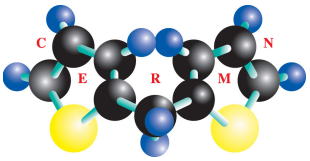
11èmes Journées de la Société Française de Chémoinformatique

5-6 Octobre 2023

Université de Caen Normandie

Table of contents

Sponsors



CODAG GREYC

UNIVERSITÉ CAEN NORMANDIE

ENSI CAEN
ÉCOLE PUBLIQUE D'INGÉNIEURS
CENTRE DE RECHERCHE

Normandie Université

cnrs
dépasser les frontières

cnrs GDR Groupement de recherche
MaDICS Masses de données, informations et connaissances en sciences

cnrs GDR Groupement de recherche
BigDataChim

Normastic

Caenlamer
NORMANDIE
COMMUNAUTÉ URBAINE

RÉGION NORMANDIE

Schrödinger

Qubit
PHARMACEUTICALS

discngine

Mot du Président de la SFCI

Bienvenue à toutes et tous aux 11e journées journées de la SFCi, évenement majeur récurrent de notre société et qui permet tous les 2 ans de rassembler les actrices et acteurs de la chémo-informatique en France du monde académique ou industriel. Ces journées sont l'occasion pour les jeunes chercheuses et chercheurs de notre domaine de présenter leurs travaux et de rencontrer leurs homologues du paysage français. C'est aussi une occasion importante pour les plus anciens d'entre nous de se retrouver avec plaisir et de suivre les avancées respectives de nos équipes.

Merci de votre confiance et de votre fidélité à nos journées et au plaisir de vous retrouver !

Avec toutes mes amitiés,

Matthieu Montes, président de la SFCi

National and Local Committees

National committee

- M. Montes, Président (CNAM, Paris)
- K. Audouze (MetaTox, UMR-S1124)
- P. Roche (CRCM, Université Aix-Marseille)
- S. Bourg (CNRS, Orléans)
- L. Colliandre (Evotec)
- A. Lepailleur (CERMN, Université de Caen Normandie)
- S. Lozano (OncoDesign Services)
- N. Lagarde (CNAM, Paris)
- V. Le Guilloux (DiscNgine)

Local committee

- L. Brun (GREYC, ENSICAEN)
- R. Bureau (CERMN, UNICAEN)
- B. Crémilleux (GREYC, ENSICAEN)
- B. Cuissart (GREYC, UNICAEN)
- B. Gaüzère (LITIS/INSA de Rouen)
- J. Giovannini (CERMN, UNICAEN)
- J.-L Lamotte (GREYC, UNICAEN)
- A. Lepailleur (CERMN, UNICAEN)
- A. Ouali (GREYC, ENSICAEN)
- J. Sopkova - de Oliveira Santos (CERMN, UNICAEN)
- A. Zimmermann (GREYC, UNICAEN)

Program

Thursday morning - joint events

9h15 Atelier DSChem GDR Madics
09h30 Frédéric Payan, SensaaS and MoleculArXiv, or the benefits of interdisciplinarity
10h00 AG GDR BigDataChim
11h00 Coffee break
11h30 AG SFCI

Thursday afternoon

12h30 Lunch buffet

13:30 Opening session

13:45 **Thierry Langer - Next Generation Pharmacophore Modeling: Tools and Applications**

Session: Ligand-based drug design

14:30 Corentin Bedart, Pilot study towards a pan-Canadian virtual chemical library for chemical probe and drug discovery
14:45 Damien Geslin, Analysis of the structure-activity relationships from a pharmacophore space. Application to polypharmacology.
15:00 Etienne Lehembre, How to interactively guide an expert in a pharmacophore structured space
15:15 Regina Pikalyova, Application of molecular cartography to DNA-Encoded Library optimization

15:30 Coffee break — Poster

16:00 **Fabrizio Costa, Heterogeneous networks integration for disease–gene prioritization with node kernels**

Session : Machine learning / Datamining in chemoinformatics

16:45 Youcef Bagdad, Machine learning for the prediction of AmpC β-lactamase inhibition to design new antimicrobial agents.
17:00 Gabriela Falcon-Cano, Development of recursive random forest algorithms to predict the aqueous solubility of new drugs
17:15 Tagir Akhmetshin, Construction of order-independent molecular fragments space with vector quantised graph autoencoder
17:30 Zeineb Si Chaib, Expanding the Domain of Applicability of Structure-Based Drug Design with IFD-M

17:45 Coffee break — Poster

18:15 **Dmitri Kireev, Interaction signatures in early-stage drug discovery**

19:30 Banquet

Friday morning

09:00 **Anne Claude Camproux, Structural analysis of targets for characterization of their binding site and flexibility to improve drug partner predictiony**

Session: Structure-based drug design (1)

09:45 Marina Botnari, hVKORC1 as a multi-faced target for Allo-Network Modulators Design

10:00 François Sindt, Protein-constrained "In Silico" synthesis of ligands from chemical reagent docking poses enables hit identification from ultra-large chemical spaces

10:15 Ahmad Elbahnsi, Efflux mechanism and conformational landscape of P-glycoprotein to better rationalize the prediction of inhibitors involved in drug-drug interactions.

10:30 Coffee break — Poster

11:00 **Dominique Douguet, An overview of approved small-molecule pharmacopeia and what it can bring to drug design**

Session: Structure-based drug design (2)

11:45 Luca Chiesa, Detection of ßadrenergic receptor agonists using machine learning and single-ligand dynamic interaction data

12:00 Xiaojun Mao, Universal Frags2Drugs (uF2D): a proteome-wide in silico fragment-based ligand design tool

12:15 Marc Farag, Structure based pharmacophore and MM-PBSA method: a tool to design small molecules acting as XIAP inhibitors.

12:30 Closing session

12:45 Lunch pack

Invited speakers

Next Generation Pharmacophore Modeling: Tools and Applications

Thierry Langer

Department of Pharmaceutical Sciences, University of Vienna, Vienna, Austria

Pharmacophore-based compound modeling, virtual screening, and bio-activity profiling has become a popular in silico technique for supporting medicinal chemists. The advanced molecular design tool LigandScout [1] has been developed to successfully address one of the most important issues in virtual screening: Enhancing early enrichment while maintaining high computational speed as well as ease of use, as shown by reference studies. [2]
As an extension of the static pharmacophore approach, we lately have focused on incorporating dynamic effects of ligand protein binding into our automated interaction determination process. Our Common Hits Approach (CHA) [3] uses multiple coordinate sets saved during MD simulations. Pharmacophore models with the same pharmacophore features are pooled and virtual screening runs are then performed with every representative pharmacophore model resulting in a consensus hit list. The recently developed GRAIL (GRids of phArmacophore Interaction fieLds) [4] method combines the advantages of traditional grid-based approaches for the identification of interaction sites with the power of the pharmacophore concept: A reduced pharmacophore abstraction of the target system enables the computation of all relevant interaction grid maps in short amounts of time. This allows one to extend the utility of a grid-based method for the analysis of large amounts of coordinate sets obtained by long-time MD simulations. In this way it is possible to assess conformation dependent characteristics of key interactions over time.
In the NeuroDeRisk project [5], we utilize these new developments, together with machine learning methods for adding quantitative pharmacophore feature weighting [6] to predict potential neurotoxic effects of drug candidates.

References :
1. Wolber G, Langer T, J Chem Inf Model. 2005; 45(1): 160
2. Karaboga AS, et al., J Chem Inf Model. 2013; 53(3):1043
3. Wieder M, et al., J Chem Inf Model. 2017; 57(2):365
4. Schuetz DA, et al., J Chem Theory Comput. 2018; 14: 4958
5. NeuroDeRisk IMI2 JU (www.neuroderisk.eu) has received funding under grant agreement No 821528
5. Kohlbacher SM, et al., J Cheminform 2021; 13, 57

Heterogeneous networks integration for disease–gene prioritization with node kernels

Fabrizio Costa

Department of Computer Science, Faculty of Environment, Science and Economy, University of Exeter, United Kingdom

The identification of disease–gene associations is a task of fundamental importance in human health research. A typical approach consists in first encoding large gene/protein relational datasets as networks due to the natural and intuitive property of graphs for representing objects' relationships and then utilizing graph-based techniques to prioritize genes for successive low-throughput validation assays. Since different types of interactions between genes yield distinct gene networks, there is the need to integrate different heterogeneous sources to improve the reliability of prioritization systems. In this lecture we will propose an approach based on three phases: first, we will merge all sources in a single network, then we will partition the integrated network according to edge density introducing a notion of edge type to distinguish the parts and finally, we will employ a novel node kernel suitable for graphs with typed edges. We will show how the node kernel can generate a large number of discriminative features that can be efficiently processed by linear regularized machine learning classifiers. Finally we will validate the accuracy of the proposed approach on 12 disease–gene associations and on a time-stamped benchmark containing 42 newly discovered associations.

Interaction signatures in early-stage drug discovery

Dmitri Kireev

Department of Chemistry, University of Missouri, Columbia, Missouri, USA

Docking-based screening of billion-compound libraries is a promising avenue for drug discovery. Though the new, larger scale comes with new caveats. In particular, the number of well-scored ligands may attain millions. This underscores the importance of hit triage tools able to increase the true positive rates to levels sufficient for hit identification from small sets of procurable compounds. And since fast and accurate calculations of binding affinity remain a distant future, we need to rely on empirical filters. Recognition of true binders through their interaction signatures with the target protein is a promising avenue. Indeed, if a ligand docked to the protein forms a network of interactions that is reminiscent of interactions made by true binders, it is also more likely to be a true binder. Previously, we developed structural protein-ligand fingerprints (SPLIF) and SPLIF-based similarity score to retain ligands with hit-like interaction signatures. SPLIF-score proved to be an effective method in hit-finding practice and benchmark studies by us and others and the fingerprints are available through a popular DeepChem resource. We now introduce DeepSPLIF, a graph-convolutional neural network (GCNN) trained on a large dataset of x-ray structures of protein-ligand complexes represented as interaction graphs to predict the likelihood for a ligand to be a true hit. The GCNN output is a fixed-size interaction signature that can be readily used for similarity/diversity analyses and pose ranking. Interaction signatures can also be useful in computer-aided fragment-based discovery. For instance, we developed FRASE-bot, a computational platform enabling de novo construction of small-molecule ligands directly in the binding pocket of the target protein. It makes use of the concept of FRAgments in Structural Environments (FRASE) and machine learning to distill 3D information relevant to the protein of interest from thousands of 3D protein-ligand complexes in the Protein Data Bank (PDB). Unlike docking that needs the user to specify the binding site, FRASE-bot analyses the whole protein structure and seeds it with ligand fragments where they would have the most native-like interaction signatures. In the talk, we discuss the past and current developments involving the concept of interaction signatures, as well as their successful applications in drug discovery.

Structural analysis of targets for characterization of their binding site and flexibility to improve drug partner prediction

Anne-Claude Camproux

Université Paris Cité Modélisation Computationnelle des Interactions Protéine-Ligand, Unité de Biologie Fonctionnelle et Adaptative - CNRS UMR 8251, Université Paris Cité, Nanterre, France

An overview of approved small-molecule pharmacopeia and what it can bring to drug design

Dominique Douguet

Institut de Pharmacologie Moléculaire et Cellulaire, CNRS UMR7275, Université Côte d'Azur, Valbonne, France

Approved drugs are the most widely studied small molecules for their function and effects on humans as well as for what the body's physiology does to the molecules. They are a rich source of information to get insight into which properties are required for a molecule to be an administered drug and to optimize the search for new therapeutics. The e-Drug3D database is a chemistry-oriented collection aimed to perform retrospective analyses of past successes and to help in various endeavors such as drug repurposing, drug design, SAR analyses or improving PK/PD predictive models.

Session: Ligand-based drug design

Pilot study towards a pan-Canadian virtual chemical library for chemical probe and drug discovery

Corentin Bedart *† [1], Conrad V. Simoben [1], Grace Shimokura [2], Seth Vigneron [3], Tabitha E. H. Wood [4], Frederick G. West [5], Robert A. Batey [2], John J. Irwin [3], Matthieu Schapira‡ [1,6]

[1] Structural Genomics Consortium, University of Toronto, Toronto, ON M5G 1L7 – Canada
[2] Department of Chemistry, University of Toronto, Toronto, ON M5S 3H6, Canada
[3] Department of Pharmaceutical Chemistry, University of California, San Francisco, CA, United States
[4] Department of Chemistry, University of Winnipeg, Winnipeg, MB R3B 2G3, Canada
[5] Department of Chemistry, University of Alberta, Edmonton, AB T6G 2G2, Canada
[6] Department of Pharmacology and Toxicology, University of Toronto, Toronto, ON M5S 1A8, Canada

Early-stage drug discovery is a complex and costly process that requires the identification of active molecules for a targeted protein. Traditional high-throughput screening methods are incompatible with academic chemistry due to the need for millions of samples and a dedicated robotic system. Over the last 20 years, virtual screening has proven to be a feasible and effective alternative to high-throughput screening and benefits from recent advancements in computational power and deep learning methods (1). As this method requires virtual libraries of compounds, large combinatorial databases of several billion compounds are now available, based on well-known chemical reactions (2).

To explore uncharted chemical spaces using innovative chemistry that could have a significant impact on drug discovery, we propose the Pan-Canadian Chemical Library which aims to develop a free-to-use informatics platform to allow chemists across Canada to enter their synthetic routes to create custom combinatorial libraries for virtual screening.

As part of a pilot study to evaluate the feasibility and effectiveness of the project, we worked in close collaboration with the chemistry groups of Pr. Batey (University of Toronto), Pr. Wood (University of Winnipeg), and Pr. West (University of Alberta), and identified five unique and innovative chemical reactions. The chemical reactions were encoded in SMARTS format, which allows the necessary substructural patterns to be transcribed into a computer-readable language. Inclusion and exclusion rules were defined for each reagent, optimized by the involvement of chemists, and the compatible reagents were obtained from the commercially available building blocks catalog provided by ZINC20, the world's largest repository of commercially available compounds (3).

The chemical reactions allowed the creation of bespoke libraries of over 115 billion compounds, including 179 million that we estimate can be synthesized quickly at low cost. An online interface was developed to allow users to browse and download the first version of the Pan-Canadian Chemical Library: https://pccl.thesgc.org. This library will be used as a basis for virtual screening of in-house targets of interest and will be provided for the international hit-finding competition CACHE organized by the Structural Genomics Consortium of Toronto where molecules

*Speaker
†Corresponding author: corentin.bedart@gmail.com
‡Corresponding author: matthieu.schapira@utoronto.ca

selected are tested experimentally (4).

(1) Baum ZJ, Yu X, Ayala PY, Zhao Y, Watkins SP, Zhou Q. Artificial Intelligence in chemistry: Current trends and future directions. Journal of Chemical Information and Modeling. 2021;61(7):3197–212.

(2) Hoffmann T, Gastreich M. The Next Level in Chemical Space Navigation: Going far beyond enumerable compound libraries. Drug Discovery Today. 2019;24(5):1148–56.

(3) Irwin JJ, Tang KG, Young J, Dandarchuluun C, Wong BR, Khurelbaatar M, et al. Zinc20-a free ultralarge-scale chemical database for Ligand Discovery. Journal of Chemical Information and Modeling. 2020;60(12):6065–73.

(4) Ackloo S, Al-awar R, Amaro RE, Arrowsmith CH, Azevedo H, Batey RA, et al. Cache (critical assessment of computational hit-finding experiments): A public–private partnership benchmarking initiative to enable the development of computational methods for hit-finding. Nature Reviews Chemistry. 2022;6(4):287–95.

Analysis of the structure-activity relationships from a pharmacophore space. Application to polypharmacology.

Damien Geslin * [1], Ronan Bureau [1], Peggy Suzanne [1], Christophe Rochais [1], Alban Lepailleur [1]

[1] Centre d'Études et de Recherche sur le Médicament de Normandie UR4258 – Université de Caen Normandie – France

Melatonin (N-acetyl-5-methoxytryptamine) is a neurohormone known for its chronobiotic and sleep-inducing properties(1). Its has been suggested that administration of melatonin could have positive effect on a fast range of health issues, including depression(2) and Alzheimer's disease(3). The antidepressant properties of agomelatine appear to result from a synergistic ation between its agonism at the MT1 and MT2 melatonin receptors and its antagonism at the 5-HT2C serotonin receptor. The aim of this study was to investigate the structure-activity relationships (SAR) of ligands for the three receptors (ChEMBL data) in order to understand their potential multi-target activities and to propose new agomelatine analogues. Thanks to our previous works(4), we were able to create a pharmacophore space based on graph edit distance (GED)(5). Analysis of this space enabled us to define polypharmacological areas for which there is a close distance between the pharmacophores associated with the three receptors. A focus was made on the corresponding families of compounds.

References

(1) Pévet, P.; Bothorel, B.; Slotten, H.; Saboureau, M. The Chronobiotic Properties of Melatonin. *Cell Tissue Res* **2002**, *309* (1), 183–191. https://doi.org/10.1007/s00441-002-0584-1.

(2) Satyanarayanan, S. K.; Su, H.; Lin, Y.-W.; Su, K.-P. Circadian Rhythm and Melatonin in the Treatment of Depression. *Current Pharmaceutical Design 24* (22), 2549–2555.

(3) Shukla, M.; Govitrapong, P.; Boontem, P.; Reiter, R. J.; Satayavivad, J. Mechanisms of Melatonin in Alleviating Alzheimer's Disease. *Current Neuropharmacology 15* (7), 1010–1031.

(4) Geslin, D.; Lepailleur, A.; Manguin, J.-L.; Vo, N.-V.; Lamotte, J.-L.; Cuissart, B.; Bureau, R. Deciphering a Pharmacophore Network: A Case Study Using BCR-ABL Data. *J. Chem. Inf. Model.* **2022**, *62* (3), 678–691. https://doi.org/10.1021/acs.jcim.1c00427.

(5) Gao, X.; Xiao, B.; Tao, D.; Li, X. A Survey of Graph Edit Distance. *Pattern Anal Applic* **2010**, *13* (1), 113–129. https://doi.org/10.1007/s10044-008-0141-y.

*Speaker

How to interactively guide an expert in a pharmacophore structured space

Etienne Lehembre * [1], Bruno Crémilleux [1], Bertrand Cuissart [1], Abdelkader Ouali [1], Albrecht Zimmermann [1], Alban Lepailleur [2], Ronan Bureau [2]

[1] Equipe CODAG - Laboratoire GREYC - UMR6072 – Groupe de Recherche en Informatique, Image et Instrumentation de Caen – France
[2] Centre d'Études et de Recherche sur le Médicament de Normandie UR4258 — Université de Caen Normandie – France

In medicinal chemistry, a pharmacophore denominates a spatial arrangement of chemical features responsible for a favorable interaction or an unfavorable 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. We enhance knowledge on the pharmacophore network organization by taking into account parents-children relations and/or grouping the pharmacophores into equivalence classes, *i.e.* sets of pharmacophores occurring in the exact same molecule group. The result of this construction is a Partial Order Graph (POG) where each vertex is a set of pharmacophores.

The current work aims at supporting a medicinal chemist in his pharmacophore analyses by interactively guiding him to the most topical pharmacophores of interest. The method is designed to be used in unexplored and unlabeled data-sets using few interactions to highlight patterns of interest, or their opposite, from the rest of the pattern space. For this purpose, with a view towards Interactive Pattern Mining (3), the method leverages the interactions of an expert regarding the POG to enhance the structure information with the expert's interest. The interaction process is composed of iterations of sampling and labeling. Each iteration diffuses the subjective interest of the expert in the search space, modifying the potential interest of its lineage and their linked patterns. The potential interest can then be exploited in the following samplings, using the structured pharmacophores space to draw out pharmacophores of interest for the expert. At the end of the process, the POG structure not only contains the labels assigned by the expert but also a weighted expression of the expert's interest. These weights complete standard quality measures as growth rate, or replace standard methods using feature vectors.

(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, Deciphering a Pharmacophore Network: A Case Study Using BCR-ABL Data. J. Chem. Inf. Model. 2022,62, 678–691.

*Speaker

(3) Van Leeuwen, Matthijs. "Interactive data exploration using pattern mining." Interactive knowledge discovery and data mining in biomedical informatics. Springer, Berlin, Heidelberg, 2014. 169-182.

Application of molecular cartography to DNA-Encoded Library optimization

Regina Pikalyova * [1], Yuliana Zabolotna [1], Dragos Horvath [1], Gilles Marcou [1], Alexandre Varnek† [1]

[1] Université de Strasbourg – Laboratoire de chémoinformatique – France

The development of DNA-Encoded Library (DEL) technology introduced new challenges for the analysis of chemical libraries. Being synthesized using combinatorial split-and-pool approach, a DEL is usually a multi-million compound library that is stored and screened as inseparable mixture. In this context, it is useful to consider a chemical library as a stand-alone chemoinformatic object – represented both as a collection of independent molecules, and yet an individual entity – in particular when they are inseparable mixtures, like DELs. Herein we introduce the concept of Chemical Library Space (CLS) in which resident items are individual chemical libraries(1). We define and compare four vectorial library representations obtained using Generative Topographic Mapping (GTM). These allow effective comparison of thousands of compound collections in terms of structural and property similarity, with the ability to tune and chemically interpret the similarity relationships. We apply various CLS encodings for the selection problem of DELs that optimally "match" a reference collection (here ChEMBL28), showing how the choice of the CLS descriptors may help to fine-tune the "matching" (overlap) criteria. Hence, the proposed CLS may represent a new efficient way for polyvalent analysis of thousands of chemical libraries. For instance, using this approach an easily accessible compound collection may be selected as a substitute of a difficult to produce "ideal" library for primary- or target-focused screening, or libraries covering novel regions of the chemical space may be chosen for library portfolio enrichment. Furthermore, we explore the cartography of the CLS using meta-(μ)GTM, where libraries themselves are mapped objects(2). Visualization on μGTM landscapes provides a bird's eye view of the CLS and comprehensive insights into inter-library similarity and intrinsic library characteristics.

(1) Pikalyova R, Zabolotna Y, Horvath D, Marcou G, Varnek A. The chemical library space and its application to DNA-Encoded Libraries. ChemRxiv. Cambridge: Cambridge Open Engage; 2023; 10.26434/chemrxiv-2023-w7x9t (JCIM accepted)

(2) Pikalyova R, Zabolotna Y, Horvath D, Marcou G, Varnek A. MetaGTM: VISUALIZATION AND ANALYSIS OF THE CHEMICAL LIBRARY SPACE. ChemRxiv. Cambridge: Cambridge Open Engage; 2023; 10.26434/chemrxiv-2023-14d8j

*Speaker

†Corresponding author: varnek@unistra.fr

Session: Machine learning / Datamining in chemoinformatics

Machine learning for the prediction of AmpC β-lactamase inhibition to design new antimicrobial agents.

Youcef Bagdad * [1], Marion Sisquellas [1,2], Maria Miteva [1]

[1] Cibles thérapeutiques et conception de médicaments – Institut de Chimie du CNRS, Centre National de la Recherche Scientifique, Université Paris Cité, Centre de Recherche Inserm – France
[2] Institut Cochin – Institut National de la Santé et de la Recherche Médicale, Centre National de la Recherche Scientifique, Université Paris Cité – France

Antimicrobial resistance is a major problem that has been growing steadily in recent years, causing millions of deaths (a). The emergence of multi-drug resistance (MDR) is particularly found among Entero-bacteriaceae such as Escherichia coli (E. coli). E.coli causes serious infections and have multiple resistance mechanisms, the most common being extended-spectrum β-lactamase (ESBL) and AmpC β-lactamase production (b)(c). One of the main mechanisms underlying resistance to β-lactam antibiotics are the AmpC β-lactamases (d). In this study, we employ in silico approaches to identify new inhibitors of AmpC β-lactamase. First, we collected 384,223 compounds experimentally tested on E.coli AmpC. The curation of these data has led to 191 inhibitors and 81720 non-inhibitors of AmpC β-lactamase. We used these compounds to develop new classification machine learning (ML) models to predict putative inhibitors of this enzyme. The ML models that we have developed can be useful for the identification of new candidates to fight against the MDR. Moreover, the analysis of the known inhibitors allowed to shed new light on the mechanisms of action of AmpC inhibitors.

References :
(a) Antimicrobial Resistance Collaborators, " Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis ", Lancet Lond. Engl., vol. 399, no 10325, p. 629 655, févr. 2022, doi: 10.1016/S0140-6736(21)02724-0.
(b) T. D. Institute, " WHO global priority pathogens list of antibiotic-resistant bacteria ", 2 décembre 2021. https://www.doherty.edu.au/news-events/news/who-global-priority-pathogens-list-of-antibiotic-resistant-bacteria/.
(c) C. A. Michael, D. Dominey-Howes, et M. Labbate, " The Antimicrobial Resistance Crisis: Causes, Consequences, and Management ", Front. Public Health, vol. 2, 2014, https://www.frontiersin.org/articles/10.3389/fpubh.2014.00145
(d) S. Lee et al., " Prevalence, Characteristics and Clonal Distribution of Extended-Spectrum β-Lactamase- and AmpC β-Lactamase-Producing Escherichia coli Following the Swine Production Stages, and Potential Risks to Humans ", Front. Microbiol., vol. 12, https://www.frontiersin.org/articles/10.3389/fmicb.2021.710747

*Speaker

Development of recursive random forest algorithms to predict the aqueous solubility of new drugs

Gabriela Falcon-Cano *† [1], Christophe Molina [1], Miguel-Angel Cabrera-Perez [2]

[1] PIKAÏROS, S.A. – 31650 Saint-Orens-de-Gameville, France
[2] Departamento de Ciencias Farmacéuticas, Facultad de Ciencias, Universidad Católica del Norte, Angamos, Antofagasta 0610, Chile

Computational prediction of aqueous solubility plays a critical role in the drug design and development process. Traditionally, the limited performance of *in silico* solubility models has been attributed to the lack of high-quality solubility data for drugs and drug-like molecules. However, recent studies suggest that the poor reliability of solubility models is not related to the quality of experimental data and that more accurate methods are needed to predict the aqueous solubility of new molecules. In this sense, the present study aims to develop a computational model based on novel modelling techniques to predict the aqueous solubility of new molecules. All steps of the modelling methodology were automated in the KNIME Analytical Platform. Recommended best practices for chemical and experimental data curation were followed to obtain a structurally diverse aqueous solubility database. Two new recursive random forest approaches were developed for data cleaning and variable selection. A consensus model based on regression and classification tree algorithms was implemented to correlate the molecular descriptors with the solubility data. The model was successfully validated by cross-validation and using an external validation set. The statistics were comparable to or better than previously published models. In addition, the model was validated against two external sets from the second 'Solubility Challenge', with results comparable to the top-ranked models from the challenge. The KNIME workflow is freely available online to ensure the reproducibility of the model and its use as a tool for the preliminary assessment of the aqueous solubility of new molecules.

*Speaker
†Corresponding author: gabriela.falcon@pikairos.com

Construction of order-independent molecular fragments space with vector quantised graph autoencoder

Tagir Akhmetshin * [1], Arkadii Lin [1], Timur Madzhidov† [2], Alexandre Varnek‡ [1]

[1] Laboratoire de Chémoinformatique – Université de Strasbourg, CNRS – France
[2] Chemistry Solutions, Elsevier Ltd – United Kingdom

Autoencoders represent a promising technique for the inverse quantitative structure-activity relationship (QSAR) task. However, undesirable bias, such as atom ordering, affects the neighbourhood behaviour of autoencoders' latent space and, consequently, usage of the latent vectors as variables in machine-learning models. Here, we present a graph-based autoencoder which implements vector quantisation operation (VQGAE). The latter allows to learn vectorial representation of molecular fragments in an unsupervised manner. The latent vectors or fragment count vectors of VQGAE are permutation invariant and perform well in similarity ranking. In QSAR benchmarks, the VQGAE's latent vectors outperform those derived by some earlier developed SMILES-based and graph-based autoencoders. Finally, VQGAE autoencoder was used in the inverse QSAR task based on genetic algorithm and machine learning models.

*Speaker
†Corresponding author: tmadzhidov@gmail.com
‡Corresponding author: varnek@unistra.fr

Expanding the Domain of Applicability of Structure-Based Drug Design with IFD-MD

Zeineb Si Chaib * [1]

[1] Senior scientist, Application science – Schrödinger GmbH

Many structure-based drug design (SBDD) methods, including free energy perturbation (FEP+), require accurate, atomic-level detail of the target protein in complex with a member of the ligand series being modeled to perform optimally. Experimental methods such as X-ray crystallography or cryo-electron microscopy (cryo-EM) can provide starting points for such predictions at or near an atomic level of resolution. However, the cost to experimentally obtain structures with new ligands ranges from trivial to extremely large.

Here, we present IFD-MD which is an integral workflow that can predict the atomic details of the structure needed for SBDD of the desired ligand series starting from a structure of the target protein with a very different ligand in the binding site or even starting from a structure of a highly homologous protein. IFD-MD uses a combination of docking algorithms, water thermodynamics, empirical scoring functions, implicit solvent force field energies and explicit solvent metadynamics trajectories to explore the motions of the target protein and simultaneously determine the relative energetics of them.

We will then see retrospective applications of IFD-MD to predicting the binding modes of co-crystal structures starting from highly dissimilar ligand complex structures or from homology models. We will also present some early examples of the impact of IFD-MD on real-world drug discovery programs and how it can significantly increase the impact of structure-based drug design.

*Speaker

Session: Structure-based drug design (1)

hVKORC1 as a multi-faced target for Allo-Network Modulators Design

Marina Botnari * [1], Julie Ledoux† [1], Luba Tchertanov‡ [1]

[1] CB - Centre Borelli - UMR 9010 – Université Paris-Saclay, Centre National de la Recherche Scientifique, Ecole Normale Supérieure Paris-Saclay – France

Human Vitamin K Epoxide Reductase (hVKORC1), a key enzyme transforming vitamin K into the form necessary for blood coagulation, requires for its activation the reducing equivalents delivered by its redox partner via thiol-disulphide exchange reactions. The protein disulphide isomerase PDI was predicted by in silico methods as the most probable hVKORC1 partner, and a PDI/hVKORC1 3D model was proposed as the first precursor complex. These results were then validated in vitro. The principal actor of hVKORC1 activation is the luminal loop (L-loop), a region often carrying numerous missense mutations.

We studied four hVKORC1 mutants with the missense mutations located on L-loop – A41S and H68Y, suggested to be involved in the resistance observed in patients, and S52W and W59R, leading to an almost complete loss of hVKORC1 activity. All proteins were studied in the fully oxidized state by in silico techniques (3D modelling and molecular dynamics (MD) simulation) to identify the mutation-induced effects on hVKORC1 inherent structural and dynamical properties.

Similar to hVKORC1WT, L-loop in mutants possesses a helical fold composed of reversible transient α- and 310-helices linked by flexible coils. These properties classify L-loop as the intrinsically disordered region (IDR) possessing great structural and conformational diversity, varying from compact globule- like (closed) to extended (open) conformations. Nevertheless, we found that L-loop fold and plasticity are mutant-specific. The conformational ensemble of disordered L-loop was described in terms of the Gibbs free energy, a more plausible approach than conventional clustering, non-adapted for IDR analysis. We found that L-loop mutations also affect transmembrane domain (TMD) helices folding and dynamics: (i) induce collective highly correlated motions of the L-loop and adjacent TMD region, and (ii) alternate locally and globally cross-correlations patterns.

The pocket search in hVKORC1WT identified, along the well-known active site pocket, a novel large allosteric pocket on L-loop that may be a potential binding site for modulators regulating L-loop plasticity. The influence of hVKORC1 mutations on the pockets appearance, their localisation and size was observed in all mutants studied.

Do mutations influence the PDI recognition by hVKORC1? To study this aspect, we designed for each mutant a precursor complex by two approaches: (i) a protein-protein automatic docking and (ii) the homology modelling using PDI/hVKORC1WT as a template. The resulting models were studied by MD simulation and characterised in terms of the binding free energy and interaction interface arranging. These descriptors delivered another therapeutic target–the interaction interface formed during the recognition/binding of PDI to L-loop. The modulators of these interactions may regulate or prevent aberrant hVKORC1 activation.

*Speaker

†Corresponding author: julie.ledoux@ens-paris-saclay.fr

‡Corresponding author: luba.tchertanov@ens-paris-saclay.fr

These two target types, allosteric sites and interaction interfaces, have been classically treated as largely distinct approaches. A new vision of drug design combines intra-protein allosteric regulation and inter-protein network control, dubbed 'allo-network drugs'. We postulated for the first time (i) the intra-protein allosteric sites in hVKORC1 and its four mutants, and (ii) the PDI/L-loop interaction interface as the targets opening novel perspectives for the development of 'allo-network drugs' required in treatment of blood disorders.

Protein-constrained "In Silico" synthesis of ligands from chemical reagent docking poses enables hit identification from ultra-large chemical spaces

François Sindt * [1], Anthony Seyller [1], Didier Rognan† [1]

[1] Laboratoire d'Innovation Thérapeutique – Université de Strasbourg, Institut de Chimie du CNRS, Centre National de la Recherche Scientifique, Centre National de la Recherche Scientifique : UMR7200 – France

The chemically accessible space for synthesis is growing exponentially year after year. With the advent of "on-demand" chemical libraries, suppliers can now offer several billion of readly-accessible molecules with a high synthesis success rate (e.g., Enamine REAL Space, 36 billion). With the continuous expansion of this chemical space, novel methods need to be developed to navigate in ultra-large chemical libraries while considering target 3D constraints.

We herewith introduce a novel method (SpaceDock) relying first on the accurate docking of REAL space chemical reagents and then on a one-step or two-steps in silico synthesis of full ligands according to topological and chemical reactivity rules, affording the identification of potential hits from billion-sized chemical spaces with limited resources and computing time. The method scales non linearly with the size of the chemical space and permits to construct up to 6 billion easily synthesizable compounds from 140 000 chemical reagents and 36 organic chemistry rules.

The above-described approach was applied to the fast structure-based identification of ligands for two receptors (estrogen receptor beta, dopamine d3 receptor) and was able to recover both known ligands (or very close analogs) as well as propose totally novel chemotypes. 31 novel hits, directly accessible in REAL space, have been synthesized and tested for their ability to the bind to the two target receptors.

Bibliography :
(1) Hartenfeller M, Eberle M, Meier P, Nieto-Oberhuber C, Altmann KH, Schneider G, Jacoby E, and Renner S. A Collection of Robust Organic Synthesis Reactions for In Silico Molecule Design. J Chem Inf Model, **2011**, *51* (12), 3093-3098.

*Speaker
†Corresponding author: rognan@unistra.fr

Efflux mechanism and conformational landscape of P-glycoprotein to better rationalize the prediction of inhibitors involved in drug-drug interactions.

Ahmad Elbahnsi * [1], Bàlint Dudàs [1], Xavier Decleves [2], Salvatore Cisternino [2], Maria Miteva [1]

[1] Inserm U1268 MCTR, CiTCoM UMR 8038 CNRS – Université Paris Cité, Institut National de la Santé et de la Recherche Médicale - INSERM, Centre National de la Recherche Scientifique - CNRS – France

[2] Inserm UMRS 1144, Optimisation thérapeutique en Neuropsychopharmacologie – Université Paris Cité, Institut National de la Santé et de la Recherche Médicale - INSERM – France

P-glycoprotein (P-gp) is one of the most studied ABC transporters which plays a key role in cancer multidrug resistance and drug-drug interactions (Fletcher et al., 2010). It represents thus an important target in drug discovery, with challenging issues related to the development of new specific inhibitors in cancer treatment. In fact, P-gp mediates the export of chemically and structurally diverse compounds, including drugs and lipids, across the plasma membrane (Aller et al., 2009; Van Helvoort et al., 1996). To ensure this process, the energy provided by ATP binding and hydrolysis permits transitions between open states (inward-facing, IF) allowing compound binding; and closed states (outward-facing, OF) allowing its eviction from the cell.

The IF and OF 3D structures of human P-gP were solved by cryo-EM (Kim and Chen, 2018; Nosol et al., 2020). However, these structures give only a partial information and do not elucidate the efflux mechanism involving large conformational changes in order to expulse very diverse drug substrates.

Recently, we developed an original enhanced sampling method for molecular dynamics (MD), namely kinetically excited targeted MD, that allowed us to reveal the transitions between the IF and OF states and the translocation pathway in BCRP, another important ABC transporter (Dudas et al., 2022). Here, we optimized this approach to generate transitory conformations along the gating cycle of P-gp and to unveil the mechanisms of substrate efflux and inhibitor interactions. Our simulations provided for the first time exploration of the P-gp transition pathway, and the studied conformational landscape revealed some crucial features about its functions and dynamics. Such data are subsequently useful to i) better define the binding sites, ii) characterize their interaction modes with known active molecules and iii) rationalize the prediction of new inhibitors and substrates.

References:

Aller et al., 2009. Structure of P-Glycoprotein Reveals a Molecular Basis for Poly-Specific Drug Binding. Science 323, 1718–1722. https://doi.org/10.1126/science.1168750

Dudas et al., 2022. ABCG2/BCRP transport mechanism revealed through kinetically excited targeted molecular dynamics simulations. Comput. Struct. Biotechnol. J. 20, 4195–4205.

*Speaker

https://doi.org/10.1016/j.csbj.2022.07.035

Fletcher et al., 2010. ABC transporters in cancer: more than just drug efflux pumps. Nat. Rev. Cancer 10, 147–156. https://doi.org/10.1038/nrc2789

Kim & Chen, 2018. Molecular structure of human P-glycoprotein in the ATP-bound, outward-facing conformation. Science 359, 915–919. https://doi.org/10.1126/science.aar7389

Nosol et al., 2020. Cryo-EM structures reveal distinct mechanisms of inhibition of the human multidrug transporter ABCB1. Proc. Natl. Acad. Sci. 117, 26245–26253. https://doi.org/10.1073/pnas.2010264117

Van Helvoort et al., 1996. MDR1 P-Glycoprotein Is a Lipid Translocase of Broad Specificity, While MDR3 P-Glycoprotein Specifically Translocates Phosphatidylcholine. Cell 87, 507–517. https://doi.org/10.1016/S0092-8674(00)81370-7

Session: Structure-based drug design (2)

Detection of β2 adrenergic receptor agonists using machine learning and single-ligand dynamic interaction data

Luca Chiesa * [1], Esther Kellenberger† [1]

[1] Laboratoire d'Innovation Thérapeutique – Université de Strasbourg, Institut de Chimie du CNRS, Centre National de la Recherche Scientifique, UMR7200 – France

G-protein coupled receptors (GPCRs) are membrane proteins responsible for signaling transduction in many biological processes. Signaling is regulated by the binding of a ligand in the orthosteric binding pocket. Agonists bind the receptor and trigger signaling, while antagonists and inverse agonist inhibit it. The β2 adrenergic receptor (ADRB2) is one of the most relevant targets for drug discovery in the GPCR family. β2-agonists have been used to treat respiratory diseases, like asthma and COPD, while the binding of antagonist is generally linked to off-target effects.

Binding information from crystallographic structures have been used to rescore or filter the results of protein-ligand docking 1. Combining experimental information with in-silico calculations allowed to distinguish active from inactive molecules, and agonists from antagonists. Such approach, while successful, is limited by the static nature of crystallographic data 2. We propose a new method based on the dynamic interactions between the receptor and a reference ligand 3, to improve the search towards ligands with a specific pharmacological profile.

Molecular dynamics (MD) simulations were used to study the interactions formed by a reference ligand with ADRB2. Relevant binding information were extracted using machine learning, specifically one class classification. The models were trained to recognize binding patterns comparable to the ones observed in the simulations, while discarding anomalous ones. The method was tested on small datasets containing well characterized molecules with different pharmacological profiles. Docking was performed on an ensemble of structures extracted from the simulations. Each docking pose was classified by the models as reference-like or not. Properly trained models were able to clearly distinguish between agonist, and antagonists and non-binders.

The proposed technique is applicable to other targets, especially other GPCRs, to post-process docking results. The trained models select only ligands presenting a binding mode comparable to the reference ligand, thus discarding ligands without the desired pharmacological profile.

References

1 de Graaf C, Rognan D. Selective Structure-Based Virtual Screening for Full and Partial Agonists of the β2 Adrenergic Receptor, J Med Chem., 2008, 51-4978.
2 Kooistra AJ, Leurs R, de Esch IJP, de Graaf C. Structure-Based Prediction of G-Protein-Coupled Receptor Ligand Function: A β-Adrenoceptor Case Study, J Chem Inf Model, 2015, 55-1045.
3 Chiesa L, Kellenberger E, One class classification for the detection of β2 adrenergic receptor agonists using single-ligand dynamic interaction data, Journal of Cheminformatics, 2022, 14-74

*Speaker
†Corresponding author: ekellen@unistra.fr

Universal Frags2Drugs (uF2D): a proteome-wide in silico fragment-based ligand design tool

Xiaojun Mao* [1], Samdani Ansar[1], Mahest Velusamy[1], Samia Aci-Sèche[1], Pascal Bonnet† [1]

[1] Institut de Chimie Organique et Analytique – Université d'Orléans, Centre National de la Recherche Scientifique – France

Computational methods are extensively employed during the initial stages of drug discovery process to reduce its duration and cost.(1) Fragment-based drug design (FBDD) prioritizes binding efficiency over global binding affinity.
In our previous work, we have developed Frags2Drugs (F2D), an innovative *in silico* FBDD tool aiming to design novel protein kinase inhibitors.(2) However, its applicability domain is limited solely to protein kinases, restricting its widespread adoption. Furthermore, F2D is based on the alignment of protein kinases, rendering it impractical for proteins belonging to distinct families.

Herein, we present Universal Frags2Drugs (uF2D), an enhanced iteration of F2D, which expands its applicability domain to encompass all protein families. The uF2D methodology comprises two distinct components: a fragment repositioning tool and a linker generator. The initial fragment repositioning phase is accomplished through the adaptation and customization of a pre-existing tool, referred to as CrystalDock.(3)

On a limited dataset consisting of 314 representative protein-ligand complexes, CrystalDock demonstrates the ability to successfully reposition the entirety of ligand fragments in approximately 90% of the cases.

We are currently updating the fragment database with CrystalDock, and evaluating multiple deep learning-based linker generators to incorporate into uF2D.

Keywords: Fragment-based drug design, Protein kinase inhibitor, Hit identification, Fragment linking

(1) A. V. Sadybekov, V. Katritch, Nature 2023, 616, 673–685.
(2) G. Peyrat, C. Bournez, P. Krezel, J.-M. Gally, S. Bourg, S. Aci-Sèche, P. Bonnet, Journal of Medicinal Chemistry, submitted
(3) J. D. Durrant, A. J Keywords. Friedman, J. A. McCammon, J. Chem. Inf. Model. 2011, 51, 2573–2580.

*Speaker
†Corresponding author: pascal.bonnet@univ-orleans.fr

Structure based pharmacophore and MM-PBSA method: a tool to design small molecules acting as XIAP inhibitors.

Marc Farag * [1] Charline Kieffer [1], Anne Sophie Voisin-Chiret [1], Jana Sopkova-de Oliveira Santos [1]

[1] Centre d'Études et de Recherche sur le Médicament de Normandie – UR4258, Université de Caen Normandie – France

XIAP (X-linked chromosome) is one of the human inhibitory apoptosis protein family including also other members like cellular IAP1/2 (cIAP 1/2), neuronal apoptosis protein (NAIP), survivin (TIAP), Apollon, melanoma IAP (ML-IAP), and IAP-like protein 2 (ILP2)1. They are considered as key regulators of cell death (apoptosis). Specifically, XIAP is known for its higher affinity to caspase enzymes released from mitochondria, through its baculoviral IAP repeats domains (BIR). Also, it contains an ubiquitin-associated domain (UBA) and a really interesting new genes (RING) with a ligase activity. XIAP-binding mechanism results in promoting cell survival regulated by the action of the second mitochondrial activator of caspases (SMAC/DIABLO). The endogenous antagonist, SMAC, binds to XIAP BIR2 or BIR3 domains releasing caspases and reactivating intrinsic signaling pathways leading to apoptosis. Overexpression of XIAP is involved in cancer and inflammatory diseases like Multiple sclerosis. That is why XIAP is considered a potential target especially, for cancer therapeutics2 and the XIAP-BIR3 domain was initially exploited to design drugs to cancer treatement. Nevertheless, the XIAP-BIR3 antagonists lacked selectivity. XIAP inhibitors also interacted with the cIAP1-2 - BIR3 domains, leading to a cIAP1-2 inhibition3 with in consequence an increase in the release of TNFa, resulting in a potentially serious side effect called "cytokine release syndrome".4,5
This project aims to build an in-silico tool to design non-peptide antagonists selective for BIR3 domains of XIAP versus BIR3 domains of cIAP1-2. In order to well understand the essential features i.e. chemical moieties required to design selective compounds for XIAP BIR3, a 3D pharmacophore model was built using Ligand Scout.6 It was based on the combination of the X-ray structure of 5M6L and a molecular dynamics simulation. Several modifications were imposed on this 3D pharmacophore in order to refine it and at the end 5 features were chosen: 2 hydrophobic ones, 2 Hbond donors, 1 Hbond acceptor. The selected pharmacophoric features were then correlated with to free energy predictions using MM-PBSA approach.

Bibliographic references:

1. Silke, J. & Vucic, D. IAP family of cell death and signaling regulators. *Methods Enzymol.* **545**, 35–65 (2014).

2. Holcík, M. Targeting endogenous inhibitors of apoptosis for treatment of cancer, stroke and multiple sclerosis. *Expert Opin. Ther. Targets* **8**, 241–253 (2004).

3. Fulda, S.; Vucic, D., Targeting IAP proteins for therapeutic intervention in cancer. *Nature reviews. Drug discovery* **2012,** 11, (2), 109-24.

*Speaker

4. Fulda, S., Smac mimetics as IAP antagonists. *Seminars in cell & developmental biology* **2015,** 39, 132-8.

5. Infante, J. R.; Dees, E. C.; Olszanski, A. J.; Dhuria, S. V.; Sen, S.; Cameron, S.; Cohen, R. B., Phase I dose-escalation study of LCL161, an oral inhibitor of apoptosis proteins inhibitor, in patients with advanced solid tumors. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **2014,** 32, (28), 3103-10

6. Wolber, G. & Langer, T. LigandScout: 3-D Pharmacophores Derived from Protein-Bound Ligands and Their Use as Virtual Screening Filters. *J. Chem. Inf. Model.* **45**, 160–169 (2005).

Posters: Structure-based drug design

P1-Design and optimization of PROTACs compounds for cosmetic industry: the Excel Project

Samdani Ansar [1], Jonathan Ribes [1], Pascal Bonnet [1], Sylvain Routier [1], Karen Plé [1], Samia Aci-Sèche [1]

[1] Institute of Organic and Analytical Chemistry (ICOA), Université d'Orléans, Pôle de chimie rue de chartres -45067 Orléans Cedex 2, France – CNRS : UMR7311

Currently, the cosmetic industry has to face increasingly complex conditions to market new products, and must ensure complete consumer safety of all ingredients. The goal of the Excel project is to develop new *in vitro* predictive cellular tools to evaluate the safety of these ingredients. In this context, we will develop compounds to chemically 'switch off' a target protein. The use of these biological systems rendered "knock out" (KO) will make it possible to determine: 1) the involvement of the target in a metabolic pathway and 2) its susceptibility to trigger a molecular initiating event (MIE) that could present a risk for the health of the consumer.

For this purpose, teams of molecular modeling, chemical synthesis and biology from Orléans formed a partnership to design and validate PROTACs (proteolysis targeting chimera) for selected protein targets in the Excel (Extinctions cellulaires sélectives pour la sécurité de composants cosmétiques) project. PROTACs are bifunctional small molecules consisting of two ligands joined by a linker: one ligand recruits and binds a protein of interest (POI) while the other recruits and binds an E3 ubiquitin ligase. Simultaneous binding of the POI and ligase by the PROTAC induces ubiquitylation of the POI and its subsequent degradation by the ubiquitin–proteasome system (UPS), after which the PROTAC is recycled to target another copy of the POI for degradation.

In our consortium, the role of the molecular modeling team consisted of optimizing the size and the composition of the linker molecule. To carry out this task, we investigated, benchmarked and applied a PROTAC modelling method developed by the CCG company in their software suite MOE1,2.

References:
1. Drummond M. L. & Williams C. I. "In Silico Modeling of PROTAC-Mediated Ternary Complexes: Validation and Application.", *J.Chem. Inf. Model.*, **2019**, 59 (4), 1634-1644. DOI:10.1021/acs.jcim.8b00872
2. Drummond M. L., Henry A., Li H. & Williams C. I. "Improved Accuracy for Modeling PROTAC-Mediated Ternary Complex Formation and Targeted Protein Degradation via New In Silico Methodologies.", *J.Chem. Inf. Model.*, **2020**, 60 (10), 5234-5254. DOI:10.1021/acs.jcim.0c00897

P2-Exploring Dynamic Flexibility for Improved Protein-Ligand Docking: A Case Study of the Androgen Receptor

Mehdi Oudahmane * [1], Matthieu Montes [1], Nathalie Lagarde [1]

[1] Laboratoire Génomique, Bioinformatique et Chimie Moléculaire (GBCM), EA 7528 – Conservatoire National des Arts et Métiers (CNAM) – 292 rue Saint Martin, 75003 Paris, France, Hesam Université

In recent years, there has been a growing emphasis on incorporating protein flexibility in protein-ligand docking strategies. However, simulating the complete flexibility of the protein can be computationally demanding and impractical for large-scale database screenings. In this study, we propose a novel protocol that intelligently identifies a subset of residues, without the need for extensive prior computational studies, whose side chains should be treated as flexible during the docking process. We subsequently assess the influence of this selection on the accuracy of docking predictions.

This study focuses on the androgen receptor (AR) (1), which is a member of the nuclear receptor (NR) family. Compounds experimentally evaluated in AR binding assays were extracted from the Environmental Protection Authority (EPA) Dataset Gateway (2) and divided in AR binding (2) and non binding (NB) compounds. AR structures were extracted from the Protein Data Ban and were evaluated using the MolProbity webserver (3). 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. For the flexible protocols, we first set as flexible the side chains of all residues within a 4Å cut-off distance from the co-crystallized ligand. Then, we developed a new protocol by investigating the docking performance associated with different combinations of 1 to 6 binding site residues with flexible side chains. These 6 residues were rationally 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 protein flexibility into account can improve docking performance. In particular, we highlighted that this event can be achieved by selecting only a small number of flexible side chains which 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.

*Corresponding author: mehdi.oudahmane@hesam.eu

(4) - Quiroga, Rodrigo, et Marcos A. Villarreal. " Vinardo: A Scoring Function Based on Autodock Vina Improves Scoring, Docking, and Virtual Screening ". PLOS ONE, 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. 6197205.

P3-First results based on docking methodologies for the identification of potential interacting small molecules to target the IL-3/IL-3Rα interface in case of atopic diseases

Lucas Rouaud [1], Alexandre Rizk [1], Alexandre De Brevern [1], Julien Diharce* [1]

[1] Biologie Intégrée du Globule Rouge – Université de la Réunion, Institut National de la Santé et de la Recherche Médicale, Université des Antilles, Université Paris Cité – France

Protein-protein interactions (PPIs) are attractive targets for drugs development as they are critical in a variety of biological processes and pathologies. As an illustration, the interleukin 3 (IL-3) and its α subunit receptor (IL-3Rα) are two proteins belonging to the cytokine (or receptor βc) family, which comprises also IL-5 and GM-CSF They are involved in several disorders like inflammatory diseases or hematological malignancies but IL-3 is specifically involved in atopic diseases, such as allergy. Those proteins are characterized by a two-subunit receptor (subunits α and β) which binds sequentially to the protein (first the α subunit and second the βc). The IL-3/IL-3Rα exhibits a low binding affinity. Interestingly, a complex formed by a mutant form of IL-3 (superkine) and IL-3Rα have emerged from the literature, with an increase of the affinity. The BASIN ANR project is therefore centered around the study of this complex, using experimental and theoretical approaches, in order to decipher in the most precise way its behavior and to eventually propose small compounds of interest with therapeutic potential in the case of atopic diseases. We already shown through our studies that the IL-3/IL-3Rα mouse interface is quite different than the human interface by evolutional analysis.1 In addition, we analyzed the dynamical behavior of the IL-3/IL-3Rα complex and shed light on a new conformation sampled, never seen before in the several crystallographic structures already solved.2 Starting from those findings, we present here our first insights of the drug discovery part of the project, targeting this particular interface: we first use the CNE dataset in order to benchmark several docking software and several scoring functions, in order to find the best solution for our system. Second, we design a complete pipeline in order to directly process a whole compounds database (preparation steps) and to realize the docking with one or several software on both partners of the complex: IL-3 and IL-3Rα. Our first results show that targeting IL-3 is more relevant that the receptor because of the shape of the interacting surface of IL-3Rα, which is not quite adapted for interacting with small compounds.
References:

1. Fogha, J., Bayry, J., Diharce, J., de Brevern A.G. Structural and evolutionary exploration of the IL-3 family and its alpha subunit receptors. Amino Acids **53**, 1211–1227 (2021).
2. Fogha, J, de Brevern, AG, Diharce, J. Conformational and binding mode assessment of the human IL-3 recognition by its alpha receptor. Proteins. 2023; 91(7): 904-919.

*Corresponding author: julien.diharce@u-paris.fr

P4-Solving the Hydrophobic-Polar model with Nested Monte Carlo Search

Milo Roucairol [1], Tristan Cazenave [??] * [1]

[1] Laboratoire d'analyse et modélisation de systèmes pour l'aide à la décision – Centre National de la Recherche Scientifique : UMR7243 / FRE3234 / UMR7024, Université Paris Dauphine-PSL – France

In this paper we present a new Monte Carlo Search (MCS) algorithm for finding the ground state energy of proteins in the Hydrophobic-Polar-model (HP model). We also compare it to other MCS algorithms not usually used on the HP model as well as to other approaches and provide an overview of the state of the art algorithms used on the HP model.

*Corresponding author: tristan.cazenave@lamsade.dauphine.fr

P5-Take your synthon-based ligand discovery to infinity and beyond with SATELLiTES, a Synthon-based Approach for the Targeted Enumeration of Ligand Libraries and Expeditious Screening

Corentin Bedart* [1], Seth Vigneron [2], John J. Irwin [2], Matthieu Schapira† [1,3]

[1] Structural Genomics Consortium, University of Toronto, Toronto, ON M5G 1L7 – Canada
[2] Department of Pharmaceutical Chemistry, University of California, San Francisco, CA, United States
[3] Department of Pharmacology and Toxicology, University of Toronto, Toronto, ON M5S 1A8, Canada

With the fast-paced growth of ultra-large virtual libraries of tens of billions of molecules, efficient exploration of chemical space will play a critical role in the identification of novel therapeutic compounds (1). A promising strategy for expanding the accessible drug-like chemical space is the synthon-based drug discovery, combining reactive small molecules known as building blocks or synthons with well-known chemical reactions. However, the number of accessible compounds surpasses the resources required to screen them computationally using conventional structure-based methods. Novel hierarchical combinatorial approaches can enable the fast exploration of ultra-large chemical libraries (2,3).

To make this method more accessible, we propose SATELLiTES, an open-source software designed to facilitate chemical space exploration and aid in the discovery of new hit compounds. Starting from a specific two-reagents chemical reaction encoded in SMARTS format and a list of commercially available or ready-to-use compatible synthons, the software operates in two successive steps.

The first step of SATELLiTES generates a Minimum Enumerated Libraries (MEL) by combining all compatible synthons where each pair includes either the smallest possible synthon, a representative chosen from the dataset, or a manually designed representative that best suits the anticipated pharmacophore. The libraries of representative compounds generated in this way cover all compatible synthons in an additive manner rather than a combinatorial one.

Once a structure-based method has been applied to select the best MEL synthons, the second step of SATELLiTES enumerates Focused Enumerated Libraries (FEL) by combining the selected MEL synthons with all chemically compatible building blocks. These libraries ensure that only compounds containing one of the best synthons are considered, significantly reducing the combinatorial chemical space that needs to be explored.

To use SATELLiTES, please visit our GitHub repository (github.com/cbedart/SATELLiTES) and embark on a journey to explore the vast chemical space.

(1) Warr WA, et al. Exploration of ultralarge compound collections for Drug Discovery. Journal

*Corresponding author: corentin.bedart@gmail.com
†Corresponding author: matthieu.schapira@utoronto.ca

of Chemical Information and Modeling. 2022;62(9):2021–34. doi:10.1021/acs.jcim.2c00224

(2) Sadybekov AA, et al. Synthon-based ligand discovery in virtual libraries of over 11 billion compounds. Nature. 2021;601(7893):452–9. doi:10.1038/s41586-021-04220-9

(3) Sadybekov AV, Katritch V. Computational approaches Streamlining Drug Discovery. Nature. 2023;616(7958):673–85. doi:10.1038/s41586-023-05905-z

P6-The pivotal role of RNA helicase A (RHA) targeted for developing new antiviral drugs

Laetitia Marty [1], Laurence Briant [1], Mickael Blaise [1], Jean-Marie Péloponèse [1], Laurent Chaloin [1]

[1] Institut de Recherche en Infectiologie de Montpellier – Centre National de la Recherche Scientifique, Université de Montpellier – France

RNA helicases (RHA) play an important role in unwinding double helix of DNA or RNA prior the transcription of cellular genes. In human cells, RHA interacts with RNA polymerase II, transcription factors and co-activators to modulate the transcription and the translation of selected genes. In parallel, many viruses have been shown to take benefit of this cellular function since RHA interacts with viral RNA and/or viral proteins allowing to enhance their replication. Indeed, RHA has been described for its contribution during the replication of RNA viruses such as HIV, HTLV-1, Dengue, Zika and SARS-CoV-2 viruses. Our project is devoted to understand the molecular mechanisms by which RHA is involved and required for viral replication and also to develop small molecules as antiviral drugs targeting this helicase in order to block the viral proliferation. We have already identified by virtual screening heterocyclic compounds targeting the core domain of RHA (outside the ATP binding site) that are able to bind to RHA and to inhibit viral replication in cell-based assays. Several rounds of lead optimizations have been started by using chemoinformatic approaches and chemical synthesis in order to improve the efficacy and the selectivity of these hits. Optimized compounds demonstrated an antiviral activity in the low micromolar range in HIV-1 infected cells. Each new series is further evaluated by molecular docking and molecular dynamics simulations to estimate their binding stability and key interactions. Correlations between *in silico* predictions and experimental data allowed to determine the minimal scaffold responsible for the antiviral activity and crystallographic studies are in progress to go further in the development of even more potent antiviral drugs.

P7-Modelling of the binding of neomycin B and its derivative into Aminoglycoside-Modifying Enzymes

Julia Revillo Imbernon [1], Esther Kellenberger* [1]

[1] Laboratoire d'Innovation Thérapeutique – Université de Strasbourg, Institut de Chimie du CNRS, Centre National de la Recherche Scientifique : UMR7200 – France

Antibiotic resistance is a major challenge for humanity in terms of public health, economics, and ecology. The World Health Organization (WHO) has called for a global response with a "one health" approach including the development of new antibiotics. Our project falls within this framework and in collaboration with Pr. Weibel (LaSyROC, UMR 7177, CNRS-University of Strasbourg), Dr. Prévost (Virulence bactérienne précoce, UR 7290, Université de Strasbourg), and Dr. Ennifar (Architecture et réactivité de l'ARN, UPR 9002, CNRS-University of Strasbourg) for the optimization of neomycin, maximizing the bactericidal effect by minimizing antimicrobial resistance and toxicity.

Neomycin is an aminoglycoside with excellent antibacterial activities against gram-negative and gram-positive bacteria. It targets bacterial ribosomal RNA and thereby inhibits translation. Resistances to neomycin mostly involve Aminoglycoside-Modifying Enzymes deactivations (AME), which are the cause of approximately 60 to 70% of acute resistances.1

The three-dimensional structures of five AMEs were modeled by homology for *E. coli* and *S. aureus* from clinical strains. For each model, neomycin and the synthetic derivative (HL_171) that showed the most interesting MIC activity were docked under constraints applied to ring II into the active site. All complexes were submitted to molecular dynamics simulation (3x 200ns). The binding modes, distances between the derivatives and the catalytic centers, and conformation of the active site were monitored and considered as indicators of the possible catalytic ability.

The objective of the simulations was to ascertain the impact of neomycin modifications on AMEs activity. Based on the analysis, two complexes exhibited a similar binding mode to neomycin B, inferring that the modifications do not influence metabolization. On the contrary, the remaining three complexes presented a dissimilar binding mode, suggesting that they could hinder AMEs metabolization activity.

1. Ramirez, M. S. & Tolmasky, M. E. Aminoglycoside Modifying Enzymes. *Drug Resist. Updat. Rev. Comment. Antimicrob. Anticancer Chemother.* **13**, 151–171 (2010).

*Corresponding author: ekellen@unistra.fr

Posters: Machine learning / Datamining in chemoinformatics

P8-Assessment of Drug-Induced Liver Injury through cell morphology and gene expression analysis.

Vanille Lejal [1], Olivier Taboureau [1], David Rouquié [2], Natacha Cerisier [1]

[1] Université Paris Cité – Unité de Biologie Fonctionnelle et Adaptative, UMR 8251, Université Paris Diderot, Université Sorbonne Paris Cité, CNRS – France
[2] Bayer Crop Science – France

Although many studies have been reported to identify biomarkers and genes/pathways signatures related to liver toxicity, Drug-Induced Liver Injuries (DILI) remain a challenge in drug discovery. Here, we explored the morphological cell perturbations induced by 575 hepatotoxic compounds with multiple DILI annotations collected from two datasets: DILIRank and eTox. Overall, these results present a combination of high-content imaging with transcriptomics data that allow to identify a relation between morphological cell perturbations and gene regulation related to DILI. First, mechanisms of action of clusters of chemicals sharing similar morphological cell perturbations were analyzed using transcriptomics data. Knowing that common signaling pathways associated with DILI could be captured from these clusters, we also developed machine learning approaches based on morphological features aiming to predict potential DILI. A model based on the ElasticNet algorithm succeeded to give hepatotoxicity predictions for several unknown chemicals showing significant morphological cell perturbations compared to controls.

P9-Cellular cartography: Decoding myotoxicity and predicting cell health through cell morphological profiling

Roman Lambert [1,2], Howard Fearnhead [1], Eva Serrano Candelas [2], Rafael Gozalbes [2]

[1] University of Galway, School of Medicine, Pharmacology & Therapeutics – Ireland
[2] ProtoQSAR SL – Spain

Advances in high-content cell microscopy allow the conception of new types of multiplex assays, by measuring rich data on morphological changes in cell populations. One of these assays, named Cell Painting, uses 6 fluorescent dyes to stain multiple cell organelles over 5 image channels. The following study is part of the TOXIFATE project (https://toxifate.net/), which aims at finding new ways to predict skeletal muscle toxicity, by combining conventional computational toxicity prediction methods such as QSPR, with cell morphology and gene expression analysis.

A panel of 30 myotoxicants was used to treat C2C12 muscle cells for 72 hours along eight doses. Cell morphology profiles were extracted with a high-throughput microscope, and normalized against untreated control cells, hence creating individual cell phenotypes. Profile aggregation allowed further analysis: hierarchical clustering unveiled information on the mode of action of the myotoxicants and separated them in meaningful classes, but also revealed unsuspected associations, with for instance sunitinib and the statins family. We also showed that these morphological profiles contain ample information on cell health, allowing the prediction of the outcomes of an external cell viability assay based on ATP content.

This study paves the way towards new ways of predicting cyto- and myotoxicity by bringing new kinds of biological descriptors related to individual cell morphology, that can be used in tandem with more conventional methods such as QSPR for building precise and efficient models describing toxicity.

P10-Implementation of soft grading systems for chemistry in a Moodle plugin.

Louis Plyer [1], Gilles Marcou* [1], Alexandre Varnek [1], Céline Perves [2]

[1] Laboratoire de Chémoinformatique, UMR 7140 CNRS, Strasbourg, France – Faculté de Chimie de Strasbourg – France
[2] DNUM – Université de Strasbourg, DNUM – France

We are developing a set of tools in the ChemMoodle project that aims to facilitate the learning of chemical drawing skills with the help of the MOODLE learning management system. In particular, we propose a soft grading system for automated correction of quiz in chemistry. This system is based on computing the similarity between the student's answer and the correct answer, provided by the teacher. The ChemMoodle plugins implement the soft grading system for the drawing of chemical structures, being sensitive to lone pairs, radicals and stereochemistry notations. Chemical reactions are now supported. This is done using atom-to-atom mapping and Condensed Graph of Reaction (CGR). CGR are pseudo-molecules recording the details of a chemical transformation. It is implemented as MOODLE plugins, using Chemdoodle engine for drawing structures, and communicating with a REST server (open source) -that can be installed and managed locally- to compute the similarity using ISIDA molecular descriptors.

*Corresponding author: g.marcou@unistra.fr

P11-Interpretable genomic space maps for S.aureus antimicrobial resistance prediction

Karina Pikalyova [1], Alexey Orlov [1], Dragos Horvath [1], Gilles Marcou [1], Alexandre Varnek* [1]

[1] Université de Strasbourg – Laboratoire de chémoinformatique – France

Increasing antimicrobial resistance (AMR) represents a global healthcare threat.(1) To decrease the spread of AMR and associated mortality, methods for rapid selection of optimal antibiotic treatment are urgently needed. Machine learning (ML) techniques based on genomic data to predict resistance phenotypes serve as a fast screening tool prior to phenotypic testing. Nonetheless, many existing ML methods lack interpretability, a particularly important feature for applications in the field of clinical decision-making. Herein, we present a methodology for visualization of genomic space and AMR prediction based on the non-linear dimensionality reduction method - generative topographic mapping (GTM)(2). This methodology was applied to predict AMR of *S. aureus* strains using a dataset of over 5000 *S. aureus* isolates with associated resistance labels retrieved from PATRIC database. All genome-based GTM models for resistance prediction against a range of 11 antibiotics possessed reasonable prediction accuracy (balanced accuracy values $\geq$0.75). At the same time, the GTMs represent complex genomic data in the form of the illustrative 2D maps. These maps were successfully used for antibiotic-wise comparison of resistance phenotypes, and were found to be efficient for the analysis of genomic biomarkers that are responsible for drug resistance development. Overall, the GTM-based methodology is a useful tool for the illustrative exploration of the genomic sequence space and modelling AMR and can complement existing ML methods for AMR prediction.(3)

(1) - Antimicrobial Resistance Collaborators, Global Burden of Bacterial Antimicrobial Resistance in 2019: A Systematic Analysis. *Lancet (London, England)* **2022**, *399* (10325), 629–655. https://doi.org/10.1016/S0140-6736(21)02724-0.
(2) - Bishop, C. M. et al., GTM: The Generative Topographic Mapping, 1998, Vol. 10. https://doi.org/10.1162/089976698300017953.
(3) - Pikalyova, K. et al. Predicting S. Aureus Antimicrobial Resistance with Interpretable Genomic Space Maps. *bioRxiv* **2023**, 2023.02.24.529878. https://doi.org/10.1101/2023.02.24.529878.

*Corresponding author: varnek@unistra.fr

P12-Machine learning model for chromatographic retention prediction

Jérémy Molineau* [1], Stéphane Bourg [2], Nicolas Boisseau [3], Maria Hideux [3], Philippe Hennig [3], Sophie Bertin [3], Fabien Mauge [3], Pascal Bonnet [1], Caroline West [1]

[1] Université d'Orléans – Institut de Chimie Organique et Analytique - ICOA UMR7311, Orléans – France
[2] CNRS Orléans – Institut de Chimie Organique et Analytique - ICOA UMR7311, Orléans – France
[3] Institut de Recherche Servier – France

Identifying impurities in drug products is a duty to ensure patient safety. This task is reserved to analytical chemists, who must separate the impurities from active components, quantify and identify them. From this standpoint, analysts are using different analytical instruments depending on the molecule's nature, usually starting from Reversed-phase Liquid chromatography (RPLC). Therefore, selecting the most adapted chromatographic technique to analyze any newly synthesized molecule is an industrial objective. It will reduce solvent and energy consumptions but also save human time on inadequate analysis.

In order to ensure this, artificial intelligence through machine learning allows adapted models construction for prediction of chromatographic analysis (1). The molecular structure should guide the selection of the best analytical method. From the molecular diversity and chromatographic data measured over several years in the partner **pharmaceutical** company, we built a model to predict the chromatographic retention for new small molecules. This model and data preparation have been conducted with Python libraries (RDKit, Scikit-learn, Pandas, ...).

In this discussion, we would like to present the construction of our first algorithm. Starting with the data preparation and going through the parsing to recover retention time. Then, we will focus on descriptors calculation and selection. At last, performances of a first regression model will be exposed using the mean absolute error as metric.

1. Fabrice Gritti, Anal. Chem. 2021, 93, 14, 5653–5664, DOI: 10.1021/acs.analchem.0c05078

*Corresponding author: jeremy.molineau@univ-orleans.fr

P13-Mining the Protein Data Bank to inspire fragment library design.

Julia Revillo Imbernone [1], Luca Chiesa [1], Esther Kellenberger* [1]

[1] Laboratoire d'Innovation Thérapeutique – Université de Strasbourg, Institut de Chimie du CNRS, Centre National de la Recherche Scientifique : UMR7200 – France

The following abstract is based on our previously published article "Revillo Imbernon, J., Chiesa, L. & Kellenberger, E. Mining the Protein Data Bank to inspire fragment library design. *Front. Chem.* **11**, (2023)." and has been adapted for presentation as a poster.
The fragment-based approach has become a popular method for drug discovery due to its ability to tackle challenging targets. The success of this approach depends on various factors, including the selection of a suitable chemical library, biophysical screening method, and high-quality fragments. It has been suggested that using promiscuous compounds can increase the likelihood of obtaining hits in screening and expand the diversity of fragment libraries. In this study, we explored the potential of promiscuous compounds for fragment-based drug design. By analyzing the Protein Data Bank, we identified 203 fragments with multiple binding modes and targeting different sites, represented by 90 scaffolds. Our set of fragments is enriched in three-dimensional scaffolds that are not widely available in commercial libraries. The dataset is available for download at 10.5281/zenodo.7554649.

1. Revillo Imbernon, J., Chiesa, L. & Kellenberger, E. Mining the Protein Data Bank to inspire fragment library design. *Front. Chem.* **11**, (2023).

*Corresponding author: ekellen@unistra.fr

P14-Optimizing electrochemical reaction yield using Bayesian optimization and sensitivity analysis

Lise Kastner [1], Bertrand Cuissart [1], Jean-Luc Lamotte [1]

[1] Equipe CODAG - Laboratoire GREYC - UMR6072 – Groupe de Recherche en Informatique, Image et Instrumentation de Caen – France

The aim is to optimize the yield of an electrochemical reaction by finding a good set of parameters in a limited number of experiments. A solution is to use active learning to search for the optimum: from a few initial observations, we use Bayesian optimization to propose new experiments to the chemist while the best yield is not reached.

However, in electrochemistry, the number of configurations of a reaction is very large: It implies that a Bayesian optimization could require to test an important number of experimental conditions before reaching a yield sufficient enough.

To solve this problem, we intend to quantify the association between the parameters of the reaction and its yield using sensitivity analysis. It will help us to focus on the most important parameters and, hopefully, reduce the number of propositions given by Bayesian optimization.

In this poster, we'll see how to apply sensitivity analysis on Bayesian optimization, in order to find good yields earlier in the process.

This work is part of the AMPERE project, in collaboration with 3 chemistry laboratories: LIMA (UMR 7042), LBM (UMR 7203), LHFA (UMR 5069).

P15-Predicting solubility, but which one ?

Pierre Llompart [1,2], Shamkhal Baybekov [1], Gilles Marcou* [1], Dragos Horvath [1], Patrick Gizzi [3], Jean-Luc Galzi [3,4], Pascal Ramos [5], Olivier Saurel [5], Claire Bourban [3], Claire Minoletti [2], Alexandre Varnek [1]

[1] Laboratory of Chemoinformatics – CNRS : UMR7140, Strasbourg – France
[2] Sanofi – SANOFI Recherche – France
[3] Laboratory of Biotechnology and cell signaling – CNRS : UMR7242, Illkirch – France
[4] ChemBioFrance - Chimiothèque Nationale, Montpellier – Institut de Chimie du CNRS – France
[5] Institut de Pharmacologie et de Biologie Structurale – CNRS : UMR5089, Toulouse – France

Water solubility measurements are categorized as thermodynamic or kinetic. Thermodynamic solubility is subdivided into pure water, apparent, and intrinsic solubility. It defines the maximum concentration of a compound at thermodynamic equilibrium and differ by the control of pH and protonation state. Kinetic solubility is preferred during early drug screening. It represents the lowest concentration at which precipitation occurs when diluting a stock solution. Despite being perceived as a crude estimate, it addresses a distinct phenomena. Public in silico models addressing kinetic solubility remain scarce, possibly due to perceived reproducibility issues and that thermodynamic solubility models could be used instead. Indeed, thermodynamic solubility models are regarded as reliable and are widely used to annotate major public databases, yet they often underperform in real-world pharmaceutical context due to data quality and applicability domain issues. On the other hand, kinetic solubility datasets are available, and models derived on them are reliable. Our study emphasizes the need to foster kinetic model development to improve early-stage drug screening reliability, urging careful data curation and consideration of physicochemical phenomena influencing the measures. Our SVM, GTM, and RF models built on the public data (https://doi.org/10.57745/CZVZIA) outperform earlier reported models. These models are implemented in the ISIDA/Predictor tool and are publicly available along with the datasets: https://chematlas.chimie.unistra.fr/WebTools/predictor_solubility.php.

*Corresponding author: g.marcou@unistra.fr

P16-Prediction of endocrine disrupting chemicals covering EATS modalities using transcriptomics and machine learning

Guillaume Ollitrault* [1], Marco Marzo [2], Emilio Benfenati [2], Alessandra Roncaglioni [2], Enrico Mombelli [3], Olivier Taboureau† [1]

[1] Unité de Biologie Fonctionnelle et Adaptative – Institut National de la Santé et de la Recherche Médicale, Centre National de la Recherche Scientifique, Université Paris Cité – France
[2] Department of Environmental Health Sciences, Laboratory of Environmental Chemistry and Toxicology, Istituto di Ricerche Farmacologiche Mario Negri IRCCS, Milan – Italy
[3] Institut National de l'Environnement Industriel et des Risques – INERIS – France

Endocrine disrupting chemicals (EDC) are chemicals that can interfere with homeostatic processes. They are a major issue for public health, they can cause adverse long-term effects such as cancer, intellectual impairment, obesity, diabetes and male infertility. The endocrine system is a complex machinery with the Estrogen (E), Androgen (A), Thyroid (T) and Steroidogenesis (S) modalities being of most importance. We developed Qualitative Gene expression Activity Relationship (QGexAR) models to assess the risk of disruption of EATS modalities for new chemicals. We used gene expression profiles from LINCS database tested on 5 cell lines used in endocrine disruption analysis the PC3 (prostate cancer), MCF7 (breast cancer), HA1E (immortalized kidney cell line), A549 (adenocarcinomic human alveolar basal epithelial cells), A375 (epithelial cell), cell lines. We optimized our prediction protocol by testing different features selection methods and classification algorithms including catboost, xgboost, Random Forest, SVM, Logistic regression, autokeras, tpot, deep learning models (three fully connected neural network models). For each endpoint the final prediction was made according to the consensus prediction of the 5 models of each cell line. With the available data we were able to develop a performant predictive model for the Estrogen receptor bind effect, the Androgen receptor binding effect, the Thyroid receptor antagonist with a consensus balance accuracy on the validation set of respectively 0.727, 0.806 and 0.802. The limited size of available data for thyroid peroxidase and steroidogenesis perturbation prediction was not sufficient to develop performant models. The features importance of the models was assessed using the inner model method and a permutation algorithm to identify known and new mechanisms of action of EATS perturbation.

*Corresponding author: guillaume.ollitrault@gmail.com
†Corresponding author: olivier.taboureau@univ-paris-diderot.fr

P17-Refinement of a ligand activity and representation of topological pharmacophores in a colored network

Maroua Lejmi [1,5], Damien Geslin [2], Bertrand Cuissart [1], Ilef Ben Slima [3], Nida Meddouri [6], Ronan Bureau [2], Alban Lepailleur [2], Amel Borgi [4], Jean-Luc Lamotte [1]

[1] Equipe CODAG - Laboratoire GREYC - UMR6072 – Groupe de Recherche en Informatique, Image et Instrumentation de Caen – France
[2] Centre d'Études et de Recherche sur le Médicament de Normandie : UR4258, Université de Caen Normandie – France
[3] ISMAIK, University of Kairaoun, SM@RTS Sfax, Tunisie
[4]ISI-LIPAH / Université de Tunis El Manar, Tunis, Tunisie
[5] FST-LIPAH / Université de Tunis El Manar, Tunis, Tunisie
[6] Laboratoire de Recherche de l'EPITA (LRE), Le Kremlin-Bicêtre, France.

Structure-Activity Relationships is a critical aspect of drug design. It enables us to examine ligand interactions and performances towards specific targets, then to design effective drugs for treating diseases or improving existing medical therapies. In this context, we specifically study the activity of ligands towards kinases using the BCR-ABL dataset. The work is dedicated to introduce a refinement method for the molecules activity. Instead of considering affinity as a binary activity, a molecule being either active or inactive, the compounds were partitioned into 4 classes according to their activity: very active, moderately active, slightly active, inactive. This activity is later used to evaluate molecular descriptors called topological pharmaophores [1]. These pharmacophores provide essential information by representing the key structural features of a molecule. Their quality is determined by measuring their "growth-rate" which corresponds to the ratio of active molecules over inactive ones, among the molecules supported by the pharmacophore. In our work, the calculation of the growth-rate is based on the classes of activity that we have created. Consequently, we will obtain three measurements of the growth rate, each one being related to a class of activity. In addition, we proposed to convert the new information of the quality of the pharmacophores into a visual representation called "The Pharmacophore Network" [2]. The latter is a graph whose nodes represent the pharmacophores and edges represent a graph-edit distance that separates them. Our goal was to structure more finely the pharmacophore space and to be able to detect visually interesting areas that can be explored. For this purpose, we integrated colors in this Pharmacophore Network, where each color refers to a class of activity.

References

[1] The Pharmacophore Network: A Computational Method for Exploring Structure-Activity Relationships from a Large Chemical Data Set, Metivier, Jean-Philippe ; Cuissart, Bertrand ; Bureau, Ronan ; Lepailleur, Alban, J. Med. Chem., volume 61, issue8, page 3551-3564, DOI10.1021/acs.jmedchem.7b01890,2018

[2] Deciphering a Pharmacophore Network: A Case Study Using BCRABL Data, Geslin, Damien ; Lepailleur, Alban ; Manguin, Jean-Luc ; Nhat-Vinh Vo ; Lamotte, Jean-Luc ; Cuissart, Bertrand ; Bureau, Ronan, J. Chem. Inf. Model., volume 62, issue3, page 678-691, DOI10.1021/acs.jcim.1c00427

Posters: Ligand-based drug design

P18-In sillico study of new UBE2N inhibitors to block the HR pathway in ovarian cancer

Côme Ghadi [1], Charline Kieffer [1], Anne Sophie Voisin-Chiret [1], Louis-Bastien Weiswald [2], Matthieu Meryet-Figuiere [2], Leonie Ibanez [2], Florian Schwalen [1], Jana Sopkova-de Oliveira Santos* [1]

[1] Centre d'Études et de Recherche sur le Médicament de Normandie – UR4258, Université de Caen Normandie – France

[2] Unité de recherche interdisciplinaire pour la prévention et le traitement des cancers – Université de Caen Normandie, CHU Caen, Centre Régional de Lutte contre le Cancer François Baclesse [Caen], Institut National de la Santé et de la Recherche Médicale – France

Every year, approximately 150,000 women worldwide lose their lives due to ovarian cancer. Recent laboratory research has revealed that inhibiting UBE2N (also known as Ubc13), a crucial mediator in the homologous recombination pathway (HR) of DNA repair, can make ovarian cancer cells more responsive to chemotherapy. Initial findings also indicate that inhibiting UBE2N (with an inhibitor called UBE2Ni) can sensitize ovarian cancer cells to PARP inhibitors (PARPi), which are proteins involved in DNA repair. PARPi has shown promising results in treating ovarian cancer patients, but its effectiveness is limited to only about one-third of patients whose cancers have defects in the HR pathway. By targeting UBE2N and disrupting HR, it is possible to significantly increase the number of patients who can benefit from PARPi treatment.

UBE2N is an E2 ubiquitin ligase that plays a role in the HR pathway of DNA repair as well as the NF-κB signaling pathway. It is an enzyme involved in the formation of nuclear Lys63-linked ubiquitin chains and requires interaction with the molecular partner Mms2 in the nucleus for DNA repair or Ubiquitin-conjugating Enzyme Variant 1a (Uev1A) in the cytoplasm for NF-κB signaling. The UBE2N/Mms2 complex facilitates the formation of ubiquitin chains by establishing a thioester bond between the ubiquitin molecule and the cysteine residue 87 in the active site of UBE2N.

The proximity of the UBE2N active site, specifically Cys 87, is influenced by a flexible loop consisting of amino acids 114-124. Movement of this loop leads to an open or closed conformation, regulating the volume of the cavity that accommodates the ligand. Currently, three inhibitors (UC-764864/65, NSC697923, and BAY 11-7082) that covalently bind to the cysteine in the UBE2N active site have been described in literature. Additionally, non-covalent inhibitors of UBE2N have been identified through in vitro screenings, but their binding mode on UBE2N is unknown.

The objective of this research is to develop new synthetic non-covalent inhibitors of UBE2N, aiming to significantly increase the number of ovarian cancer patients who can benefit from PARPi treatment. To achieve this, an in silico strategy was employed. By analyzing the available three-dimensional structures of UBE2N, two interesting pockets were identified for designing competitive inhibitors. The first pocket is located near Cys87, which is targeted by covalent inhibitors, and the second pocket is found on the binding surface with Mms2/Uev1A. In silico screening of CERMN's chemolibrary, containing 19,000 molecules, was conducted using a docking approach on these two sites to identify potential new non-covalent inhibitors. Three PDB structures of UBE2N (PDB code: 4ONM, 6UMP, 3HCU) representing different conformations

*Corresponding author: jana.sopkova@unicaen.fr

of the flexible loop near the cysteine site were used for the screenings. Subsequently, molecular dynamics simulations were employed to assess the stability of the interactions between the selected candidates and UBE2N. These simulations allowed for the identification of promising candidates based on interaction energy and complex stability. This groundbreaking project serves as the foundation for an ambitious program to synthesize innovative UBE2N inhibitors, ultimately enhancing the therapeutic care available to ovarian cancer patients who are currently not benefiting from PARPi treatment.

P19-The IMS library: an in-house collection based on chemical diversity, scaffold novelty and synthetic accessibility

Teodora Djikic-Stojsic* [1], Guillaume Bret [1], Gaëlle Blond [1], Nicolas Girard [1], Clothilde Le Guen [1], Claire Marsol [1], Martine Schmitt [1], Séverine Schneider [1], Frederic Bihel [1], Bonnet Dominique [1], Michaela Gulea Michaela Gulea [1], Esther Kellenberger† [1]

[1] Laboratoire d'Innovation Thérapeutique, UMR7200 CNRS-Université de Strasbourg, 74 route du Rhin, Illkirch-Graffenstaden 67400 – France

The initial phase of rational drug discovery process is, to a large extent, based on screening (both HTS and VS) a collection of chemicals (library), searching for a hit, that would bind to a target of interest. It is commonly believed that the wider the chemical space is covered by the library, the better the chance that a hit compound would be found.1 However, even though obtainable chemical spaces are expanding rapidly, containing billions of possible compounds, the usage of large data sets remains challenging. Working with such a large chemical space presents a constant compromise between cost, size, and time.2,3 Accordingly, diversity, novelty and drug-likeliness of the library are presenting themselves as more important than its size, in order to secure a better coverage, and to explore new areas of chemical space.4,5
Thanks to joint efforts of chemoinformaticians and medicinal chemist, we have developed two original and diverse, in-house, ready-to-use screening libraries.

Essential chemical library (eIMS) was created from the collection of previously synthetized compounds. It comprises of 578 compounds, with 80% purity, which are plated and ready for HTS. Diverse compounds were chosen by clustering, using molecular fingerprints, while their originality was determined by comparison to our in-house database of purchasable, drug-like compounds, collected from 25 trusted suppliers.

Virtual chemical library (vIMS) consists of 783.000 virtual compounds, derived from the most interesting and original scaffolds from the Essential library. Novel compounds were generated by adding different substituents on the specified connection points of the central scaffold, following the rules determined by medicinal chemists, to ensure their synthetic accessibility. Compounds were further filtered based on their chemical properties and substructures, to remove reactive and/or peculiar structures. Finally, we compared our virtual library to the Enamine REAL Space6, and concluded that around 90% of generated compounds could not be found there. Additional advantage of this library is the fact that it is based on well-known synthetic pathways and expert knowledge of medicinal chemists, which secures the easy synthesis of novel derivatives and facilitates hit-to-lead optimization.

*Corresponding author: djikicstojsic@unistra.fr
†Corresponding author: ekellen@unistra.fr

Posters: Polypharmacology

P20-Distinguishing molecular pathways between myoblasts and myotubes

Bibi Ameerah Furjun* [1,2], Roman Lambert [1,2], Eva Serrano-Candelas [1], Howard O. Fearnhead [2], Rafael Gozalbes [1,3]

[1] ProtoQSAR SL, Parque Tecnológico, Valencia – Spain
[2] Department of Pharmacology and Therapeutics, University of Galway – Ireland
[3] MolDrug AI Systems SL, Valencia – Spain

Defects in skeletal muscle cell proliferation and differentiation are associated with various skeletal muscle pathologies. Key molecular signatures induced by myotoxicants in myoblast differentiation and fusion remain to be characterized. In the context of the EU project "TOXIFATE", this research relies on a transcriptomics approach identifying and characterizing gene-expression changes brought about by known toxicants in myoblasts and myotubes to further identify the molecular signatures associated with drug-induced myotoxicity. By using a differential gene expression analysis on internal Tempo-seq data, we have mapped out the different genes that emerged from the comparative analysis between myoblasts and myotubes. Currently, we are using a functional analysis approach to link these genes to distinct signatures in cell fate in the two different muscle cell types as a way to predict drug-induced myotoxicity. Combined with the pathological pathways involved, such predictive toxicological approaches could provide insight into different cell death modalities implied in muscle cells.

*Corresponding author: afurjun@protoqsar.com

P21-Prediction of adverse drug reactions due to genetic predisposition using deep neural networks

Bryan Dafniet [1]

[1] Unité de Biologie Fonctionnelle et Adaptative – Institut National de la Santé et de la Recherche Médicale, Centre National de la Recherche Scientifique, Université Paris Cité – France

Drug development is a long and costly process, often limited by the toxicity and adverse drug reactions (ADRs) caused by drug candidates. Even on the market, some drugs can cause strong ADRs that vary depending on an individual polymorphism. The development of Genome-wide association studies (GWAS) allowed the discovery of genetic variants of interest that may cause these effects. In this study, the objective was to investigate a deep learning approach to predict genetic variations potentially related to ADRs. We used single nucleotide polymorphisms (SNPs) information from dbSNP to create a network based on ADR-drug-target-mutations and extracted interaction matrices to build deep Neural Networks (DNN) models. These DNNs predicted the association of a compound with a potential adverse effect category based on the MedDRA System Organ Classes (SOCs) with an average balanced accuracy of 0.61 having only information about mutations as variables. Including molecular fingerprints representing structural features of the drugs did not improve the performance of the models. To our knowledge, this is the first model that exploits DNN to predict ADR-drug-target-mutations. Although some improvements are suggested, these models can be of interest to analyze multiple compounds over all of the genes and polymorphisms information accessible and thus pave the way in precision medicine.

Keywords: drugs, adverse drug reactions, Single Nucleotide Polymorphisms, deep neural network, genetic variations

Session Data Science for Chemistry (atelier du GDR MaDICS)

SensaaS and MoleculArXiv, or the benefits of interdisciplinarity

Frédéric Payan * [1]

[1] Laboratoire d'Informatique, Signaux et Systèmes de Sophia Antipolis (I3S), Université Côte d'Azur / CNRS Sophia Antipolis, France

The Mediacoding team in the I3S laboratory has been working for more than three decades in the field of digital data processing, and more specifically on the problems of representation, coding, compression, transmission and visualization of multimedia objects such as images, videos, surface or volume meshes, 3D point clouds, etc.

Members of the team regularly take part in interdisciplinary projects, combining digital signal processing and computer science with biology, chemistry, geosciences, etc. During this presentation, I will talk about two current interdisciplinary projects, linked to chemistry and biology, in which Mediacoding members are taking part. Firstly, the SenSaaS project (https://github.com/SENSAAS/sensaas), which has resulted in a promising eponymous tool for aligning and comparing molecular shapes and sub-shapes, and secondly the PEPR MoleculArXiv (https://www.ins2i.cnrs.fr/fr/pepr-molecularxiv), which aims to store digital data on synthetic DNA.

These two projects are fine examples of success based on active collaboration between laboratories from different disciplines, within the same university for the SenSaaS project (I3S, a research laboratory dedicated to information and communication sciences, and IPMC, a research institute dedicated to molecular and cellular pharmacology), or at international level for the PEPR MoleculArXiv, which involves 20 laboratories from different disciplines.

*Speaker

List of participants

- Aci-Sèche Samia, samia.aci-seche@univ-orleans.fr
- Akhmetshin Tagir, tagirshin@gmail.com
- Audouze Karine, karine.audouze@inserm.fr
- Aulinka Chimprasit, chimprasit@unistra.fr
- Badel Anne, anne.badel@u-paris.fr
- Bagdad Youcef, bagdad.youcef.ybg@gmail.com
- Bedart Corentin, corentin.bedart@gmail.com
- Bonachera Fanny, f.bonachera@unistra.fr
- Botnari Marina, marina.botnari@etu.u-paris.fr
- Bourg Stéphane, stephane.bourg@cnrs.fr
- Bret Guillaume, guillaume.bret@unistra.fr
- Brun Luc, luc.brun@greyc.ensicaen.fr
- Bureau Ronan, ronan.bureau@unicaen.fr
- Camproux Anne-Claude, anne-claude.camproux@u-paris.fr
- Cazenave Tristan, Tristan.Cazenave@dauphine.fr
- Chaloin Laurent, laurent.chaloin@irim.cnrs.fr
- Charifi-Hoareau Guillaume, guillaume.charifi@ext.inserm.fr
- Chiesa Luca, luca.chiesa@unistra.fr
- Colas Aurélie, aurelie.colas@airliquide.com
- Colloc'h Nathalie, colloch@cyceron.fr
- Costa Fabrizio, f.costa@exeter.ac.uk
- Couronne, Christophe christophe.couronne@unicaen.fr
- Crémilleux Bruno, bruno.cremilleux@unicaen.fr
- Cuissart Bertrand, bertrand.cuissart@unicaen.fr
- Dafniet Bryan, bryan.dafniet@gmail.com
- Dekimeche Katia, katia.dekimeche@schrodinger.com
- Diharce Julien, julien.diharce@u-paris.fr
- Djikic-Stojsic Teodora, djikicstojsic@unistra.fr
- Douguet Dominique, douguet@ipmc.cnrs.fr
- Elbahnsi Ahmad, ahmad.elbahnsi@inserm.fr
- Falcon-Cano Gabriela, gabriela.falcon@pikairos.com
- Farag Marc, marc.ragui@unicaen.fr
- Furjun Bibi Ameerah, a.frjn.a@gmail.com
- Geslin Damien, damien.geslin@unicaen.fr
- Ghadi Côme, comeghadi@gmail.com
- Giovannini Johanna, johanna.giovannini@unicaen.fr
- Guillaume Valentin, valent.guillaume@gmail.com
- Horvath Dragos, dhorvath@unistra.fr
- Jacquemard Célien, jacquemard@unistra.fr
- Kastner Lise, lise.kastner@unicaen.fr
- Kellenberger Esther, ekellen@unistra.fr
- Kireev Dmitri, dmitri.kireev@missouri.edu
- Klimchuk Olga, o.klimchuk@unistra.fr
- Lagarde Nathalie, nathalie.lagarde@lecnam.net
- Lambert Heather, heather.lambert@pikairos.com
- Lambert Roman, romanlamb@outlook.fr
- Lamotte Jean-Luc, jean-luc.lamotte@unicaen.fr
- Langer Thierry, thierry.langer@univie.ac.at
- Le Corre Laurent, laurent.le-corre@parisdescartes.fr

- Le Guilloux Vincent, Vincent.leguilloux@discngine.com
- Lehembre Etienne, etienne.lehembre@unicaen.fr
- Lejal Vanille, vanille.lejal@etu.u-paris.fr
- Lejmi Maroua, maroua.lejmi@unicaen.fr
- Lepailleur Alban, alban.lepailleur@unicaen.fr
- Llompart Pierre, pierre.llompart@etu.unistra.fr
- Lozano Sylvain, slozano@oncodesign-services.com
- Mao Xiaojun, xiaojun.mao@univ-orleans.fr
- Marcou Gilles, g.marcou@unistra.fr
- Meddouri Nida, nida.meddouri@epita.fr
- Molina Christophe, contact@pikairos.com
- Molineau Jeremy, jeremy.molineau@univ-orleans.fr
- Montes Matthieu, matthieu.montes@cnam.fr
- Muller Christophe, chris.muller@evotec.com
- Ollitrault Guillaume, guillaume.ollitrault@gmail.com
- Ouali Abdelkader, abdelkader.ouali@unicaen.fr
- Oudahmane Mehdi, mehdi.oudahmane@hesam.eu
- Paccagnini Manon, Manon.Paccagnini@u-bourgogne.fr
- Payan Frédéric , fpayan@i3s.unice.fr
- Pikalyova Karina, karinapikalyova@gmail.com
- Pikalyova Regina, reginapikalyova@gmail.com
- Plyer Louis, louis.plyer@unistra.fr
- Rahali Inés Sabine, ines-sabine.rahali@u-paris.fr
- Reboul Etienne, etienne.reboul@etu.u-paris.fr
- Renault Nicolas, nicolas.renault@univ-lille.fr
- Revillo Imbernon Julia, revilloimbernon@etu.unistra.fr
- Roche Philippe, philippe.roche@inserm.fr
- Rognan Didier, rognan@unistra.fr
- Roucairol Milo, milo.roucairol@dauphine.eu
- Si Chaib Zeineb zeineb.sichaib@schrodinger.com
- Sindt Francois, f.sindt@unistra.fr
- Sopkova-de Oliveira Santos Jana, jana.sopkova@unicaen.fr
- Stanovic Stevan, stevan.stanovic@ensicaen.fr
- Taly Antoine, taly@ibpc.fr
- Urvas Lauri, urvas@unistra.fr
- Varnek Alexandre, varnek@unistra.fr
- Zimmermann Albrecht, albrecht.zimmermann@unicaen.fr