Bari "Aldo Moro", via E. Orabona, 4, I-70125 Bari, Italy.
[3] CNR – Institute of Crystallography, Via Amendola 122/o, 70126 Bari, Italy
[4] Department of Pharmacy - Pharmaceutical Sciences, University of Bari "Aldo Moro", via E. Orabona, 4, I-70125 Bari, Italy

We present *DeLA-Drug*,[1] a recurrent neural network (RNN) model composed of two Long Short-Term Memory (LSTM) layers and conceived for data-driven generation of drug-like compounds. *DeLA-Drug* captures the syntax of SMILES strings of more than 1 million molecules belonging to the ChEMBL28 database and generates analogues starting from a single user-defined query compound by employing a new strategy called Sampling With Substitutions (SWS). The generative model preserves drug-likeness and synthetic accessibility of the known bioactive compounds belonging to the ChEMBL28 repository. The absence of any time-demanding fine-tuning procedure enables *DeLA-Drug* to perform a fast generation of focused libraries for further high-throughput screening and makes it a suitable tool for performing *de-novo* design even in low-data regimes. *DeLA-Drug*, available as a free web platform (http://www.ba.ic.cnr.it/softwareic/deladrugportal/), can help medicinal chemists interested in generating analogues of compounds already available in their laboratories and, for this reason, good candidates for an easy and low-cost synthesis.

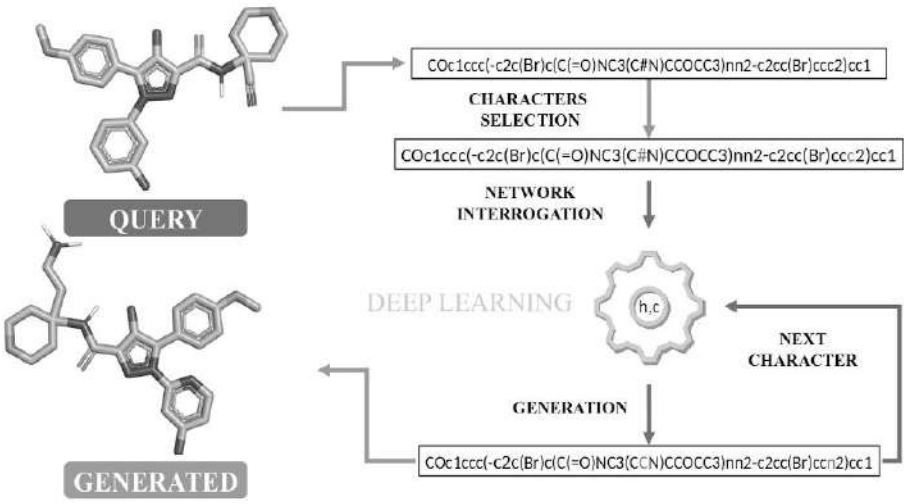


Figure 1. Main steps of the DeLA-Drug workflow.

Bibliography :

[1] Creanza, T. M.; Lamanna, G.; Delre, P.; Contino, M.; Corriero, N.; Saviano, M.; Mangiatordi, G. F.; Ancona, N. (2022), DeLA-Drug: A Deep Learning Algorithm for Automated Design of Druglike Analogues. *J. Chem. Inf. Model.*, *62* (6), 1411–1424. https://doi.org/10.1021/acs.jcim.2c00205.

[P19] Prediction of Compound Plasma Concentration Time Profiles after Oral Administration in Mice Using Random Forests

Koichi HANDA[1,2], Peter WRIGHT[1], Michiharu KAGEYAMA[2], Takeshi IIJIMA[2], Andreas BENDER[1]

[1] Centre for Molecular Informatics, Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge, CB2 1EW, UK
[2] Toxicology & DMPK Research Department, Teijin Institute for Bio-medical Research, Teijin Pharma Limited, 4-3-2 Asahigaoka, Hino-shi, Tokyo 191-8512, Japan

For the setting of optimal dose in clinical stage, PK/PD understanding for both efficacy and safety is imperative, and it needs information of the concentration over time explicitly, like time above MIC. However, pharmacokinetic (PK) parameters such as clearance (CL) and volume of distribution (VD), which are too oversimplified to understand PK/PD relationship, have been the subject of recent *in silico* predictive models [1, 2]. Then, although the intravenous PK data is easy to be dealt, deriving profiles for oral administration (p.o.) is pertinent as 80% of all dosage forms are oral [3]. In this study, despite its increased complexity due to a variety of absorption processes [4], we predicted p.o concentration time profiles by developing novel *in silico* models of plasma concentration at consecutive time points after oral administration.

To this end, we used mouse p.o. PK data obtained for 871 compounds (17 projects) under a standardized protocol (single dose: 2 µmol/kg, 7 time points for blood sampling). For explanatory variables, MACCS Keys as well as *in silico* predicted human VD (hVD), *in-vitro* Caco2 permeability, solubility, mouse intrinsic clearance (mCLint), and unbound fraction of mouse plasma (mfu) were used. The predictive accuracy of Random Forest models (RF), 2-compartment models using estimated CL and VD (CPM), and average models (as a control experiment, AVM) were investigated using 5-fold cross validation (5-fold CV) and leave-one-project-out cross validation (LOPO-CV).
The average predictive accuracy of RF in 5-fold CV was the best among the models studied. The RMSE at 0.25 h, 1 h, and 8 h were RF: 0.500, 0.612, 0.509; CPM: 0.611, 0.756, 0.547, and AVM: 0.758, 0.827, 0.776, respectively. This RF was further validated using LOPO-CV (ranges of predictive accuracy of 8 projects having over 25 compounds): 0.494-0.772 at 0.25 h, 0.523-1.024 at 1 h, and 0.296-1.305 at 8 h. We next investigated the importance of the PK explanatory variables using the GINI index. At 0.25 h, the importance of Caco2 permeability was 2.1 times higher than the next most important variable (mCLint). At 2 to 8 h, the importance of hVD, mCLint, and mfu were at least 1.8 times higher than the next one (Caco2 permeability). Given the good acurracy of the *in silico* model, this suggests that these parameters contribute differently towards the efficient model abstraction of key absorption and distribution processes at different time points after administration, and this is consistent with the prior knowledge of ADME research. Although the predictivity depended on each project, due to its combination of sufficient accuracy and speed of prediction we found the model to be fit-for-purpose for practical lead optimization.

Bibliography:

[1] Wang Y. et. al. In Silico Prediction of Human Intravenous Pharmacokinetic Parameters with Improved Accuracy. J. Chem. Inf. Model. (2019), 59, 9, 3968–3980
[2] Iwata H. et. al. Prediction of Total Drug Clearance in Humans Using Animal Data: Proposal of a Multimodal Learning Method Based on Deep Learning. J Pharm Sci. (2021);110(4):1834-1841.
[3] Morawietz G. et. al. RITA--Registry of Industrial Toxicology Animal-data--progress of the working group. Exp Toxicol Pathol. (1992). 44(6):301-309.
[4] Arthur J. et. al. CHAPTER 13 - PHARMACOKINETICS. Pharmacology and Therapeutics. (2009),193-202.

[P20] Flexible Protein-Ligand Docking Protocol: a Case Study on the Androgen Receptor

Mehdi OUDAHMANE, Matthieu MONTES, Nathalie LAGARDE

Laboratoire Génomique, Bioinformatique et Chimie Moléculaire, EA 7528, Conservatoire National des Arts et Métiers, 292 rue Saint Martin, 75003 Paris, France, Hesam Université

New protein-ligand docking strategies have emerged over time, shifting from approaches considering the protein as completely rigid to approaches taking into account the protein flexibility. In this study, we present a protocol to rationally select a subset of residues which side chains should be considered as flexible during the docking with no requirement of prior computational studies. We then evaluate the impact of this choice on the docking performances.

This study focus on the androgen receptor (AR) [1], which is a member of the nuclear receptor (NR) family. AR binding (B) and non binding (NB) compounds were extracted from the Environmental Protection Authority (EPA) Dataset Gateway [2]. AR structures were extracted from the Protein Data Bank. The AR structures were evaluated using the MolProbity webserver [3] and by inspection of the electronic density maps and 8 structures were selected accordingly. Different docking softwares (Autodock Vina, smina and GNINA) and scoring functions (vina, vinardo and dkoes_scoring) [4,5] were then evaluated using both rigid and flexible protein protocols. The 2 AR structures associated with respectively the best and the worst AUC using the rigid protein docking protocol were then submitted to the flexible protein docking protocol. To do so, we first set as flexible the side chains of all residues within a 4Å cut-off distance from the co-crystallized ligand. In a second time, we studied the docking performance associated with different combination of 1 to 6 residues with flexible side chains. These 6 residues were selected according to the intrinsic flexibility of their side chain, the alternate conformation observed within different AR structures and their position in the binding site.

We demonstrated that taking into account the protein flexibility enable to enhance docking performance. In particular, we highlighted that this can be achieved by selecting only a small number of flexible side chains (a combination of 3 residues). This is a crucial point to generate results for large compounds libraries with reasonable computational times.

Bibliography:

[1] Weikum, Emily R., et al. « The Nuclear Receptor Superfamily: A Structural Perspective: The Nuclear Receptor Superfamily ». Protein Science, vol. 27, no 11, November (2018), p. 1876-92.

[2] Mortensen, Holly M., et al. « The 2021 Update of the EPA's Adverse Outcome Pathway Database ». Scientific Data, vol. 8, no 1, December (2021), p. 169.

[3] Chen, Vincent B., et al.« MolProbity : all-atom structure validation for macromolecular crystallography ». Acta Crystallographica Section D Biological Crystallography, vol. 66, no 1, January (2010), p. 12-21.

[4] Quiroga, Rodrigo, et Marcos A. Villarreal. « Vinardo: A Scoring Function Based on Autodock Vina Improves Scoring, Docking, and Virtual Screening ». PLOS ONE, édité par Heinrich Sticht, vol. 11, no 5, may (2016).

[5] Axerio-Cilies, Peter, et al. « Inhibitors of Androgen Receptor Activation Function-2 (AF2) Site Identified through Virtual Screening ». Journal of Medicinal Chemistry, vol. 54, no 18, September (2011), p. 6197-205.

[P21] Assessment of the Overlap Between the 'On-The-Shelf' Drug-Like Space and Ultra-Large 'On-Demand' Combinatorial Libraries

Mariana PEREBYINIS and Didier ROGNAN

Laboratoire d'Innovation Thérapeutique (LIT), UMR7200 CNRS-Université de Strasbourg, F-67400 Illkirch

Ultra-large 'on-demand' combinatorial libraries [1, 2] are revolutionizing virtual screening strategies aimed at identifying innovative hit compounds or guide fast hit to lead optimization. Due to their size (several billion molecules), these libraries are encoded as fragment spaces defined by starting building blocks and organic chemistry yielding the fully enumerated compounds, and required ad-hoc browsing similarity search algorithms [3]. The pairwise maximum common substructure (MCS) similarity across commercial ultra-large fragment spaces has recently been addressed and shown to be surprisingly low [4]. However, the MCS similarity to commercially available drug-like libraries ('on-the-shelf' physically-available compounds) remains unknown.

We therefore assembled a library of 9.3 million drug-like compounds from 25 trustable suppliers using a series of in-house druggability filters. This 'on-the-shelf' chemical space was next represented by a list of 2.4 million Bemis-Murcko scaffolds [5] that were searched, using the SpaceMACS algorithm [6], in three on-demand fragment spaces: Enamine's REAL (19 billion compounds), Otava's CHEMryia (11 billion compounds), and WuXi's GalaXi (2.1 billion compounds). Surprisingly, only the REAL fragment space significantly overlapped the commercial MCS space, suggesting that the later 'on-demand' fragment space is the most suitable for most hit to lead follow-up studies. The remaining two spaces (CHEMryia, GalaXi) are however interesting notably for primary hit identification among chemical spaces clearly orthogonal to that covered by commercially available screening decks. Potential biases in the current analysis will be discussed.

Bibliography :

[1] Gastreich, M. et al. Drug Discov. Today, (2019), 24, 1148-1156.
[2] Gorgulla, C. et al. Nature (2020), 580, 663–668.
[3] Warr, W.A. et al. J. Chem. Info. Model., (2022), https://doi.org/10.1021/acs.jcim.2c00224
[4] Bellmann, L. et al. J. Chem. Inf. Model. (2022), 62, 553-566.
[5] Bemis, G.W. et al. J. Med. Chem. (1996), 39, 2887-2893.
[6] Schmidt, R., Klein, R. and Rarey, M. J. Chem. Info. Model. (2021), https://doi.org/10.1021/acs.jcim.1c00640

[P22] Structure-Guided Design of Novel Tubulin Binders – Towards Site-Specific Cysteine Targeting

H. PEREZ-PEÑA[1,2*], M. SHEVELEV[1*], A-C. ABEL[2,3*], Z. BOIARSKA[2,3*], F. BONATO[2,4*], A. SOLIMAN[4*], M. Á. OLIVA[4], F. J. DÍAZ[4], A. VARNEK[1], M. O. STEINMETZ[3], D. HORVATH[1], S. PIERACCINI[2], D. PASSARELLA[2], A. E. PROTA[3]

[1] *Laboratoire de Chémoinformatique, UMR 7140, Université de Strasbourg, 1 rue Blaise Pascal, 67000 Strasbourg (France)*
[2] *Department of Chemistry, Università degli Studi di Milano, Via Golgi 19, 20133 Milan (Italy)*
[3] *Laboratory of Biomolecular Research, Paul Scherrer Institute Forschungsstrasse 111, 5232 Villigen PSI (Switzerland)*
[4] *Centro de Investigaciones Biológicas Margarita Salas Consejo Superior de Investigaciones Científicas, Ramiro de Maeztu 9, 28040 Madrid CSIC (Spain)*

Microtubule (MT)-targeting agents (MTAs) have been shown to be potent modulators of cellular growth of outstanding practical importance (in particular anticancer agents like paclitaxel, maytansine). However, these are no homogeneous family – neither chemically, nor mechanistically speaking – and therefore discovery of new species with distinct action mechanisms may open novel perspectives. The European ITN network TubInTrain has therefore dedicated a work package to the rational discovery of novel α-tubulin binders.

Seven small molecule binding sites on tubulin were known[1], until they were recently complemented with 11 novel ones through a comprehensive crystallographic fragment screen[2]. Furthermore, it was demonstrated that it is possible to “grow” a fragment binding to a novel site at the inter-dimer interface, into a fully active tubulin targeting agent (dubbed “Todalam”), able to modulate MT dynamics[3].

First, virtual screening based on targeted substructure replacement of Todalam, pharmacophore modelling and docking was successfully applied in order to discover alternative chemotypes (scaffold hopping) able to target the Todalam site. Alternative binders were validated by obtention of their tubulin-bound X-ray structures and in vitro MT-polymerization assays.

It was noticed that Todalam site binders are proximal to residues αCys4 and αCys200 of a-tubulin. This inspired us, knowing that only one covalent MT binder is known so far (pironetin[4]). Resulting covalent Todalam site binders could serve as molecular probes to investigate further MT dynamics and breakdown associated with neurodegeneration.

In order to discover ways to covalently link the so-far discovered binding scaffold to the cysteines, we performed a virtual screening of fragment-like compounds with Cys-binding warheads from commercial libraries and fitting the chemical synthesis protocols that can be performed in our laboratory. The top-scoring products were then filtered by a strict selection process, heavily based on the binding pose (conservation of ligand-site contacts undergone by Todalam). Their stability within the binding pocket was assessed by molecular dynamics simulations. Promising candidates will be synthetized, subjected to X-ray crystallography experiments and in vitro MT-polymerization assays.

Bibliography :

[1] (a) M. O. Steinmetz, A. E. Prota, Trends Cell Biol (2018), 28, 776-792;
(b) S. Matthew, Q. Y. Chen, R. Ratnayake, C. S. Fermaintt, D. Lucena-Agell, F. Bonato, A. E. Prota, S. T. Lim, X. Wang, J. F. Diaz, A. L. Risinger, V. J. Paul, M. A. Oliva, H. Luesch, Proc Natl Acad Sci USA (2021), 118.
[2] T. Mühlethaler, D. Gioia, A. E. Prota, M. E. Sharpe, A. Cavalli, M. O. Steinmetz, Angew Chem Int Ed Engl (2021), 60, 13331-13342.
[3] T. Mühlethaler, Milanos, L., Ortega, J. A., Blum, T. B., Gioia, D., Prota, A. E., Cavalli, A., Steinmetz, M. O., submitted.
[4] A. E. Prota, J. Setter, A. B. Waight, K. Bargsten, J. Murga, J. F. Diaz, M. O. Steinmetz, J Mol Biol (2016), 428, 2981-2988.

[P23] HIV-1 Drug Resistance Profiling Using Amino Acid Sequence Space Cartography

Karina PIKALYOVA[1], Alexey ORLOV[1], Arkadiy LIN[1], Oolga TARASOVA[2], Gilles MARCOU[1], Dragos HORVATH[1], Vladimir POROIKOV[2], Alexander VARNEK[1]

[1] Laboratoire de Chémoinformatique, UMR 7140, Université de Strasbourg, 1 rue Blaise Pascal, Strasbourg 67000, France
[2] Institute of Biomedical Chemistry, Pogodinskaya Str., 10/8, Moscow 119121, Russia

The global epidemic of human immunodeficiency virus (HIV) infection is one of the major healthcare problems worldwide. The development of drug resistance decreases the efficacy of antiretroviral therapy that plays a pivotal role in HIV treatment. As a result, an optimal choice of the treatment regimens is required to avoid the emergence of the resistance. Computational methods allowing to predict resistance profiles based on genotypic data represent a perspective alternative to time- and labour-consuming experimental measurements. However, current computer-based approaches for drug resistance predictions are either not suitable for emerging HIV strains with complex mutational patterns [1] or lack interpretability, which is of primary importance in clinical practice. In this study, we propose a methodology for drug resistance profiling, which addresses these drawbacks (Figure 1). The methodology is based on an interepretable machine learning method, Generative Topographic Mapping (GTM) [2]. GTM was applied to build illustrative maps of HIV viral protein sequence space and to model quantitative genotype-phenotype relationships (QGPR). The sequence and associated resistance data were retrieved from the Stanford HIV drug resistance database [3]. The sequences of the three viral proteins integrase (IN), protease (PR), and reverse transcriptase (RT), encoded as variable length k-mers counts vectors and associated with drug resistance profiles for 20 anti-HIV drugs were used for modeling. The QGPR models represented as GTM-based resistance landscapes enabled us to predict HIV drug resistance with accuracy comparable with other machine learning methods:random forest, support vector machine, gradient boosting. The average balanced accuracy for PR inhibitors was 0.89±0.01, for IN inhibitors 0.85±0.01, for non-nucleoside RT inhibitors 0.73±0.01, and nucleoside RT inhibitors 0.84±0.01. The GTM-based resistance and mutation landscapes have been shown to be effective for the in-depth investigation of the resistance profiles through examination of the relationships between mutation patterns and drug resistance. The role of specific mutations (e.g. V32I, L10F, and L33F in HIV PR) for the resistance development was predicted and retrospectively validated using literature data. Overall, this work highlights potential perspectives of the GTM application in the field of bioinformatics as an interpretable machine learning tool for the illustrative sequences space exploration. [4]

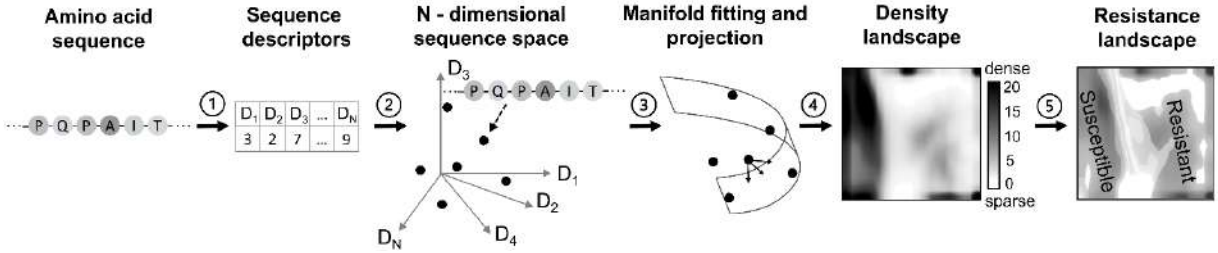


Figure 3. Key steps of the GTM method applied to amino acid sequence space. Each amino acid sequence is encoded by a numerical vector (1) defining its position in the N-dimensional descriptor space (2). The flexible manifold is fitted in a way to approach the data points followed by projection of the data points onto the manifold (3). The unbending of the manifold yields a 2D map. Each projected datapoint is characterized by a probability to be located in the nodes of a rectangular grid superposed with the manifold. (4) Each node is then associated ("coloured") with a weighted average of resistance values of residing sequences. Ensemble of the coloured nodes forms the resistance landscape (5).

Bibliography :

[1] Y. Singh, Healthc. Inform. (2017)Res. 23, 271.
[2] C. M. Bishop, M. Svensén, C. K. I. Williams, GTM: The Generative Topographic Mapping, (1998)
[3] S.-Y. Rhee *et al.*, Nucleic Acids Res. (2003) 31, 298.
[4] K. Pikalyova *et al.*, Bioinformatics (2022) btac090.

[P24] Exploration of the Chemical Space of DNA-Encoded Libraries

Regina PIKALYOVA[1], Yuliana ZABOLOTNA[1], Dmitriy VOLOCHNYUK[2,3], Dragos HORVATH[1], Gilles MARCOU[1], Alexandre VARNEK[1]

[1]*University of Strasbourg, Laboratory of chemoinformatics, Strasbourg, France.*
[2]*Institute of Organic Chemistry, National Academy of Sciences of Ukraine, Kyiv, Ukraine.*
[3]*Enamine Ltd., Kyiv, Ukraine.*

DNA-Encoded Library[1] (DEL) technology has emerged as an alternative method for the discovery of bioactive molecules in medicinal chemistry. It enables quick synthesis and screening of compound libraries of enormous size. Despite its growing popularity, very few reports are devoted to the analysis of DEL chemical space. Therefore, in this work, we aimed to (1) computationally generate and analyze the ultra-large chemical space of DELs and (2) thereupon estimate individual DEL relevance for primary biological screening when little or no information about a biological target and its binders is available[2]. Around 2500 DELs containing 2,5B compounds were designed from commercially available building blocks and DEL-compatible reactions using eDesigner[3], which is a freely available computational tool for DELs generation. The resulting DELs were analyzed and compared to the ChEMBL[4] library of biologically tested molecules using Generative Topographic Mapping[5] (GTM) to identify the DEL(s) significantly occupying ChEMBL space, thus best suited for primary screening (figure 1). A so-called "golden" DEL most suitable for this task allowing to cover 60% of ChEMBL chemical space was found. Different combinations of DELs were also analyzed to find a set of mutually complementary libraries that can be pooled together and used for screening, allowing to attain even higher coverage of ChEMBL chemical space. In such a way, sets of three and five DELs containing the maximum possible percentage of biologically relevant chemotypes were selected. The developed approach is not limited to the selection of libraries for primary screening, given a library of active molecules against a particular biological target it can be also used to design focused DELs.

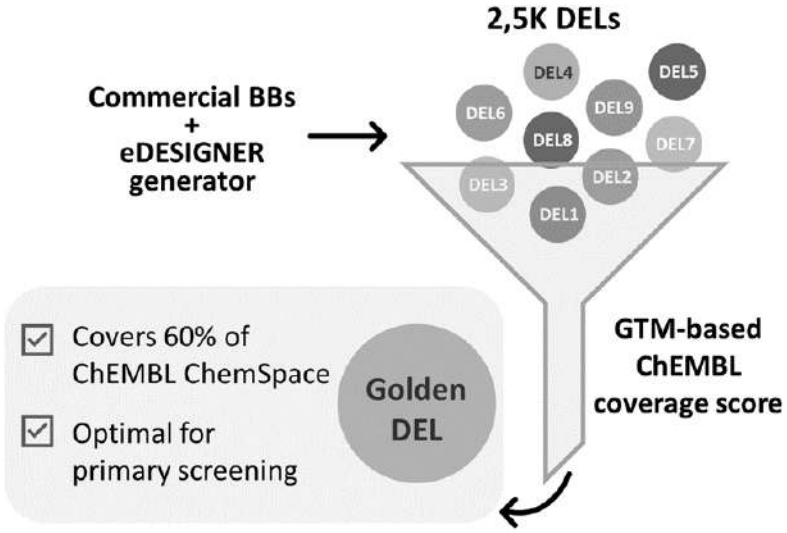


Figure 1. General workflow of GTM-based selection of the optimal DEL for primary biological screening from DELs generated by eDesigner.

Bibliography :

[1] S. Brenner et al., Proc. Natl. Acad. Sci. USA (1992), 89, 5381–5383.
[2] R. Pikalyova et al., Mol. Inf. (2022), 2100289, https://doi.org/10.1002/minf.202100289.
[3] A. Martin et al., Commun. Chem. (2020), 3, 1–9, ARTN 127.
[4] D. Mendez et al., Nucleic Acids Res. (2019), 47, D930–D940.
[5] C. M. Bishop et al., Neural Computation (1998), 10, 215–234.

[P25] Implementation of a Soft Grading System for Chemistry in a Moodle Plugin

Plyer LOUIS[1], Marcou GILLES[2]

[1] *Université de Strasbourg, Faculté de Chimie*
[2] *Université de Strasbourg, Faculté de Chimie, Laboratoire de Chémoinformatique - UMR7140*

We report a novel approach for grading chemical structure drawings for remote teaching, integrated in the Moodle platform. Typically, existing online platforms use a binary grading system, which often fails to give a nuanced evaluation of the answers given by the students. Therefore, such platforms are unevenly adapted to different disciplines. This is particularly true in the case of chemical structures, where most questions simply cannot be evaluated on a true/false basis. Specifically, a strict comparison of candidate and expected chemical structures is not sufficient when some tolerance is deemed acceptable. To overcome this limitation, we have developed a grading workflow based on the pairwise similarity score of two considered chemical structures. This workflow is implemented as a Moodle plugin, using the Chemdoodle engine for drawing structures, and communicating with a REST server to compute the similarity score using molecular descriptors. The plugin is easily adaptable to any academic user; both embedding and similarity measures can be configured.

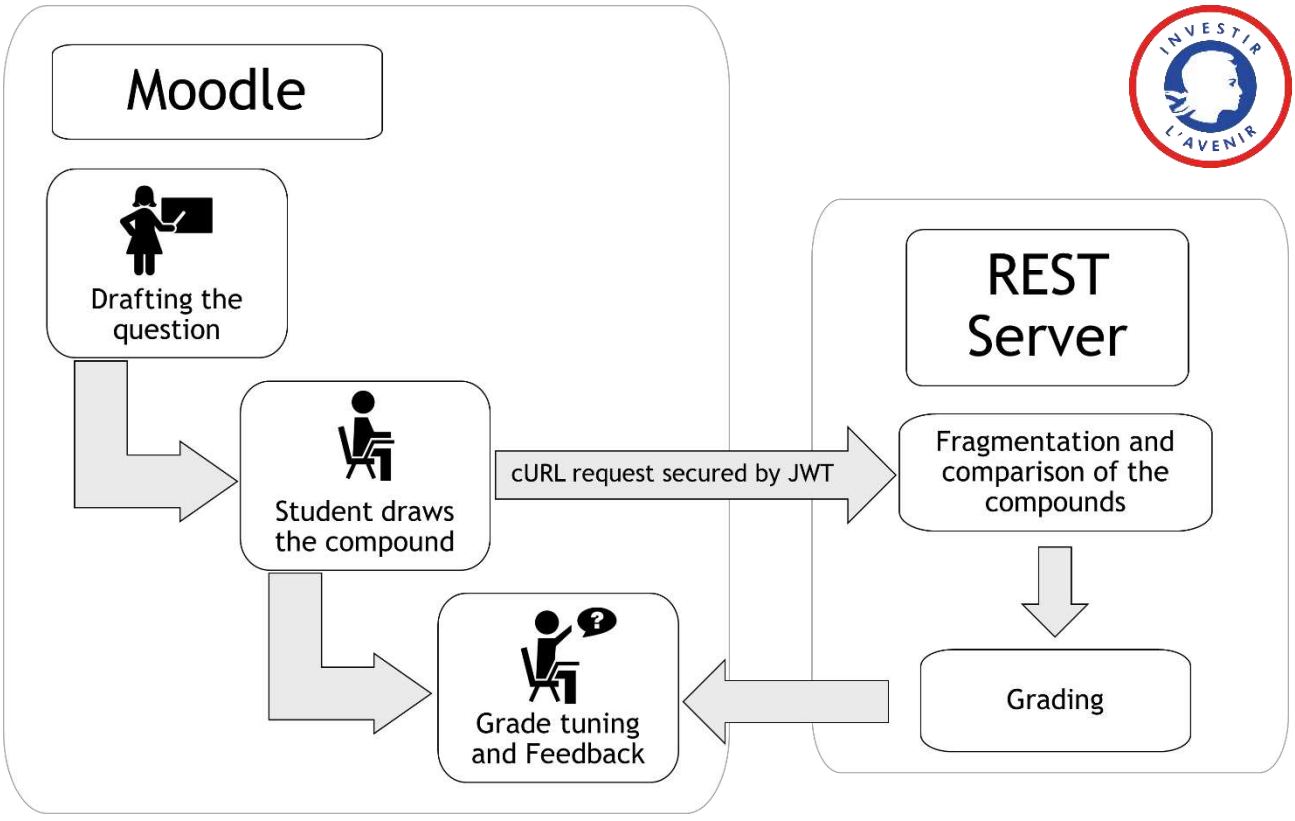


Figure 1. Workflow of the plugin. The teacher drafts a question, which is proposed to the student via the Moodle interface. The student and expected answers are sent to a server for comparison and soft grading. The grade is returned to Moodle for evaluation and feedback to the student.

[P26] Comprehensive Analysis of Commercial Fragment Libraries

Julia REVILLO IMBERNON[1], Célien JACQUEMARD[1], Guillaume BRET[1], Gilles MARCOU[2] and Esther KELLENBERGER[1]

[1] Laboratoire d'Innovation Thérapeutique, UMR 7200, CNRS-Université de Strasbourg, 74 Route du Rhin, 67400 Illkirch-Graffenstaden, France
[2]Laboratory of Chemoinformatics, UMR 7140, CNRS-Université de Strasbourg, 4 rue Blaise Pascal, 67000 Strasbourg, France

For the last 25 years, Fragment-Based Drug Discovery (FBDD) has widely increased in popularity and proven its interest by connecting many worlds, from computational chemistry to biophysics.[1] It has become an alternative to High-Throughput Screening (HTS) and has the advantage of covering a large chemical space with a small number of fragments while providing structural information for the elaboration of hit into druglike compound.[2]

This work aims to analyze the composition of commercial fragment libraries. We focused on important topics on FBDD: molecular obesity[3], three-dimensionality[4] and chemical diversity.

We collected the fragments of 86 freely-available libraries from 14 suppliers. We determined, for the full ensemble of fragments, the chemical descriptors related to the Rule of 3[5] (MW ≤ 300, logP ≤ 3, hydrogen bond donor ≤ 3, hydrogen bond acceptor ≤ 3), and three-dimensional descriptors (PBF[6], SASA, 3D-PSA). To assess the chemical diversity of libraries, we studied the number and frequencies of chemical scaffolds, and analyzed the fragment space using Generative Topographic Map[7] (GTM).

We studied 754 646 molecules, 512 284 after filtering the duplicates. The library size ranges from 80 to 172 723 compounds. The small libraries, containing a maximum of 2000 molecules, are the most interesting with respect to experimental testing. The analysis of the 2D and 3D descriptors showed that MW and logP distributions are globally well balanced in small libraries and that there is a bias towards flat molecules. The scaffold analysis revealed a sur-representation of very simple scaffolds as well as many scaffolds present in only one molecule. Finally, the analysis of the GTM landscapes allowed the systematic comparison of the libraries by pairs. It also allowed to evaluate whether a library is representative of the full fragments set.

In conclusion, our results provide guidelines for the selection or the design of an adequate library for a specific project.

Bibliography:

[1] Romasanta A. K. S., van der Sijde P., Hellsten I., Hubbard R. E., Keseru G. M., van Muijlwijk-Koezen J., de Esch I. J. P, Drug Discovery Today 23 (2018) 1596-1609.
[2] Lipinski C., Hopkins A., Nature 432 (2004) 855-861.
[3] Hann M. M., Keserü G. Nat Rev Drug Discov 11 (2012) 355–365.
[4] Lovering F., Bikker J., Humblet C., J. Med. Chem. 52 (2009) 6752-6756.
[5] Congreve M., Carr, R., Murray, C. & Jhoti, H., Drug Discov. Today 8 (2003) 8, 876– 877.
[6] Firth N. C., Brown N., Blagg J., J. Chem. Inf. Model 52 (2012) 52, 2516-2525.
[7] Kireeva N., Baskin I., Gaspar H. A., Horvath D., Marcou G., Varnek A., Mol. Inf. 31 (2012) 301-312.

[P27] Chemo-Biological Analysis Applied to the Olfaction Field

Marylène RUGARD[1], Anne TROMELIN[2], Karine AUDOUZE[1]

[1] Université Paris Cité, Inserm UMR S-1124, 45 Rue des Saints-Pères, 75006 Paris, France
[2] Centre des Sciences du Goût et de l'Alimentation, AgroSup Dijon, CNRS, INRA, Université Bourgogne Franche-Comté, 9E Boulevard Jeanne d'Arc, F-21000 Dijon, France

Human are surrounded by numerous odorant molecules in the air from a limited number of olfactory receptors (1), and the odor perceived from these molecules does not comes from the isolated molecules but from their mixture (2). According to the odor perceived from the mixture, we differentiate the heterogeneous perception which is the distinction between the odor components of the mixture and the homogeneous perception that corresponds to the perception of a single odor from the mixture preventing the distinction between odor components of the mixture. There are two types of homogeneous perceptions: the masking and the blending mixture. Masking occurs when one of the mixture components covers the other constituents and we can only smell the odor of one of the mixture components. And, we will speak of blending mixture when the perceived odor is different from each of the mixture components. In the present study, we focus on both types of homogeneous perception with a red cordial (RC) mixture involving vanillin, isoamyl acetate, frambinone, ethyl acetate, beta-ionone, beta-damascenone (blending mixture) (3) and a woody-fruity mixture (masking mixture) involving whiskey lactone (WL) and isoamyl acetate (IA) (4). The aim was to improve the understanding about homogeneous perception combining classification and pharmacophore approaches. By collecting from different data sources, we built a large dataset listing more than 5000 odorant molecules associated with their odors. Using the fingerprints representing the structure of these molecules, we performed a dimension reduction by the Uniform Manifold Approximation and Projection and then a Self-organizing map (SOM) classification on the coordinates of the reduced dimensions that led to define specific clusters. Pharmacophore modeling was performed to determine the presence of common features between the molecules for each of the mixtures. We used (i) subsets of the components of the aroma mixtures; and (ii) subsets of about ten of molecules selected on the basis of their odor notes in the SOM clusters containing the components of the mixtures. About the masking, relevant hypotheses were obtained from the WL and IA subsets, unlike the subsets constituted by the RC mixture components, from which no relevant hypotheses were generated. These results suggest that WL and IA could have common binding sites while the components of RC mixture have probably not. Consequently, two different ways for the formation homogeneous perception could exist. The configural perception of the masking could take place at the peripheral level whereas for the RC mixture, the configural perception might rather require the signal integration at higher levels, in the olfactory bulb and/or in the brain.

This work was supported by Agence Nationale de la Recherche (ANR-18-CE21-0006-01 MULTIMIX).

Bibliography :

[1] Buck L, Axel R. A novel multigene family may encode odorant receptors: A molecular basis for odor recognition. Cell. avr (1991);65(1):175-87.
[2] Thomas-Danguin T, et al. The perception of odor objects in everyday life: a review on the processing of odor mixtures. Front Psychol. (2014); 5:504.
[3] Le Berre E, et al. Perceptual Processing Strategy and Exposure Influence the Perception of Odor Mixtures. Chemical Senses. 24 oct (2007);33(2):193-9.
[4] Atanasova B, et al. Perceptual Interactions in Odour Mixtures: Odour Quality in Binary Mixtures of Woody and Fruity Wine Odorants. Chemical Senses. 1 mars (2005);30(3):209-17.

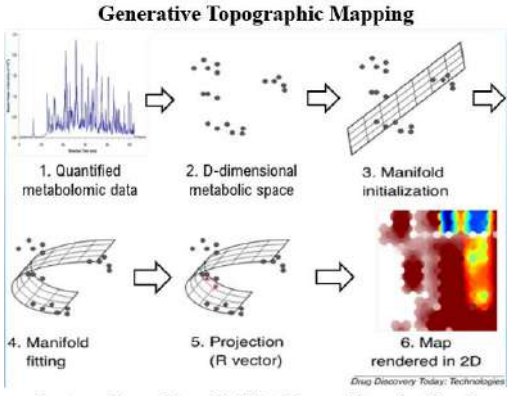
[P28] Visualization and Analysis of Metabolomic Space of Alzheimer's Disease Using Generative Topographic Mapping

Sai Prashanth SANTHAPURI[1,2], Dragos HORVATH[1], Silvia MARIN[2], Josep CENTELLES[2], Marta CASCANTE[2] and Alexandre VARNEK[1]

[1] Laboratory of Chemoinformatics, UMR 7140, CNRS University of Strasbourg, 4 rue Blaise Pascal, 67000 Strasbourg, France
[2] Laboratory of Integrative Systems Biology, Metabolomics and Cancer, Department of Biochemistry and Molecular Biomedicine, Faculty of Biology, Universitat de Barcelona, Av Diagonal 643, 08028 Barcelona, Spain

Alzheimer's disease (AD) is a genetic and sporadic neurodegenerative disease occurring as a consequence of accumulation of β-amyloid and tau proteins in the brain. Conclusive evidence suggests metabolic deficiencies (such as glucose metabolism, mitochondrial dysfunction, etc.) contribute to, or at least are co-occurring with AD pathogenesis. We report the analysis of a dataset of mass spectra obtained using the Biocrates P180 kit and mass spectrometry for the quantification of biologically active molecules (for instance: amino acids, biogenic amines, lipids) within the AD samples (blood, tissues, etc.). The aim is to detect a correlation between metabolite levels and AD state of a patient (diseased / healthy).

Metabolomic data has been obtained from Alzheimer's disease neuroimaging initiative (ADNI) and AD Knowledge portal. The metabolomic data represents a multidimensional space. In this project, we apply a novel non-linear dimensionality reduction method called Generative topographic mapping (GTM) to explore the metabolomic space in AD. GTM can be used to understand whether and how the metabolic profiles of patients in various stages of AD are different and thus separated on the map. By contrast to standard Principal Component Analysis typically used in metabolomic studies, GTM has certain key advantages: it is a non-linear method using fuzzy logics to map every item (patient defined by his or her vector of metabolites) on relevant map "nodes", with various degrees of patient-node association (these degrees of associations are termed "responsibilities"). Based on these responsibilities, map nodes can be "colored" as a function of the clinical status (diseased / healthy in a most simple binary scenario) of associated patients. Resulting fuzzy clinical class landscapes can be analyzed to (a) understand the metabolite concentration signatures that are specific to patients in each clinical class and (b) tentatively predict the clinical status of any new patient based on measured metabolite levels – by projecting the metabolite concentration vector on the map and "reading" the predicted clinical class of matching nodes.



Courtesy: Dragos Horvath, Gilles Marcou, Alexandre Varnek

[P29] Predicting Ligand Binding to Nuclear Receptors Using a Pipeline Combining Docking and Pharmacophore Models

Asma SELLAMI, Matthieu MONTES, Nathalie LAGARDE

Laboratoire GBCM, EA 7528, Conservatoire National des arts et métiers, 2 rue Conté, 75003, Paris, France, HESAM Université.

Endocrine disrupting chemicals (EDCs) are compounds able to penetrate the body and to interfere with the functions of the endocrine system (1). EDCs are considered as a public health threat since human exposure to these compounds have been associated with increased risk of several diseases (2,3). It has been shown that EDCs can act through direct binding to nuclear receptors (NR) which leads to either inhibition or overactivation of the hormonal activity (4). Early detection of potential EDCs becomes an imperative as it is a guarantee of safety for several fields including pharmaceutical, food industry and agriculture. Several health and environmental authorities have been investigating suspicious compounds through experimental testing. However, this remains a challenging task due to the considerable number of compounds to be evaluated. *In silico* methods can be used in complement to prioritize compounds for experimental testing (5).

In this work, we propose a pipeline combining structure-based (SB) and ligand-based (LB) models to predict potential EDCs based on their ability to bind six nuclear receptors: AR, ERα, ERβ, GR, PPARγ and TRα. The pipeline output enables to categorize query compounds into “high”, “intermediate”, “medium” and “low” risk of being NRs binding compounds and thus, accordingly to the direct mechanism, potential EDCs. To build the pipeline, data was collected from the EPA Comptox dashboard (6) gathering structurally diverse compounds experimentally tested in multiple endpoints against several protein receptors. The dashboard was filtered to only keep the compounds tested in binding assays against each studied NR leading to six individual datasets. Each one of these datasets was then employed to build docking, SB and LB pharmacophores models. Each model was optimized, and their combinations have been assessed to select the protocol associated with the best performances for each receptor. The best performances among the six studied NRs were obtained with the ERβ data set for which the combination of docking and pharmacophore models reached high sensitivity, specificity and accuracy values (0.8, 0.6 and 0.65 respectively).

Bibliography :

[1] Lee, H.-R.; Jeung, E.-B.; Cho, M.-H.; Kim, T.-H.; Leung, P. C. K.; Choi, K.-C. J. Cell. Mol. Med., 17 (2013), 1–11.
[2] Fillol, C.; Oleko, A.; Saoudi, A.; Zeghnoun, A.; Balicco, A.; Gane, J.; Rambaud, L.; Leblanc, A.; Gaudreau, É.; Marchand, P.; Le Bizec, B.; Bouchart, V.; Le Gléau, F.; Durand, G.; Denys, S. Environ Int 147 (2021) 106340.
[3] Kumar, M.; Sarma, D. K.; Shubham, S.; Kumawat, M.; Verma, V.; Prakash, A.; Tiwari, R. Front Public Health 8 (2020) 553850.
[4] Combarnous, Y.; Nguyen, T. M. D. Toxics 1 (2019), n°5.
[5] Montes-Grajales, D; Olivero-Verbel, J. Toxicology 327 (2015), 87–94.
[6] Williams, A. J.; Grulke, C. M.; Edwards, J.; McEachran, A. D.; Mansouri, K.; Baker, N. C.; Patlewicz, G.; Shah, I.; Wambaugh, J. F.; Judson, R. S.; Richard, A. M. J Cheminform 9 (2017) n° 61.

[P30] Machine-Learning for Predicting Material Properties with Atomistic Potentials

Shokirbek SHERMUKHAMEDOV and Michael PROBST

Institute of Ion Physics and Applied Physics, University of Innsbruck, Technikerstrasse 25, 6020 Innsbruck, Austria

The interaction of plasma and hot gases, for example argon (Ar) with walls and the tungsten (W) divertor is one of the main problems for using tokamaks as fusion devices since heavy atoms can contaminate the plasma and cause disruptions. Despite a large amount of experimental and computational data on tungsten, the sputtering yields caused by low-energy Ar atoms are still poorly understood [1]. The best way to describe such systems requires fitting of very accurate atomistic potential energy functions based on quantum chemical calculations. High Dimensional Neural Network Potentials (NNPs) have shown to be very efficient for such purposes [2]. In this work we trained a feedforward neural network potential with two hidden layers of 25 nodes each. The input nodes consist of weighted radial and angular Behler-Parrinello type symmetry functions [3]. 6230 configurations containing 72 and 144 W atoms and one Ar atom each served as reference data. 5655 total energies and 1 140 705 forces are used to train the final NNP. The remaining 10% of the structures were used as test set. Within 50 training steps, the RMSE in the test set converged to 0.94 meV/atom for energies and 0.17 eV/Å for atomic forces. The corresponding values in the training set are 0.97 meV/atom and 0.19 eV/Å. The range of potential energies and forces per atom in the training data is quite large. The correlation of the NNP vs. the reference energies and the corresponding forces together with their distributions are shown in figure 1. The trained NNP will be used for further simulating the sputtering by molecular dynamics.

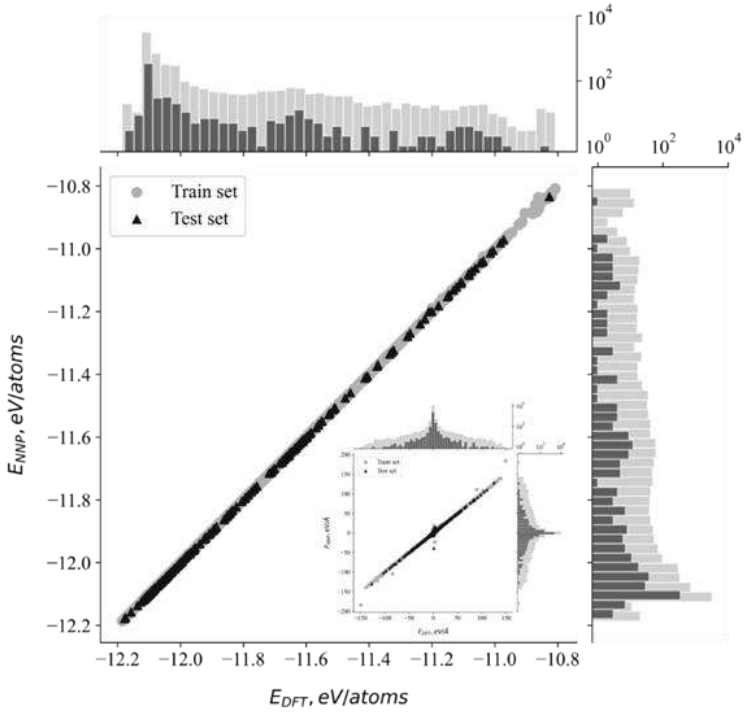


Figure 1. Two-dimensional histograms of DFT-calculated and NNP-predicted atomic energies and forces (subplot).

Bibliography:

[1] M.Rieth, R.Doerner, A.Hasegawa, , Y.Ueda, and M. Wirtz. J. Nucl. Mater. 519 (2019) 334-368.
[2] A.Singraber, J.Behler, and C.Dellago. J. Chem. Theory Comput. 15 (2019) 1827–1840
[3] M.Gastegger, L.Schwiedrzik, M.Bittermann, et al. J. Chem. Phys. 148 (2018) 241709

[P31] Protein-Applied Computer Vision and Deep Generative Linking to Generate Potent Kinase Inhibitors: Influence of Fragments Definition

François SINDT, Merveille EGUIDA and Didier ROGNAN

Laboratoire d'Innovation Thérapeutique (LIT), UMR7200 CNRS-Université de Strasbourg, F-67400 Illkirch, France

We recently described a novel computational fragment-based drug design strategy to grow ligands inside a target cavity.[1] First, the query cavity, represented as an image by a point cloud with key shape and pharmacophoric properties, is aligned by a computer vision method to an archive of images describing fragmented ligands-bound subpockets from the Protein Data Bank.[1] The fragments of the most similar PDB subpockets are then directly positioned in the query cavity using the corresponding pocket transformation matrices. Last, suitable connectable atoms of already oriented fragment pairs are linked by a deep generative model[2] to yield fully connected molecules. The above-described approach was applied to design a focused inhibitor library targeting human cyclin dependent kinase 8 (CDK8) and yielded several nanomolar inhibitors.[3] We herein investigate how the fragmentation scheme influence the collection of available fragments/subpockets and drives the final library design.

Bibliography:

[1] Eguida M, Rognan D. A Computer Vision Approach to Align and Compare Protein Cavities: Application to Fragment-Based Drug Design. J Med Chem. (2020), 63:7127-7142.
[2] Imrie F, Bradley AR, van der Schaar M, Deane CM. Deep Generative Models for 3D Linker Design. J Chem Inf Model. (2020), 60:1983-1995
[3] Eguida M., unpublished data

[P32] Study of Odorants Sharing the Odor Notes of an Aroma Blending Mixture by a Pharmacophore Approach

Anne TROMELIN[1], Florian KOENSGEN[1], Marylène RUGARD[2], Karine AUDOUZE[2], Thierry Thomas DANGUIN[1], Elisabeth GUICHARD[1]

[1]Centre des Sciences du Goût et de l'Alimentation, CNRS, INRAE, Institut Agro, Université de Bourgogne Franche-Comté, F-21000 Dijon, France
[2]Université de Paris, INSERM UMR-S1124, F-75006 Paris, France

Odors perceived in our environment are mainly the result of mixtures of odorants whose the specific mechanisms involved in their processing remain poorly understood [1].

In previous studies performed at INRAE-CSGA [2], the perception of a mixture of ethyl isobutyrate (Et-iB, strawberry-like odor, STR) and ethyl maltol (Et-M, caramel-like odor, CAR) was investigated in comparison with a reference (allyl hexanoate, Al-H, pineapple-like odor, PNA) chosen to evoke an odor close to the one expected in the mixture. The binary specific mixture of Et-iB and Et-M was judged as more typical of a pineapple odor than the individual components.

Some studies highlight the significance of the biological function of odorant, that is to say their odor, to understand odorant discrimination [3, 4]. From this perspective, we selected molecules having two of the three odors STR, CAR or PNA from the large odorants database that we recently used to perform a multivariate statistical analysis [5]. The pharmacophore generation were performed separately using 9 molecules STR-CAR on the one hand, and 4 molecules STR-PNA on the second hand. We used Common Feature Pharmacophore Generation protocol (Discovery Studio 4.5, Biovia) to generate pharmacophore hypotheses [6]; the maximum number of generated hypotheses for each run was set to 10. In our study, the pharmacophoric features considered are hydrogen bond acceptors (HBA/HBA-lip) and hydrophobic regions (HY/HY-Al).

All the hypotheses generated from both the STR-CAR and the STR-PNA sets are made up of 2 HBA/HBA-lip. Besides, STR-CAR hypotheses have only 1 HY/HY-Al while there are 2 for STR-PNA hypotheses. We compared the best significant hypotheses generated from STR-CAR and STR-PNA. A distance close to 8 Å between the centers of at least one HY and one HBA is common to STR-CAR and STR-PNA models. The pharmacophore comparison of the two models revealed a satisfactory mapping of the features.

The obtained result is in line with the scheme of olfactory coding, and is consistent with the hypothesis that molecules sharing the odors involved in a blending mixture could recognize a common set of ORs.

This work was supported by the Agence Nationale de la Recherche (ANR-18-CE21-0006-01 MULTIMIX) and INRAE-TRANSFORM department (ANS CAMELIA).

Bibliography:

[1] T. Thomas-Danguin, C. Sinding et al., Front. Psychol., 5 (2014).
[2] E. Le Berre, N. Beno et al., Chem. Senses, 33 (2008) 389-395.
[3] L.M. Ma, Q. Qiu et al., Proc. Natl. Acad. Sci. U. S. A., 109 (2012) 5481-5486.
[4] E. Poivet, N. Tahirova et al., Sci. Adv., 4 (2018).
[5] A. Tromelin, C. Chabanet et al., Flavour Frag. J., 33 (2018) 106-126.
[6] O.O. Clement, A. Mehl, in: O.F. Güner (Ed.) Pharmacophore perception, Development and Use in Drug Design, International University Line, La Jolla, 2000, pp. 69-84.

[P33] Applicability of Graph Neural Networks to Binding Affinities Prediction from Protein-Ligand Structures

Mikhail VOLKOV [1, 2], Joseph-André TURK [2], Nicolas DRIZARD [2], Nicolas MARTIN [2], Brice HOFFMANN [2], Didier ROGNAN [1]

[1]*Laboratoire d'Innovation Thérapeutique, UMR7200 CNRS-Université de Strasbourg, 74 Route du Rhin, F-67400, Illkirch, France*
[2]*Iktos, 65 rue de Prony, 75017, Paris, France*

Accurate prediction of binding affinities in protein-ligand complexes from their 3D-structures remains a major challenge in drug discovery. Repetitive claims that deep learning methods represent a significant breakthrough in quantifying protein-ligand binding have been contradicted by independent groups warning about potential biases in ligand and protein composition of training/test sets of experimentally determined structures of protein-ligand complexes such as the PDBBind dataset [1]. In the current study we use message passing graph neural networks [2] , which can be trained on independent representations of ligand and protein as well as on their interaction pattern. We show that explicit description of protein-ligand non-covalent interactions does not provide any advantage with respect to simple ligand or protein descriptors. Simple models, inferring binding affinities of test samples from that of the closest ligands or proteins in the training set, already exhibit good performances suggesting that memorization largely dominates true learning in deep neural networks applied to the binding affinity prediction problem. The current study suggests considering only non-covalent interactions while omitting their protein and ligand atomic environments. The simple interaction-only model appears to be robust enough to be applicable to various external test sets with an acceptable accuracy. Removal of all hidden biases from accessible datasets probably requires much denser protein-ligand training matrices and a coordinated effort of the drug design community to solve the necessary protein-ligand structures.

Bibliography:

[1] Liu Z.; Li Y.; Han L.; Li J.; Liu J.; Zhao Z.; Wang R. Bioinformatics, 31(3) (2015) 405-412.
[2] Gilmer J.; Schoenholz S. S.; Riley P. F.; Vinyals O.; Dahl G. E. Proceedings of the 34th International Conference on Machine Learning, PMLR 70 (2017) 1263-1272.

[P34] Interpreting Neural Network Models for Toxicity Prediction by Extracting Learned Chemical Features

Moritz WALTER[1], Samuel J. WEBB[2], Valerie J. GILLET [1]

[1]Information School, University of Sheffield, Regent Court, 211 Portobello, S1 4DP, Sheffield, UK.
[2]Lhasa Limited, Granary Wharf House, 2 Canal Wharf, LS11 5PS, Leeds, UK

Deep Neural Network (DNN) models have become a popular machine learning technique for toxicity prediction of chemicals. Due to their complex structure it is difficult to understand predictions made by these models which limits confidence. Current approaches to tackle this problem such as SHAP or integrated gradients provide insights by attributing importance to input features of individual compounds.[1,2] While these methods have produced promising results, they do not shed light on representations of compounds in hidden layers. In the realm of image classification models, feature visualization has been developed as a tool to depict features learned in specific hidden layer neurons.[3] Detecting complex chemical features learned in hidden layers might complement current approaches for interpreting DNN models used for toxicity prediction.

The present study focuses on feedforward neural networks with RDKit's Morgan fingerprints as input. A novel method was developed to automatically extract chemical features responsible for activation of hidden neurons. This method leverages both information about training compounds strongly activating hidden neurons and learned model parameters. Using Ames mutagenicity as a well-understood toxicity endpoint, the method was able to extract known toxicophores. Moreover, extracted substructures can be mapped onto test compounds to obtain model explanations incorporating hidden layer representations of compounds. Using toxicophores from the Derek expert system[4] as ground truth, the explanatory capability of the approach was evaluated using attribution AUCs as metric.[5] For the majority of compounds provided explanations match the ground truth very well with AUC values above 0.8.

The proposed method may be used to extract novel toxicophores by leveraging chemical features encoded in DNN models. Furthermore, understanding of model predictions is increased by providing explanations complementary to those obtained with established attribution methods. While not explored in the present study, the proposed method could be adapted to other DNN architectures such as graph-convolutional neural networks.

Bibliography:

[1] Rodriguez-Perez, R., et al.; Interpretation of Compound Activity Predictions from Complex Machine Learning Models Using Local Approximations and Shapley Values. J. Med. Chem. 63 (2020) 8761-8777.
[2] Preuer, K., et al.; Interpretable Deep Learning in Drug Discovery. arXiv :190302788v2 (2019).
[3] Olah, C., et al.; Feature Visualization. Distill (2017).
[4] Marchant, C., et al.; In Silico Tools for Sharing Data and Knowledge on Toxicity and Metabolism: Derek for Windows, Meteor and Vitic. Toxicol. Mech. Methods. 18 (2008) 177-187.
[5] McCloskey, K., et al.; Using Attribution to Decode Binding Mechanisms in Neural Network Models for Chemistry. PNAS. 116 (2019) 11624-11629.

[P35] GTM-Based Analysis of The Chemical Space of the Chimiothèque Nationale

Polina OLENEVA, Yuliana ZABOLOTNA, Dragos HORVATH, Gilles MARCOU, Fanny BONACHERA, Alexandre VARNEK

Laboratoire de Chémoinformatique, UMR7140 CNRS, Institut Le Bel, University of Strasbourg, 4 rue Blaise Pascal, 67081, Strasbourg, France

Chimiothèque Nationale (CN) — The French National Compound Library — is a library of small molecules and natural products.[1] Containing today more than 83K compounds prepared for purchase in a format convenient for biological screening, this collection might be a potential source of promising hits for drug design. However, there is a lack of detailed analysis of CN, especially with respect to other libraries often used in medicinal chemistry.

In order to identify the place of CN in the chemical space of screening and biologically relevant compounds, the library was compared with ZINC in-stock collection[2] (9M purchasable molecules offered by different suppliers) and the ChEMBL library[3] (1,6M biologically studied compounds). The analysis included physico-chemical properties, Bemis-Murcko (BM) scaffolds[4] and chemical space overlap of those libraries on the 2D Generative Topographic Maps (GTM)[5]. In addition, to estimate the synthetic uniqueness of CN collection, the retrosynthetic analysis was performed with the help of an in-house tool Synt-On[6]. In such a way, the synthesizability of the CN compounds by means of combinatorial chemistry combined with commercially available BBs was assessed.

Moreover, hierarchical GTM ("zooming")[7,8] was applied in order to generate n ensemble of maps enabling different levels of navigation of the chemical space of CN – from the general overview in the main map to the structural patterns in the local areas of the chemical space. The map was separated into small zones and new GTMs were generated focusing only on the compounds extracted from the most populated regions. Areas with a low number of residents (less than 1000 compounds) were analyzed with respect to BM scaffolds. Around 20 physicochemical property landscapes were also precomputed and all results were made available in the framework of ChemSpace Atlas (https://infochm.chimie.unistra.fr). On this webpage, users can browse through the thousands of CN compounds, analyze their properties, compare them with molecules from other libraries and perform the analogue search for the small dataset of user-defined compounds.

Bibliography :

[1] F. Mahuteau-Betzer, Medecine/Sciences 2015, 31, 417–422.
[2] J. J. Irwin, K. G. Tang, J. Young, C. Dandarchuluun, B. R. Wong, M. Khurelbaatar, Y. S. Moroz, J. Mayfield, R. A. Sayle, Journal of Chemical Information and Modeling 2020, 60, 6065–6073.
[3] D. Mendez, A. Gaulton, A. P. Bento, J. Chambers, M. de Veij, E. Félix, M. P. Magariños, J. F. Mosquera, P. Mutowo, M. Nowotka, M. Gordillo-Marañón, F. Hunter, L. Junco, G. Mugumbate, M. Rodriguez-Lopez, F. Atkinson, N. Bosc, C. J. Radoux, A. Segura-Cabrera, A. Hersey, A. R. Leach, Nucleic Acids Research 2019, 47, D930–D940.
[4] G. W. Bemis, M. A. Murcko, The Properties of Known Drugs. 1. Molecular Frameworks, 1996.
[5] C. M. Bishop, M. Svensén, C. K. I. Williams, The Generative Topographic Mapping, 1998.
[6] Y. Zabolotna, D. M. Volochnyuk, S. v Ryabukhin, K. Gavrylenko, D. Horvath, K. Olga, O. Oksiuta, G. Marcou, A. Varnek, SynthI: A New Open-Source Tool for Synthon-Based Library Design, 2021.
[7] P. Tino, I. Nabney, Hierarchical GTM: Constructing Localized Nonlinear Projection Manifolds in a Principled Way, n.d.
[8] A. Lin, B. Beck, D. Horvath, G. Marcou, A. Varnek, Journal of Computer-Aided Molecular Design 2020, 34, 805–815.

[P36] ChemSpace Atlas: Multiscale Chemography of Ultra-Large Libraries for Drug Discovery

Yuliana ZABOLOTNA, Fanny BONACHERA, Dragos HORVATH, Olga KLIMCHUK, Arkadiy LIN, Gilles MARCOU, Alexandre VARNEK

Laboratoire de Chémoinformatique, UMR7140 CNRS, Institut Le Bel, University of Strasbourg, 4 rue Blaise Pascal, 67081, Strasbourg, France

Nowadays, drug discovery is inevitably intertwined with the usage of large compound collections. Understanding of their chemotype composition and physicochemical property profiles is of the highest importance for successful hit identification. However, the development of efficient polyfunctional tools allowing multi-faceted analysis of the constantly growing chemical libraries is complicated by the Big Data challenges.

Thus, we present ChemSpace Atlas (https://infochm.chimie.unistra.fr/) - an intuitive polyvalent tool for the ultra-large chemical space exploration and analysis with respect to medicinal chemistry problems. Being based on the hierarchical ensemble of tens of thousands of Generative Topographic Maps(GTM)[1], it provides access to ChEMBL, ZINC and COCONUT collections allowing easy navigation through the hundreds of millions of compounds from a global bird's eye view to structural pattern detection[2]. ChemSpace Atlas functionality is not limited to a simple visualization of the similarity relationships in the chemical space but it also allows users to analyze physicochemical properties and biological activities, perform polypharmacological profiling (around 750 biological activities), analogs search, and detailed structural analysis with the help of MCSs and scaffolds. In Figure 1 one can see the example of the activity visualization page for CDK4 (CHEMBL331) ligand series from ChEMBL.

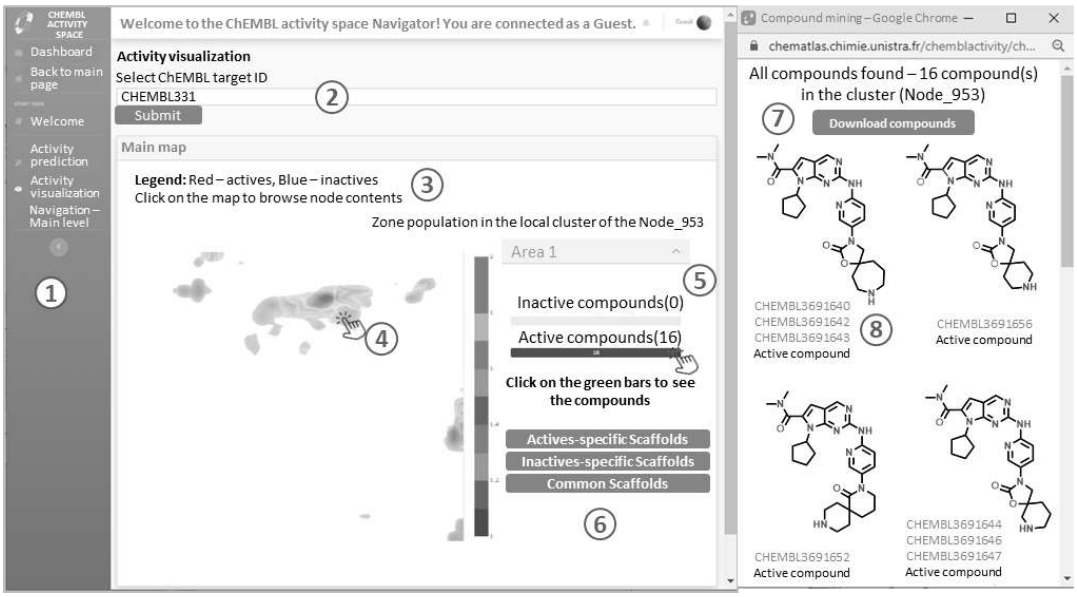


Figure 4. Activity visualization page of the ChEMBL activity space Navigator on the example of CDK4 (CHEMBL331) ligands. 1) navigation bar; 2) target selection menu; 3) legend of the map; 4) interactive activity landscape; 5) zone population information (if green, bars become clickable and corresponding compounds can be displayed); 7) button for downloading compounds from the selected area of the chemical space; 8) direct links to the source database (here ChEMBL).

Bibliography:

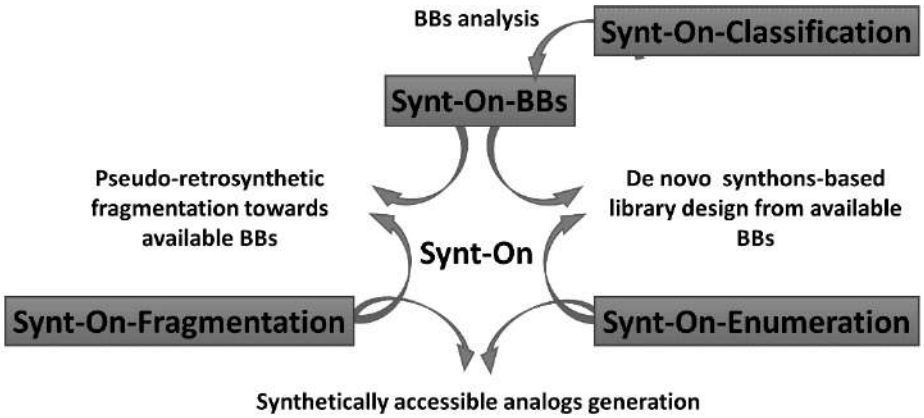
[1] C. M. Bishop, M. Svensén, C. K. I. Williams, Neural Comput. (1998), 10, 215-234.
[2] A. Lin, B. Beck, D. Horvath, G. Marcou, A. Varnek, J. Comput. Aided Mol. Des. (2019).

[P37] Synt-On: A New Open-Source Tool for Synthon-Based Library Design and Building Blocks Analysis

Yuliana ZABOLOTNA[1], Dmitriy M. VOLOCHNYUK[3,6], Sergey V. RYABUKHIN[4,6], Kostiantyn GAVRYLENKO[5,6], Dragos HORVATH[1], Olga KLIMCHUK[1], Oleksandr OKSIUTA[3,7], Gilles MARCOU[1], Alexandre VARNEK[1,2]

[1] University of Strasbourg, Laboratoire de Chemoinformatique , 4, rue B. Pascal, Strasbourg 67081, France
[2] Institute for Chemical Reaction Design and Discovery (WPI-ICReDD), Hokkaido University, Kita 21 Nishi 10, Kita-ku, 001-0021 Sapporo, Japan
[3] Institute of Organic Chemistry, National Academy of Sciences of Ukraine, Murmanska Street 5, Kyiv 02660, Ukraine
[4] The Institute of High Technologies, National Taras Shevchenko University of Kyiv, 64 Volodymyrska Street, Kyiv 01601, Ukraine
[5] Research-And-Education ChemBioCenter, National Taras Shevchenko University of Kyiv, Chervonotkatska str., 61, 03022 Kyiv, Ukraine
[6] Enamine Ltd. 78 Chervonotkatska str., 02660 Kyiv, Ukraine
[7] Chemspace, Chervonotkatska Street 78, 02094 Kyiv, Ukraine

Most of the existing computational tools for de novo library design are focused on the generation, rational selection, and combination of promising structural motifs to form members of the new library. However, the absence of a direct link between the chemical space of the retrosynthetically generated fragments and the pool of available reagents makes such approaches appear as rather theoretical and reality-disconnected. In this context, we present Synthons Interpreter (*Synt-On* former *SynthI*), a new open-source toolkit for de novo library design that allows merging those two chemical spaces into a single synthons space.[1] Here synthons are defined as actual fragments with valid valences and special labels, specifying the position and the nature of reactive centers. They can be issued from either the "break-up" of reference compounds according to 38 retrosynthetic rules (*Synt-On-Fragmentation Module*) or real reagents (building blocks or BBs), after leaving groups withdrawal or transformation (*Synt-On-BBs Module*). Such an approach not only enables the design of synthetically accessible libraries and analogs generation (*Synt-On-Enumeration Module*) but also facilitates BBs analysis in the medicinal chemistry context (*Synt-On-Classification and Synt-On-BBs Modules*). Synt-On code is publicly available at https://github.com/Laboratoire-de-Chemoinformatique/Synt-On.



Bibliography:

[1] Zabolotna Y., Volochnyuk D.M., Ryabukhin S.V., Gavrylenko K., Horvath D, Klimchuk O., Oksiuta O., Marcou G., and Varnek A. J. Chem. Inf. Model. Article ASAP. (2021). ttps://doi.org/10.1021/acs.jcim.1c00754.

[P38] Multi-Instance Learning Approach to Predictive Modeling of Catalyst Enantioselectivity

Dmitriy ZANKOV [1,2], Timur MADZHIDOV [2], Alexander VARNEK [1]

[1] *Laboratory of Chemoinformatics, University of Strasbourg, Strasbourg, 4 rue B. Pascal, France*
[2] *Chemoinformatics and Molecular Modeling, Kazan Federal University, Russia*

The production of enantiomerically pure organic compounds is a hot topic of modern organic chemistry. Enantioselective catalysis is a powerful technology for the synthesis of enantiomerically pure compounds using special organic catalysts. Chemoinformatics is an appealing technology aiming to empower experimentalists in the quest for developing new catalysts. Preliminary theoretical research enables the identification of the most promising catalysts before their experimental testing, reducing the time and overheads needed to find an appropriate catalyst.

We developed a new chemoinformatics-based protocol for constructing accurate models for the prediction of catalyst enantioselectivity. The catalysts were represented by multiple conformations, which were encoded with new 3D descriptors developed in our group and probed in predicting the biological activity of molecules [1]. Models were constructed with multi-instance neural networks. Multi-instance (MI) machine learning algorithms can be applied to process the multiple conformations (instances) of a catalyst. In the multi-instance approach, a molecule (catalyst) is presented by a bag of instances (i.e., a set of conformations), and a label (a selectivity value) is available only for a bag (a molecule), but not for individual instances (conformations). The multi-conformation models were compared with single-conformation models constructed with the lowest-energy catalyst conformation.

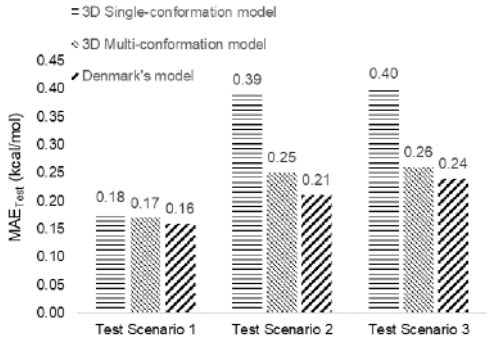


Figure 1: Mean Absolute Error (MAE, kcal/mol) obtained for test sets 1-3.

The 2D, single- and multi-conformation models were built on the training set of 384 data points resulted from a combination of 24 catalysts with 16 reactions. The models were validated on three test sets selected according to different scenarios: (a) new reactions with known catalysts, (b) known reactions with new catalysts, and (c) new reactions with new catalysts. Thus, Test set 1 contained 216 instances resulted from a combination of 24 catalysts from the training set with 9 new reactions, Test set 2 included 314 instances (19 new catalysts / 16 training reactions) and Test set 3 contained 171 instances (19 new catalysts / 9 new reactions). Performances of 2D, single-conformation and multi-conformation models (Mean Absolute Error, MAE) in comparison with those of the model by Zahrt et al. [2] are given in Figure 1. These results demonstrate the importance of accounting for all representative catalyst conformations in predictive modeling.

Bibliography:

[1] Zankov et al., QSAR Modeling Based on Conformation Ensembles Using a Multi-Instance Learning Approach. JCIM, 2021, 61, 4913
[2] Zahrt et al., Prediction of higher-selectivity catalysts by computer-driven workflow and machine learning. Science, 2019, 363, 6424

[P39] Exploration of Bioinformatic Domain Based on Data Mining, Reaction Predictions and Enzyme Promiscuous Predictions

Chonghuan ZHANG[1], Alexei LAPKIN[1,2]

[1] Department of Chemical Engineering and Biotechnology, University of Cambridge, Philippa Fawcett Drive, Cambridge CB3 0AS, United Kingdom
[2] Cambridge Centre for Advanced Research and Education in Singapore, CARES Ltd, 1 CREATE Way, CREATE Tower #05-05, 138602 Singapore

Conventional chemical engineering intakes fossil fuels as raw materials and undergoes harsh synthesis to produce pharmaceuticals and industrial chemicals. In contrast, synthetic biology is interested in bio-renewable feedstocks at various stages of metabolic pathway, and produce target substances in cell organisms or with cell-free biocatalyst, which improves redox efficiency and moderates reaction conditions. Addition of synthetic biology into chemical engineering certainly opens opportunities for more efficient and sustainable chemical productions.[1] The development of synthetic biological routes relies on integration of bioinformatic data, to understand high level functions and connections of the biological systems. With bioinformatics, scientists explored the biological spaces to assist complex molecule (bio)-synthesis. However, our previous work[2] indicates due to the lack of biological molecules and reactions data, the current bioinformatic space was so sparse compared with the chemoinformatic space, the reaction network of organic chemistry (NOC), which comprises only 3.5% of the hybridised organic and biological domain. This limits opportunities to find sustainable alternatives for chemical reactions.

Herein, we proposed a workflow pipeline to explore the sparse synthetic biological domain: we predicted feasible biological molecules and reactions, and populated the current biological dataset. This was specifically based on data mining from a biological database, KEGG, and extended on our previous work[2] of hybrid chemical and biological reaction network. It started from applying a biological-molecule-specific graph method[3] to find reactions centres of KEGG recorded molecules, and detecting enzymatic transformations from KEGG recorded reactions. Based on the referenced reactions, the enzymatic transformations were predicted to functionally and structurally similar molecules. In total, 8,857 molecules and 12,422 reactions were predicted. Confidence scores were given to the predicted reactions to judge the goodness of predictions.

To mediate the novel reactions, instead of designing novel heterogeneous enzymes, which was believed to be uncertain and time-consuming, we focused on enzymatic promiscuity, which expand the specificity of the database recorded enzymes to potential substrates. A machine learning tool, an autoencoder based classifier, was applied to learn from gene sequences of enzymes and molecular structures of enzymes, and suggest promiscuous enzymes to bind with the substrates in the predicted reactions. The well-trained classifier exhibited 83% accuracy from test reaction-enzyme dataset, and case studies indicate the model expertise (top accuracy) to suggest oxidoreductase (EC1) and hydrolases (EC3) enzymes on the test reactions. The binding possibilities of the suggested enzyme from the classifier were given to the prediction reactions. The populated biological space was analysed to investigate its feasibility to guide bioprocess design.

Bibliography:

[1] Weber, J. M.; Guo, Z.; Zhang, C.; Schweidtmann, A. M.; Lapkin, A. A., Chem. Soc. Rev. (2021), 50 (21), 12013-12036.
[2] Zhang, C.; Lapkin, A. A., Hybridizing Organic Chemistry and Synthetic Biology Reaction Networks for Optimizing Synthesis Routes. ChemRxiv (2022).
[3] Hattori, M.; Tanaka, N.; Kanehisa, M.; Goto, S., Nucleic Acids Res. (2010), 38 (suppl_2), W652-W656.

[P40] Computational Elucidation of GPCR Allosteric Modulators

Sonja PETER, Chris de GRAAF, Ijen CHEN

Heptares Therapeutics Limited, Granta Park, Great Abington, Cambridge, CB21 6DG, United Kingdom

G protein-coupled receptors (GPCRs), the largest family of cell signalling trans-membrane proteins, are regulated by diverse small molecules. Allosteric modulators interact with binding sites topologically distinct from the orthosteric ligand binding sites, which can be the extracellular, intracellular, or extrahelical allosteric sites [1, 2, 3], just to name a few. However, the identification of allosteric binding sites has been challenging due to the high diversity of binding modes and protein plasticity upon ligand binding. The advances in GPCR structural biology [4] have allowed to examine allosteric modulators at a more precise ngle. In this poster, I describe the development and the the application of structural bioinformatics and chemogenomics methods to assess ligands from structural GPCRome [5]. The analyses from both structural and ligand perspectives have been applied to both public and proprietary ligand-bound GPCR structures. This includes druggability assessment for the ligand-binding sites and chemoinformatics analysis on the bound ligands. Results are discussed to get a better understanding of all the flavors of allosteric ligands in relation to the orthosteric binders. This work aims to provide an overview and characterize the current GPCR allosteric binding site structure landscape with exciting potential for GPCR drug discovery.

This project receives funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 956314.

Bibliography :

[1] Chris de Graaf et al. “Extending the structural view of class B GPCRs”. In: Trends in biochemical sciences 42.12 (2017), pp. 946–960.
[2] Zhaotong Cong et al. “Structural perspective of class B1 GPCR signaling”. In: Trends in Pharmacological Sciences (2022).
[3] Attila Egyed, Dóra Judit Kiss, and György M Keseru. “The impact of the secondary binding pocket on the pharmacology of Class A GPCRs”. In: Frontiers in Pharmacology (2022), p. 468.
[4] Miles Congreve et al. “Impact of GPCR structures on drug discovery”. In: Cell 181.1 (2020), pp. 81–91.
[5] Márton Vass et al. “Chemical diversity in the G protein-coupled receptor superfamily”. In: Trends in Pharmacological Sciences 39.5 (2018).

[P41] Machine Learning to Discover Antibiotics Against *Klebsiella Pneumoniae*

Marina VILLACAMPA-FERNÁNDEZ[1], Gianluca POLLASTRI[2], Séamus FANNING[3], Denis C. SHIELDS[1]

(1) School of Medicine, Conway, University College Dublin
(2) School of Computer Science, University College Dublin
(3) School of Public Health, Physiotherapy and Sports Science, University College Dublin

The continued efficacy of antibiotics is currently uncertain due to the global dissemination of antibiotic-resistance determinants. Without urgent action, it is projected that by the year 2050 deaths attributable to these infections will reach 10 million. We aimed to discover new antibiotic compounds targeting Klebsiella pneumoniae by applying a chemoinformatics algorithm suited to machine learning. We classified the antibiotic activity of a set of 12,250 compounds based on their minimum inhibitory concentration (MIC) against *Klebsiella pneumoniae*, where 75% of the compounds were classified as inactive. For each molecule, a set of 80 relevant features that appropriately describe the molecular structure was selected. With this dataset, an ensemble model was obtained from a selection of artificial neural networks, with different dimensions, depth and parameters to predict the antibiotic activity probabilities. Results show that the Stochastic Gradient Descent optimizer destabilised the models, whilst the Adam optimizer is faster and more stable. We found that networks of medium sizes (around 6,000 parameters) were optimal. We present the predictions of the model for a database of known antibiotics and for a database of known drugs. From these, the most interesting drugs will be tested in the lab.

This research was supported by SFI 18/CRT/6214 and the EU's Horizon 2020 Marie Sklodowska-Curie grant H2020-SCA-COFUND-2019-945385.

[P42] Pharmacophores vs Circular Fingerprints with Learned Feature Transformation Before Clustering. Comparative Studies on Bcr-Abl Data

Hajar REHIOUI[1] , Abdelkader OUALI[1] , Alban LEPAILLEUR[2] , Jean-Luc LAMOTTE[2], Albrecht ZIMMERMANN[1] , Ronan BUREAU[2], Bertrand CUISSART[1]

[1] Normandie Univ., UNICAEN, CNRS – UMR GREYC, 14000 Caen, France
[2] Centre d'Etudes et de Recherche sur le Médicament de Normandie, Normandie Univ, UNICAEN, CERMN, 14000 Caen, France

We aim at computing the different parts of a dataset of molecules, considering both their chemical structures and their activity towards a biological receptor – in this case Bcr-Abl. To this aim, we start by representing a molecule with three molecular fingerprinting either based on state of the art ECFP4 descriptors, on the FCFP4s or a new description from our laboratories that uses frequent pharmacophores associated to a dataset [1,2]. Then we perform two learned feature transformations by passing the data through two feed forward neural network (FFNN), a supervised FFNN and an unsupervised one [3]. Those two feature transformations allow to weight the data. Finally, we relied on clustering techniques to identify different families of ligands in the studied dataset.
The partitions obtained by using the former computation initiate a comparative study between the different fingerprinting methods: the pharmacophores and ECFP4 and FCFP4 representations. The results have demonstrated that our pharmacophores obtain the best results and allow to distinguish between three principal families of ligands in addition to efficient results on pharmacophoric decoys.

UNION EUROPÉENNE
Fonds européen pour le développement régional

Our work is part of the European project "SCHISM", funded by the European Union within the framework of the Operational Programme ERDF/ESF 2014-2020

Bibliography :

[1] Métivier J-P., Cuissart B. , Bureau R. and Lepailleur A., The Pharmacophore Network: A Computational Method for Exploring Structure−Activity Relationships from a Large Chemical Data Set, J. Med. Chem. (2018), 61, 3551−3564
[2] D. Geslin, A. Lepailleur, J.-L. Manguin, N.-V. Vo, J.-L. Lamotte, B. Cuissart, R. Bureau, J. Chem. Inf. Model. (2022), 62, 678–691.
[3] Yan D., Zhou X., Wang X., & Wang R., An off-center technique: Learning a feature transformation to improve the performance of clustering and classification. Information Sciences, (2019),503, 635-651

[P43] Large-step scaffold hopping benchmark

Philippe Pinel[1, 2], Stéphanie Labouille[1], Brice Hoffmann[1], Véronique Stoven[2]

[1]*Iktos SAS, 75017 Paris, France*
[2]*Center for Computational Biology, Mines ParisTech, PSL Research University, 75006 Paris, France*

An important issue in drug discovery is the ability to identify new molecules with optimized biological and chemical characteristics for a given protein target, potentially belonging to a different structural family compared to known binders. In this context, scaffold hopping refers to the identification of molecules having similar binding modes against a target but dissimilar core chemical structures.

There are several degrees to scaffold hopping[1], going from the simplest cases where atom types in heterocycles are changed - called small-step scaffold hopping - to the hardest situations involving molecules with novel core structures: *large-step scaffold hopping*. Although those cases are uncommon, they are of high interest for the development and evaluation of methods able to handle molecules belonging to remote chemical spaces.

In the specific context where the 3D structure of the protein is not known, only ligand-based methods can be applied. Evaluating the ability of these methods to identify these chemically distant molecules yet able to form the same interactions is of high interest for drug design. For this purpose, we explored the PDB BIND[2] database to search for large-step scaffold hopping situations using protein-ligand interaction fingerprints. We considered the situations where molecules with different scaffolds bind to the same protein, as assessed by both the absence of a common generic Murcko scaffold and a low Tanimoto similarity according to their Morgan fingerprints, and a high Tanimoto similarity according to their interaction fingerprints. Besides, we ensure that no common substructure between the two ligands is responsible for their similar binding modes, to avoid easily explainable cases.

Out of the few hundred situations gathered, we selected a diverse set representing 87 different proteins. For each pair of molecules in a large-step scaffold hopping situation, we selected appropriate decoys extracted from the ZINC[3] database, having similar physical and chemical properties to the ligands. We then conducted similarity searching experiments using traditional ligand-based representations, including 2D and 3D pharmacophore representations, to evaluate their ability to rank with a higher similarity the scaffold hopping pair compared to the decoys.

On this benchmark dataset, we observed that on average, 3D representations outperformed 2D representations, even if conformers are generated instead of the crystallographic conformation. However, in only 15% of the cases the molecule in scaffold hopping was ranked in the top 5% most similar molecules compared to the reference ligand. This demonstrates a large room for improvement for methods aiming at tackling large-step scaffold hopping in the context of ligand-based drug design.

Bibliography:

[1] Sun H.; Tawa G.; Wallqvist A. Classification of scaffold-hopping approaches. Drug Discovery Today. Volume 17, Issues 7–8, 2012, 310-324
[2] Wang R.; Fang X.; Lu Y.; Wang S. The PDBbind Database: Collection of Binding Affinities for Protein-Ligand Complexes with Known Three-Dimensional Structures. J. Med. Chem. 2004, 47, 2977-2980
[3] Irwin J.; Shoichet B. ZINC − A Free Database of Commercially Available Compounds for Virtual Screening. J. Chem. Inf. Model. 2005, 45, 1, 177–182

List of participants

Speakers

Alex TROPSHA	alex_tropsha@unc.edu
Artem CHERKASOV	artc@mail.ubc.ca
Connor COLEY	ccoley@mit.edu
Emilio BENFENATI	emilio.benfenati@marionegri.it
Gisbert SCHNEIDER	gisbert.schneider@sec.ethz.ch
Hanoch SENDEROWICZ	hsenderowitz@gmail.com
Johannes KIRCHMAIR	johannes.kirchmair@univie.ac.at
Jose MEDINA-FRANCO	medinajl@unam.mx
Jürgen BAJORATH	bajorath@bit.uni-bonn.de
Matthias REREY	rarey@zbh.uni-hamburg.de
Nikolaus STIEFL	nikolaus.stiefl@novartis.com
Ola ENGKVIST	Ola.Engkvist@astrazeneca.com
Olivier TABOUREAU	olivier.taboureau@univ-paris-diderot.fr
Pavel POLISHCHUK	pavlo.polishchuk@upol.cz
Thierry LANGER	thierry.langer@univie.ac.at

Participants

Fbelo	jas@fct.unl.pt	Universidade NOVA de Lisboa, Portugal
AKHMETSHIN Tagir	tagir.akhmetshin@etu.unistra.fr	University of Strasbourg, France
ANSBACHER Tamar	tamarans@tauex.tau.ac.il	Tel Aviv University, Hadassah Academic College, Israel
ASGARKHANOVA Farah	farah.asgarkhanova@etu.unistra.fr	University of Strasbourg, France
AUDOUZE Karine	karine.audouze@u-paris.fr	Université Paris Cité, France
BAYBEKOV Shamkhal	shamkhal.baybekov@etu.unistra.fr	University of Strasbourg, France
BERENGER Francois	berenger@edu.k.u-tokyo.ac.jp	The University of Tokyo, Japan
BILOVOLSCHI Bianca	Bianca.bilovolschi@astx.com	Astex Pharmaceuticals, United Kingdom
BONACHERA Fanny	fbonachera@unistra.fr	University of Strasbourg, France
BORT William	william.bort@gmail.com	University of Strasbourg, France
BORUP Rasmus	rmb@chem.ku.dk	University of Copenhagen, Denmark
BOURG Stéphane	stephane.bourg@cnrs.fr	ICOA - CNRS, AR2T, France
BRET Guillaume	gbret@unistra.fr	University of Strasbourg, France
BRIAND Mael	mael.briand@helsinki.fi	University of Helsinki, Finland
BUREAU Ronan	ronan.bureau@unicaen.fr	CERMN, Universite de Caen Normandie, France
BUSHIRI David	bushiri@fmp-berlin.de	Forschungsinstitut für Molekulare Pharmakologie (FMP), Germany
CAMPROUX Anne-Claude	anne-claude.Camproux@u-paris.fr	Université Paris Cité, France
CHEDIK Lisa	lisa.chedik@inrs.fr	INRS, France
CHEN Ya	ya.chen@univie.ac.at	University of Vienna, Austria
CHIESA Luca	luca.chiesa@unistra.fr	University of Strasbourg, France
CLARKE Paul	paul.clarke@etu.unistra.fr	University of Strasbourg, France
CUISSART Bertrand	bertrand.cuissart@unicaen.fr	GREYC - University of Caen Normandy, France
DAPIAGGI Federico	federico.dapiaggi@syngenta.com	Syngenta Crop Protection, Switzerland
DIETRICH Jonas	jonas.dietrich@uni-jena.de	Jonas Dietrich, Sweden
DION Charles	charles.henry.dion@iktos.com	Iktos, France
DJIKIC Théodora	djikicstojsic@unistra.fr	University of Strasbourg, France
DREANO Loïc	loic.dreano@helsinki.fi	University of Helsinki, Finland
EGUIDA KOSSIWA Merveille	keguida@unistra.fr	University of Strasbourg, France
EL KHAOUDI ENYOURY Hocine	hocine.el-khaoudi-enyoury@etu.unistra.fr	University of Strasbourg, France
FAUST Jonas	jonas_faust@outlook.com	Heidelberg University, Germany
GAMBACORTA Nicola	nicola.gambacorta1@uniba.it	University of Bari, Italy
GANTZER Philippe	p.gantzer@icredd.hokudai.ac.jp	University of Strasbourg, France
GASTREICH Marcus	gastreich@biosolveit.de	BioSolveIT GmbH, Germany
GAWALSKA Alicja	alicja.gawalska@doctoral.uj.edu.pl	Jagiellonian University Collegium Medicum, Poland
GESLIN Damien	damien.geslin@unicaen.fr	Centre d'Etudes et de Recherche sur le Médicament de Normandie, France
GIOVANNINI Johanna	johanna.giovannini@unicaen.fr	Centre d'Etudes et de Recherche sur le Médicament de Normandie, France
GRENUT Pierre	pierre.grenut@iktos.com	Iktos, France

GULIN Martin Rafael	martin-rafael.gulin@etu.unistra.fr	University of Strasbourg, France
HANDA Koichi	ko.handa@teijin.co.jp	University of Cambridge, United Kingdom
HOFFMANN Brice	brice.hoffmann@iktos.com	IKTOS, France
HORVATH Dragos	dhorvath@unistra.fr	University of Strasbourg, France
IBRAHIM Hamza	hamzaibrahiem@gmail.com	Universität des Saarlandes, Germany
IWATA Michio	iwata121@bio.kyutech.ac.jp	Kyushu Institute of Technology, Japan
JACQUEMARD Célien	jacquemard@unistra.fr	University of Strasbourg, France
JAMROZIK Marek	marek.jamrozik@doctoral.uj.edu.pl	Jagiellonian University Medical College, Poland
JONCZYK Jakub	jakub.jonczyk@uj.edu.pl	Jagiellonian University Medical College, Poland
KAITOH Kazuma	kaito168@bio.kyutech.ac.jp	Kyushu Institute of Technology, Japan
KELLENBERGER Esther	ekellen@unistra.fr	University of Strasbourg, France
KIUZMIN Viktor	theorchem@gmail.com	A. V. Bogatsky Physico-Chemical Institute, Odessa, Unraine
KLIMCHUK Olga	o.klimchuk@chimie.u-strasbg.fr	University of Strasbourg, France
KURILOV Roman	r.kurilov@dkfz.de	German Cancer Research Center,Heidelberg, Germany
LACOUR Antoine	antoine.lacour@helmholtz-hips.de	Helmholtz Institute for Pharmaceutical Research Saarland, Germany
LAGARDE Nathalie	nathalie.lagarde@lecnam.net	Conservatoire National des Arts et Métiers, France
LAMANNA Giuseppe	giuseppe.lamanna@uniba.it	Istituto di Cristallografia, CNR, Bari, Italy
LEGEHAR Ashenafi	ashenafi.legehar@helsinki.fi	University of Helsinki, Finland
LEHEMBRE Etienne	etienne.lehembre@unicaen.fr	GREYC, Normandy University Caen, France
LEPAILLEUR Alban	alban.lepailleur@unicaen.fr	Université de Caen Normandie, France
LIBOUBAN Pierre-Yves	Pierre-Yves.libouban@univ-orleans.fr	Institut de chimie organique et analytique, France
LLOMPART Pierre	pierre.llompart@etu.unistra.fr	University of Strasbourg, France
MARCOU Gilles	g.marcou@unistra.fr	University of Strasbourg, France
MARTINEZ-MAYORGA Karina	kmtzm@unam.mx	Institute of Chemistry-Merida, UNAM, United States
MIDDLETON James	jmiddleton6@sheffield.ac.uk	University of Sheffield, United Kingdom
MOLZA Anne-Elisabeth	anne-elisabeth.molza@univ-reunion.fr	Universite de La Reunion, Reunion
MONTES Matthieu	matthieu.montes@cnam.fr	Conservatoire National des Arts et Metiers, France
OLENEVA Polina	polina.oleneva@etu.unistra.fr	University of Strasbourg, France
OMRAN Abir	abir.omran@univie.ac.at	University of Vienna, Austria
OUDAHMANE Mehdi	mehdi.oudahmane@hesam.eu	Laboratoire GBCM - Cnam, France
PASQUINI Marta	marta.pasquini@syngenta.com	Syngenta Crop Protection AG, Switzerland
PEREBYINIS Mariana	perebyinis.mar@gmail.com	University of Strasbourg, France
PEREZ-PENA Helena	helena.perez-pena@etu.unistra.fr	University of Strasbourg, France
PETER Sonja	sonja.peter@SoseiHeptares.com	ALLODD (Sosei Heptares), United Kingdom
PETITJEAN Michel	petitjean.chiral@gmail.com	Université Paris Cité, France
PHAM Long-Hung	pdlhung.d18@ump.edu.vn	University of Medicine and Pharmacy at Ho Chi Minh City, Vietnam,
PIKALYOVA Karina	karina.pikalyova@etu.unistra.fr	University of Strasbourg, France

PIKALYOVA Regina	regina.pikalyova@etu.unistra.fr	University of Strasbourg, France
PINEL Philippe	philippe.pinel@iktos.com	Iktos, France
PLYER Louis	louis.plyer@unistra.fr	University of Strasbourg, France
PODLIPNIK Crtomir	crtomir.podlipnik@fkkt.uni-lj.si	University of Ljubljana, Slovenia
RADAEVA Mariia	mradaeva@prostatecentre.com	University of British Columbia, Canada
RAJAONSON Ella	ella.rajaonson@mail.utoronto.ca	University of Toronto, Canada
REHIOUI-KARINE Hajar	hajar.rehioui-karine@unicaen.fr	CoDaG , GREYC, University of Caen Normandy, France
REVILLO-IMBERNON Julia	jrevilloimbernon@unistra.fr	University of Strasbourg, France
RINALDO David	david.rinaldo@schrodinger.com	Schrödinger, Germany
ROGNAN Didier	rognan@unistra.fr	University of Strasbourg, France
RUGARD Marylène	marylene.rugard@gmail.com	T3S Unit - Université Paris Cité, France
SANTHAPURI Sai Prashanth	saisanthapuri@ub.edu	University of Strasbourg, France
SELLAMI Asma	asma.sellami@lecnam.net	Cnam, France
SEUMER Julius	jseumer@chem.ku.dk	University of Copenhagen, Denmark
SHAHANE Ganesh Sanjay	ganesh.shahane@fujitsu.com	Fujitsu UK, United Kingdom
SHERMUKHAMEDOV Shokirbek	2shermux@gmail.com	University of Innsbruck, Austria
SHEVELEV Maxim	mdshev7@gmail.com	University of Strasbourg, France
SIDOROV Pavel	pavel.sidorov@icredd.hokudai.ac.jp	WPI-ICReDD, Hokkaido University, Japan
SIMON Nicolas	nicolas.simon5@etu.unistra.fr	University of Strasbourg, France
SINDT François	francois.sindt@etu.unistra.fr	University of Strasbourg, France
SINDT François	francois.sindt@etu.unistra.fr	University of Strasbourg, France
SOUSA Bruna Katiele de Paula	brunakatiele@discente.ufg.br	Universidade Federal de Goiás, Brazil
STAUB Ruben	ruben.staub@icredd.hokudai.ac.jp	ICReDD (Hokkaido University), Japan
SULPIZIO Giustino	giustino.i.sulpizio@gmail.com	EPFL, Switzerland
THOMPSON Mark	thok@zhaw.ch	Zurich University of Applied Sciences, Switzerland
TROMELIN Anne	anne.tromelin@inrae.fr	INRA-CSGA, France
URVAS Lauri	urvas@unistra.fr	University of Strasbourg, France
VARNEK Alexandre	varnek@unistra.fr	University of Strasbourg, France
VILLACAMPA FERNANDEZ Marina	marina.villacampafernandez@ucdconnect.ie	University College Dublin, Ireland
VOLKOV Mikhail	mvolkov@unistra.fr	University of Strasbourg, France
WALTER Moritz	mwalter4@sheffield.ac.uk	University of Sheffield, United Kingdom
ZABOLOTNA Yuliana	zab.yul94@gmail.com	University of Strasbourg, France
ZANKOV Dmitrii	dmitry.zankov@etu.unistra.fr	University of Strasbourg, France
ZHANG Chonghuan	cz332@cam.ac.uk	University of Cambridge, United Kingdom
ZIBERNA Lara	lara.ziberna@hotmail.com	University of Strasbourg, France


cnrs


Université
de Strasbourg


INVESTIR
L'AVENIR


SFCi


cnrs
GDR
Groupement
de recherche
BigDataChim


Université
franco-allemande
Deutsch-Französische
Hochschule


Collège
doctoral
européen


IKTOS
Artificial Intelligence
for new drug design


BioSolveIT


Celtic
L'eau minérale idéale