du Mercredi 29 Septembre au Vendredi 1er Octobre

CONGRÈS 2021

GGMM SFCi

LILLIAD
LEARNING CENTER
INNOVATION

VILLENEUVE D'ASCQ

TABLE OF CONTENTS

The organization of this event would not have been possible without the help of our sponsors

The event was organized in partnership with

FOREWORDS

It is a great pleasure to welcome you in Villeneuve d'Ascq for the joint GGMM and SFCi days. The two major meetings of the French molecular modeling community have been combined due to the (bio)hazards of the recent and unwelcomed health situation, but lady Luck and its lesser avatar, serendipity, are also a main drive for discovery. We hope you will enjoy the stay, the double dose of science in a single injection and gathering with friends old and new.

The organizing committee

Bonjour à toutes et tous !

Nous sommes très heureux de vous accueillir au centre LILLIAD de l'Université de Lille à Villeneuve d'Ascq pour une nouvelle odyssée d'un congrès couplé entre le Groupe Graphique et Modélisation Moléculaire et la Société Française de Chémoinformatique !

La période pandémique nous a tous éloigné les uns des autres nous obligeant à annuler certains évènements (l'école de modélisation du GGMM en 2020) ou à les décaler !!! Le GGMM se tenant habituellement au printemps, nous avons fait le choix de le proposer en cette période de fin Septembre pour espérer un mode présentiel, ou tout du moins hybride pour ceux qui ne voudraient ou ne pourraient venir ! Ce choix semble finalement le bon ! On va se retrouver pour parler sciences, modélisation, chémoinfo ! Et surtout pouvoir évoquer nos travaux respectifs, le tout dans une ambiance habituelle de bienveillance, de rires et de camaraderie.

Nous avons fait le choix d'associer le 23ème congrès du GGMM aux 10èmes journées de la SFCi pour vous proposer des sessions de deux communautés qui ne sont pas si éloignées et qui utilisent des méthodologies et des outils de bioinformatique structurale par des prismes différents mais en aucun cas opposés !

Dans cette optique, nous avons le plaisir de vous accueillir dans les Hauts de France pour ce jumelage scientifique. Nous sommes très heureux de vous proposer le programme retenu par le comité scientifique, et nous espérons que vous passerez d'excellents moments tout à la fois scientifiques et amicaux !

Bienvenue chez les Ch'tis !

Manuel DAUCHEZ & Matthieu MONTES, présidents du GGMM et de la SFCi

COMMITTEES

Organisation board

- Amaury Farce (MCU, Université de Lille, U1286)
- Nicolas Renault (IR, Université de Lille, U1286)
- Corentin Bedart (Doctorant, ParImmune, U1286)
- Manuel Dauchez (PU, Université de Reims Champagne-Ardennes, UMR 7369)
- Matthieu Montès (PU, CNAM, GBCM)

Scientific committee

- Samia Aci-Sèche (CR, CNRS, ICOA UMR 7311)
- Karine Audouze (MCU, Université de Paris, U1124)
- Laurent Chaloin (DR, CNRS, UMR 9004)
- Isaure Chauvot de Beauchene (CR, CNRS, LORIA)
- Philippe Chavatte (PU, Université de Lille, U1286)
- Nathalie Colloch (CR, CNRS, ISTCT UMR 6030)
- Manuel Dauchez (PU, Université de Reims Champagne-Ardennes, UMR 7369)
- Annick Dejaegere (PU, Université de Strasbourg, IGBMC)
- Gilles Marcou (MCU, Université de Strasbourg, UMR 7140)
- Claire Minoletti (Chercheur, Sanofi)
- Matthieu Montès (PU, CNAM, GBCM)
- Gautier Moroy (MCU, Université de Paris, U1133)

Detailed program

Wednesday, 29 September 2021

12:00 - 13:45	**Reception & Opening of «Journées GGMM»**	
13:45 - 14:00	**Welcome introduction**	
Session #1 : Visualization and graphism		
14:00 - 14:45	PEZESHKIAN Weria	Simulating Realistic Membrane Shapes
14:45 - 15:00	GALOCHKINA Tatiana	MEDUSA: web server for protein flexibility prediction from sequence
15:00 - 15:15	GONZALEZ-ALEMAN Roy	BitQT: A Graph-Based Approach to the Quality Threshold Clustering of Molecular Dynamics
15:15 - 15:30	BEDART Corentin	SINAPs: A software tool for analysis and visualization of interaction networks of molecular dynamics simulations
15:30 - 15:45	LANGENFELD Florent	Comparative Evaluation of Shape Retrieval Methods on Macromolecular Surfaces: An Application of Computer Vision Methods in Structural Bioinformatics
15:45 - 16:00	GELLY Jean-Christophe	Flexible protein structural alignment for non trivial comparisons
16:00 - 16:15	Flash posters - GGMM #1 - Session A	
16:15 - 17:00	Poster session & Coffee break	
Session #2 : Simulation of biosystems		
17:00 - 17:45	DE RUYCK Jérôme	How can oncoprotein Ets-1 interact with DNA repair enzyme PARP-1? A molecular modelling approach to design cancer progression inhibitors
17:45 - 18:00	BLANC Florian	Mechanism and energetics of proton-powered c-ring rotation in mitochondrial ATP synthase
18:00 - 18:15	FAGNEN Charline	PHF6 aggregation process responsible for Alzheimer's disease investigated by molecular dynamics
18:15 - 18:30	BARTOCCI Alessio	Identification of allosteric modulatory sites in the Glycine receptor by coarse-grained and atomistic Molecular Dynamics simulations
18:30 - 18:45	GHOULA Mariem	Molecular Dynamics Simulations reveal the conformational changes and the allosteric behavior in the human Insulin Degrading Enzyme
18:45 - 19:00	BELLAICHE Adam	Does temperature contribute to enhance the aggregation risk of antibodies? A molecular dynamics study on a representative biodrug
19:00 -	Dinner	

Thursday, 30 September 2021

Session #3 : Methodological developments		
08:45 - 09:00	ARAUJO-ROCHA Mario	Towards a theory-driven design of a DNA-based aptasensor
09:00 - 09:15	AUFFINGER Pascal	Revealing Short-Range Imbalances in the AMBER Lennard-Jones potential for Sugar...Base Lone-pair ... Pi contacts in Nucleic Acids
09:15 - 10:00	BARBE Sophie	Computational Protein Design
10:00 - 10:15	Flash posters - GGMM #2 - Session B	
10:15 - 11:00	Poster session & Coffee break	
11:00 - 11:15	BOUCHIBA Younes	Computational Design of miniprotein binders
11:15 - 11:30	TUBIANA Thibault	Mapping amino acids at protein-membrane interfaces to update the current membrane binding model
11:30 - 11:45	MILAN-RODRIGUEZ Paula	Amphipathic helix folding in membranes: Markov State Models to decipher the mechanism
11:45 - 14:00	**Opening of «Journées SFCI» & Lunch break**	
Session #4 : Integrative modeling		
14:00 - 14:45	MITEVA Maria	Integrated Mechanistic and Machine Learning Approach to predict Inhibitors of Drug Metabolizing Enzymes
14:45 - 15:00	Flash posters - SFCI #1 - Session C	
15:00 - 15:15	KHAKZAD Hamed	Integrative structural biology revealed how GAS M1 protein inhibits the IgG1 Fc-receptor binding site
15:15 - 15:30	ZABOLOTNA Yuliana	ChemSpace Atlas: empowering ultra-large library exploration
15:30 - 15:45	ALFERKH Lina	Biasing RNA coarse-grained folding simulations with Small-Angle X-ray Scattering (SAXS) data
15:45 - 16:00	DUDAS Balint	Insights into the transport mechanism of BCRP through Molecular Dynamics with excited Normal Modes simulations
16:00 - 16:15	SACQUIN-MORA Sophie	Investigating the interaction of enzymes with functionalized surfaces: Lessons from multiscale modeling approaches
16:15 - 16:45	Flash posters - SFCI #2 & PV - Session D	
16:45 - 17:30	Poster session & Coffee break	
Session #5 : Data-driven drug discovery		
17:30 - 18:15	DETROYER Ann	The rise of in silico methods in Cosmetic Safety Assessment : an industry perspective
18:15 - 18:30	BENKAIDALI Lydia	A machine learning model of CYP3A4 ligand selectivity based on geometric modeling of active site access channels
18:30 - 18:45	EGUIDA Merveille	Focused Library Design via Fragment-Bound Subpocket Alignment and Deep Generative Linking: A Proof-of-Concept for CDK8 Inhibitors
18:45 - 19:00	GHEERAERT Aria	Investigating conformational changes between perturbed molecular dynamics: a network approach
19:00 - 19:30	Break	
19:30 -	Extra dinner	

Friday, 1 October 2021

09:00 - 09:15	HOFFMANN Brice	Structure-guided de novo drug design using deep generative modeling, a case study
09:15 - 09:30	HLADIS Matej	Representation learning to overcome scarce data in machine learning. Application to chemosensory receptors
Session #6 : Current topics in chemoinformatics		
09:30 - 10:15	HORVATH Dragos	Big Data Fast Chemoinformatics Model to Predict Generalized Born Radius and Solvent Accessibility as a Function of Geometry
10:15 - 10:45	Coffee break	
10:45 - 11:00	PEYRAT Gautier	Application of Frags2Drugs for the fragment-based drug design of macrocyclic kinase inhibitors
11:00 - 11:15	REHIOUI Hajar	Improving SAR analysis via pharmacophoric feature selection and feature transformation
11:15 - 11:30	SELLAMI Asma	Predicting potential endocrine disrupting chemicals binding to Estrogen receptor α using a combination of structure based and ligand based in silico methods
11:30 - 11:45	TELLES de SOUZA Paulo	Drug Design with Martini 3 Coarse-Grained Model
11:45 - 12:00	TURK Joseph-André	Understanding Structure-Activity Relationship With Interpretability Methods for Molecular Activity and Chemical Property Prediction
12:00 - 12:30	GGMM award conference	
12:30 - 12:45	Poster awards & Closing	
12:45 - 14:00	Lunchbox & Board meetings : GGMM, SFCI, GDR BigDataChim	

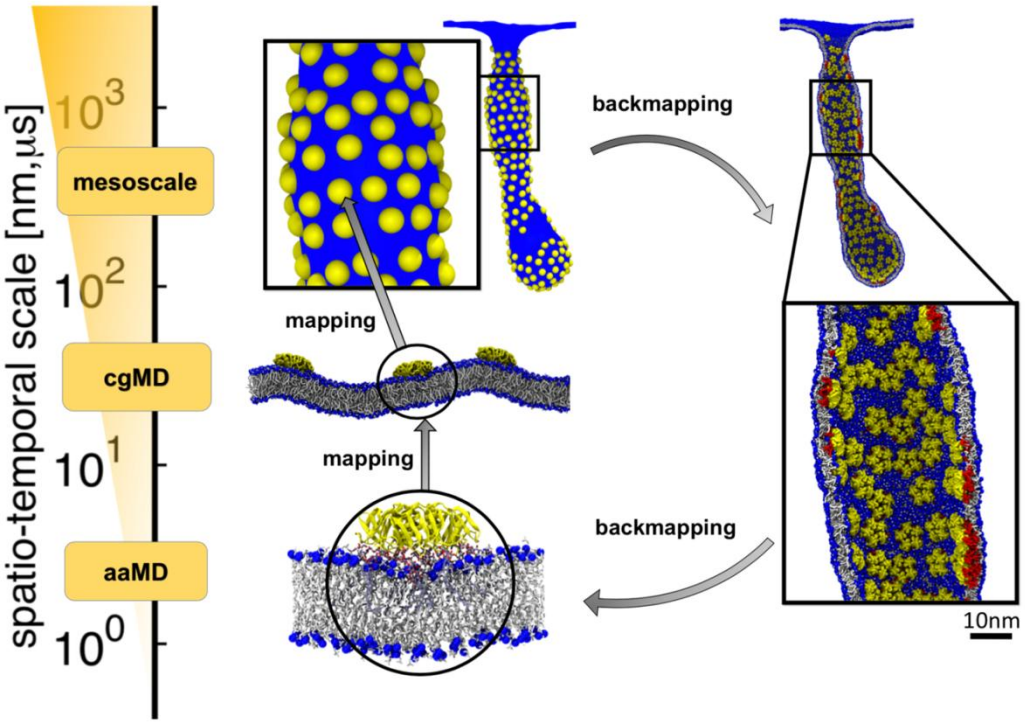
KEYNOTE LECTURES

KL1
Simulating Realistic Membrane Shapes

Weria PEZESHKIAN

Groningen Biomolecular Sciences and Biotechnology Institute and Zernike Institute for Advanced Materials, University of Groningen, Groningen, The Netherlands. w.pezeshkian@rug.nl

Abstract: Biomembranes are essential functional elements of the cell architecture, actively participating in vital cellular processes. A central feature of living cell membranes is their flexible shape that undergoes constant transformations at a length scale much greater than their thickness. Uncovering the mechanisms that underlie the shape remodelling of these macromolecular structures is essential for understanding their biological functions and providing fundamental scientific bases for combatting diseases, engineering artificial cells, and designing drug delivery vehicles. One of the important drivers of membrane remodelling is the cooperative action of membrane proteins. However, it remained unclear how exactly proteins play roles in achieving or stabilization the characteristic shape of cellular membranes. I will present our recent advances in exploring the coupling between membrane shape and lateral protein organizations using multiscale computer simulations.

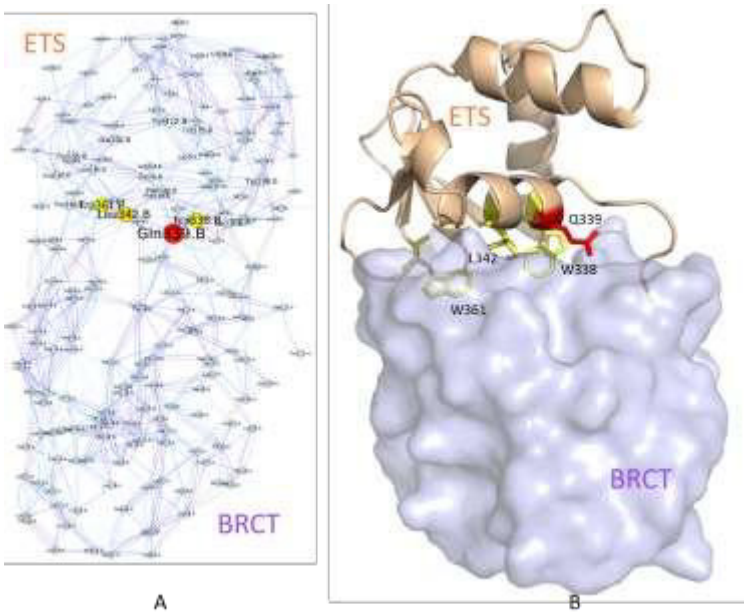


KL2
How can oncoprotein Ets-1 interact with DNA repair enzyme PARP-1? A molecular modelling approach to design cancer progression inhibitors

Jérôme de Ruyck, Guillaume Brysbaert, Marc Aumercier and Marc F. Lensink

Univ. Lille, CNRS UMR8576 UGSF – Unité de Glycobiologie Structurale et Fonctionelle, F-59000 Lille, France. jerome.de-ruyck@univ-lille.fr

Abstract: The Ets-1 oncoprotein is a transcription factor that promotes target gene expression in specific biological processes. Typically, Ets-1 activity is low in healthy cells, but elevated levels of expression have been found in cancerous cells, specifically related to tumor progression. Like the vast majority of the cellular effectors, Ets-1 does not act alone but in association with partners. Given the important role that is attributed to Ets-1 in major human diseases, it is crucial to identify its partners and characterize their interactions. In this context, two DNA repair enzymes, PARP-1 and DNA-PK, have been identified recently as interaction partners of Ets-1. We here identify their binding mode by means of protein docking. The results identify the interacting surface between Ets-1 and the two DNA repair enzymes centered on the α-helix H1 of the ETS domain, leaving α-helix H3 available to bind DNA. We rationalize the binding mode using a series of computational analyses, including alanine scanning, molecular dynamics simulation and residue centrality analysis. Our study constitutes a first but important step in the characterization, at the molecular level, of the interaction between an oncoprotein and DNA repair enzymes.



KL3
Critical role of environmental effects in regulation of the functioning of integral membrane proteins

A.A. Polyansky[1],4, P.E. Volynsky1, D.E. Nolde[1,2], A.S. Kuznetsov[1,2,3], Yu.A. Trofimov[1,3], N.A. Krylov[1,2], A.O. Chugunov[1,2,3], E.V. Bocharov[1,3], R.G. Efremov[1,2,3].

[1]Shemyakin–Ovchinnikov Institute of Bioorganic Chemistry RAS, Moscow, Russia; [2]Higher School of Economics, Moscow, Russia; [3]Moscow Institute of Physics and Technology (State University), Dolgoprudny, Russia; [4]University of Vienna, Vienna, Austria. r-efremov@yandex.ru

Abstract: In addition to the barrier role, biomembranes effectively perform a number of specific and finely regulated functions that make them not only indispensable for ensuring the normal functioning of the cell, but also extremely promising in terms of creating new generations of drugs. One of the most important properties of cell membranes is the modulation of the work of integral membrane proteins (MPs) due to the effects of the water-lipid environment on the structural and dynamic behavior of their transmembrane (TM) domains. In this work, based on the results of computational experiments, the physicochemical and dynamic properties of the TM domains of a number of receptor tyrosine kinases (RTK) and thermosensitive ion channels of the TRPV family embedded in model lipid membranes of various compositions are analyzed. It is established that local rearrangements of the membrane environment play an important role in the behavior of the TM domains of proteins, causing specific clustering and binding of lipids. It is demonstrated that the membrane environment seriously affects the transitions between different functional states of MPs. It is shown that lipids significantly contribute to the free energy of the association of TM helices, and this process has mainly entropic character. The detailed balance of the different energy contributions depends strongly on the membrane composition and the amino acid sequence of the protein. The "trigger" role of individual lipid molecules in the opening of the TRPV1 and TRPV3 ion channels was revealed. Thus, MPs and their water-lipid environment determine functioning of cell membranes, mutually strongly affecting each other and consistently reacting to external influences.

The work on the modeling of RTK TM domains is supported by the Russian Science Foundation (18-14-00375). The work on mapping the properties of ion channel pore domains was supported by the Russian Foundation for Basic Research (19-04-00350).

KL4
Computational Protein Design

Sophie BARBE

Toulouse Biotechnology Institute, TBI, Université de Toulouse, CNRS, INRAE, INSA, ANITI Toulouse, France. Sophie.Barbe@insa-toulouse.fr

Abstract: Proteins are responsible for almost all molecular processes that are essential for life on Earth. They have therefore undergone long-term natural optimization to adjust their properties to the requirements of living organisms. This optimization does not facilitate the use of natural proteins outside of their natural niche or for alternative purposes. Protein engineering strategies have thus been developed to tailor proteins to the specific demands generated by applications in health, biotechnologies, green chemistry, food & feed, and cosmetics for example. In most cases, protein engineering remains a daunting task, owing to the huge sequence space that needs to be explored especially when multiple-point mutants need to be considered. In the line of directed evolution (2018 Chemistry Nobel price), computational protein design aims at providing original proteins with improved or radically new capacities, but without the restraints of experimental approaches. Indeed, with 20 natural amino acids, designing even a simple protein of 100 amino acids requires to find a suitable amino acid sequence in a huge space of 20 100 possible sequences. A space from which only a minute fraction can be explored by experimental assays (usually far less than 10 12). In this talk, I will present our advances in the development of computational protein design methods based on hybrid techniques combining molecular modelling and artificial intelligence, focusing more particularly on the introduction of some recent functionalities allowing for multi-state design, data-driven design, binding energy estimation as well as diverse sequence library generation [1-5]. This will be illustrated with associated experimentally tested designs in the health and biotechnology fields.

1. Viricel C., de Givry S., Schiex T., Barbe S. Cost function network-based design of protein–protein interactions: predicting changes in binding affinity, Bioinformatics, Volume 34, Issue 15, 01 August 2018, Pages 2581–2589, https://doi.org/10.1093/bioinformatics/bty092
2. Simoncini D., Zhang KY J., Schiex T., Barbe S. A structural homology approach for computational protein design with flexible backbone, Bioinformatics, Volume 35, Issue 14, July 2019, Pages 2418–2426, https://doi.org/10.1093/bioinformatics/bty975
3. Vucinic J., Simoncini D., Ruffini M., Barbe S., Schiex T. Positive multistate protein design, Bioinformatics, Volume 36, Issue 1, 1 January 2020, Pages 122–130, https://doi.org/10.1093/bioinformatics/btz497
4. Ruffini M., Vucinic J., de Givry, S., Katsirelos G., Barbe S., Schiex T. Guaranteed Diversity and Optimality in Cost Function Network Based Computational Protein Design Methods. Algorithms 2021, 14, 168. https://doi.org/10.3390/a14060168
5. Bouchiba Y., Cortés J., Schiex T., Barbe S. Molecular flexibility in computational protein design: an algorithmic perspective, Protein Engineering, Design and Selection, Volume 34, 2021, gzab011, https://doi.org/10.1093/protein/gzab011

KL5
Integrated Mechanistic and Machine Learning Approach to predict Inhibitors of Drug Metabolizing Enzymes

Maria A. MITEVA

INSERM U1268 «Medicinal Chemistry and Translational Research », CiTCoM UMR 8038 CNRS - Univ. ParisFac., Pharmacie de Paris. maria.mitev@inserm.fr

Abstract: Drug metabolizing enzymes (DME) play a key role in the metabolism, elimination and detoxification of xenobiotics, drugs and endogenous molecules. While their principal role is to detoxify organisms by modifying compounds, such as pollutants or drugs in some cases they render their substrates more toxic thereby inducing adverse drug reactions, or their inhibition can lead to drug-drug interactions. We focused on Cytochrome P450 (CYP) responsible for the metabolism of 90 % drugs and on sulfotransferases (SULT), phase II conjugate drug metabolizing enzymes, acting on a large number of drugs, hormones and natural compounds. We established an original in silico approach that integrates structure-based and machine learning modeling and developed a new software DrugME to predict CYP and SULT inhibitors. This approach allowed the identification of new drug inhibitors and substrates of CYP2C9. Such strategy would improve the prediction of drug-drug interactions for clinical practice and drug development pipelines.

KL6
The rise of in silico methods in Cosmetic Safety Assessment: an industry perspective

Ann DETROYER

L'Oréal, Research & Innovation, 1 Avenue Eugène Schueller, Aulnay-sous-Bois, France. ann.detroyer@rd.loreal.com

Changes imposed by the international regulation (Reach and 7th Cosmetics Amendment) and the introduction of safety assessment notions in the early development stages of new ingredients have strongly accelerated the development of in vitro and in silico methods for safety assessment of ingredients. Developing these alternative methods is an objective of the L'Oréal Research group to resolve 21st century challenges on safety assessment. The cosmetic sphere has to take into account major specificities compared to other industries, owing to the large diversity of its products and ways of application. These include amongst others a wide and diverse variety of ingredients, physicochemical properties, and cover major toxicological endpoints related to skin. Few official recommendations are available with regards to in silico methods for assessing the safety of cosmetic ingredients, but their rise in this field is undoubtedly ongoing. We attempt, here, to expose some examples of applications of such methods in such context (some developed by L'Oréal Research and its collaborative partners*) together with the lessons learned with regards to questions on big data v. smart data, use of AI, computational chemistry/biology they include. Next, perspectives and opportunities will be discussed.

*ex. In silico mechanistically-based profiling module for acute oral toxicity ; D. Nedelcheva et al. ; https://doi.org/10.1016/j.comtox.2019.100109 ; Skin sensitisation testing in practice: Applying a stacking meta model to cosmetic ingredients; F.Tourneix et al. ; https://doi.org/10.1016/j.tiv.2020.104831

KL7
”Big Data" Fast Chemoinformatics Model to Predict Generalized Born Radius and Solvent Accessibility as a Function of Geometry

Dragos HORVATH[a*], Gilles MARCOU[a,] Alexandre VARNEK[a]

[a] Laboratory of Chemoinformatics, UMR 7140 University of Strasbourg/CNRS, 4 rue Blaise Pascal, 67000 Strasbourg, France. dhorvath@unistra.fr

The Generalized Born[1,2] (GB) solvent model is offering the best accuracy/computing effort ratio and yet requires drastic simplifications to estimate of the Effective Born Radii (EBR), in bypassing a too expensive volume integration step. EBR are a measure of the degree of burial of an atom, and not very sensitive to small changes of geometry. However, unlike dynamics simulations which proceed by fs steps hardly impacting on EBRs, stochastic sampling of by evolutionary algorithms (as performed by our S4MPLE[3,4] tool) must update them at each step. Therefore, a Quantitative Structure-Property Relationship (QSPR) has been developed in order to express the EBRs as a function of both topological neighborhood and the geometric occupancy of the space around atoms. A training set of 810 molecular systems, starting from fragment-like, to drug-like compounds, small proteins, host-guest systems and ligand-protein complexes has been compiled. For each species, S4MPLE generated several hundreds of random conformers. For each atom in each geometry of each species, its “standard” EBR was calculated by numeric integration and associated to topological and geometric descriptors of the atom neighborhood. This training set (EBR, atom descriptors) involving >5M entries was subjected to a boot-strapping multilinear regression process with descriptor selection. In parallel, the strategy was repurposed to also learn atomic solvent-accessible areas (SA), based on the same descriptors. Resulting linear equations were challenged to predict EBR and SA values, respectively, for a similarly compiled external set of >2,000 new molecular systems. Solvation energies calculated with estimated EBR and SA match "standard" energies within the typical error of a force-field based approach (a few kcal/mol). Given the extreme diversity of molecular systems covered by the model, this simple EBR/SA estimator covers a vast applicability domain.

1. Onufriev, A. V.; Case, D. A., Generalized Born Implicit Solvent Models for Biomolecules. Annual Review of Biophysics, Vol 48 2019, 48, 275-296.
2. Onufriev, A.; Case, D. A.; Bashford, D., Effective Born radii in the generalized Born approximation: The importance of being perfect. Journal of Computational Chemistry 2002, 23, 1297-1304.
3. Hoffer, L.; Saez-Ayala, M.; Horvath, D.; Varnek, A.; Morelli, X.; Roche, P., CovaDOTS: In Silico Chemistry-Driven Tool to Design Covalent Inhibitors Using a Linking Strategy. Journal of Chemical Information and Modeling 2019, 59, 1472-1485.
4. Hoffer, L.; Chira, C.; Marcou, G.; Varnek, A.; Horvath, D., S4MPLE-Sampler for Multiple Protein-Ligand Entities: Methodology and Rigid-Site Docking Benchmarking. Molecules (Basel, Switzerland) 2015, 20, 8997-9028.

GGMM award
Modelling protein interfaces using evolutionary information

Chloé Quignot[a], Hélène Bret[a], Pierre Granger[a], Julien Rey[b], Guillaume Postic[b], Samuel Murail[b], Pablo Chacón[c], Pierre Tufféry[b], Raphaël Guerois[a], Jessica Andreani[a]

[a] Université Paris-Saclay, CEA, CNRS, Institute for Integrative Biology of the Cell (I2BC), 91198 Gif-sur-Yvette, France
[b] Université de Paris, CNRS UMR 8251, INSERM U1133, RPBS, Paris 75205, France
[c] Department of Biological Physical Chemistry, Rocasolano Institute of Physical Chemistry C.S.I.C, Madrid, Spain

Protein complexes are of fundamental importance in most biological processes and mainly carry out their function in networks. The structure of their interface can give us crucial information to understand the mechanisms behind these processes. As the experimental determination of 3D complex structures is not always possible, in silico predictions are very helpful to study how two proteins interact. In protein-protein docking, we generate many possible interface models (sampling step) and score them in order to choose the most plausible ones.
My work focuses on the improvement of the prediction power of docking and scoring methods, in particular by drawing on co-evolutionary information[1]. Our docking server, InterEvDock[2,3], first benefitted from this information through InterEvScore[4], a scoring function combining evolutionary information through homologous sequence alignments at residue level with a simple coarse-grained statistical potential. In InterEvDock, InterEvScore is used together with complementary scores FRODOCK and SOAP-PP to output a consensus of 10 interface models. InterEvDock accepts oligomeric structure or sequence inputs. The latter are modelled automatically through homology modelling when a suitable template can be found for at least one of the subunits. The user can also integrate prior knowledge about the interaction as constraints.
More recently, I implemented a more efficient and higher-resolution way of integrating evolutionary information into scoring. I managed to derive the implicit evolutionary information present in the sequence alignments to an atomic level of detail using modelled homologous interfaces. This explicit representation is directly compatible with atomic-scale scoring and significantly increases the success in predictive performance from 32% to 40% on a large benchmark[5] with time scales similar to the previous versions of InterEvDock[6]. Additional evolutionary information can be extracted from covariation-based contact maps used by InterEvDock as an alternative scoring scheme during docking.
Our pipeline has shown its value in many applications such as collaborations with experimentalists or as a tool in our "human participation" in the international CAPRI (Critical Assessment of Predicted Interactions) protein docking challenge[7] for which our group ranked first in the 2016-2019 edition. We are constantly working on improving our server towards a more efficient and user-friendly service. Further integration of DCA-like methods and other coevolution-based methods with machine learning and deep learning algorithms will likely prompt future progress and expand the range of applications.

[1] Andreani J., Quignot C., Guerois R. Structural prediction of protein interactions and docking using conservation and coevolution. WIREs Comput. Mol. Sci. 2020; 10:e1470.
[2] Quignot C., Rey J., Yu J., Tufféry P., Guerois R., Andreani J. InterEvDock2: an expanded server for protein docking using evolutionary and biological information from homology models and multimeric inputs. Nucleic Acids Res. 2018; 46:W408–W416.
[3] Quignot C., Postic G., Bret H., Rey J., Granger P., Murail S., Chacón P., Andreani J., Tufféry P., Guerois R. InterEvDock3: a combined template-based and free docking server with increased performance through explicit modeling of complex homologs and integration of covariation-based contact maps. Nucleic Acids Res. 2021; 49:W277–W284.
[4] Andreani J., Faure G., Guerois R. InterEvScore: a novel coarse-grained interface scoring function using a multi-body statistical potential coupled to evolution. Bioinformatics. 2013; 29:1742–1749.
[5] Yu J., Guerois R. PPI4DOCK: large scale assessment of the use of homology models in free docking over more than 1000 realistic targets. Bioinformatics. 2016; 32:3760–3767.
[6] Quignot C., Granger P., Chacón P., Guerois R., Andreani J. Atomic-level evolutionary information improves protein-protein interface scoring. Bioinformatics. 2021
[7] Nadaradjane A.A., Quignot C., Traoré S., Andreani J., Guerois R. Docking proteins and peptides under evolutionary constraints in critical assessment of prediction of interactions rounds 38 to 45. Proteins. 2020; 88:986–998.

ORAL COMMUNICATIONS

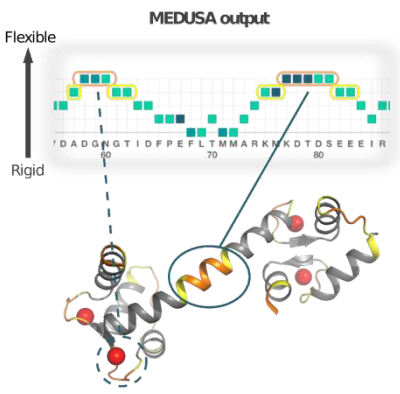
OC1
MEDUSA: web server for protein flexibility prediction from sequence

Yann Vander Meersche[1,2], Gabriel Cretin[1,2], Alexandre G. de Brevern[1,2], Jean-Christophe Gelly[1,2] and Tatiana Galochkina[1,2]

[1] Université de Paris, Inserm UMR_S 1134 BIGR, INTS, 6 rue Alexandre Cabanel, 75015 Paris, France
[2] Laboratoire d'Excellence GR-Ex, 75015 Paris, France

Background: Information on protein structural and dynamical properties is essential to understand their mechanism of function. Flexibility of different protein regions has a crucial impact on protein stability, as well as on interactions with other molecules. Due to complexity and high price of the experimental description of protein structure, the gap between a number of known protein sequences and a number of proteins with available flexibility information continues to grow. Data-driven computational tools have potential to provide important information on proteins without resolved structure by predicting protein flexibility profile directly from amino acid sequence.
Objectives
In the current study we developed a deep learning based prediction tool named MEDUSA for the protein flexibility prediction from amino acid sequence.

Methods: We have considered protein flexibility in terms of the B-factor obtained in the X-ray crystallography. For each position of the amino acid sequence we predict its expected flexibility in two, three and five classes using deep convolutional neural networks. The protein sequences were encoded using evolutionary information extracted from homologous protein sequences, as well as physico-chemical properties of individual amino acids. The performance of our model was estimated in 10-fold cross validation on the dataset filtered by structural similarity to ensure independence between train and test datasets.



Results and conclusion: We have implemented MEDUSA[1] as a web server (https://www.dsimb.inserm.fr/MEDUSA) as well as a standalone utility (https://github.com/DSIMB/medusa). MEDUSA outperforms the state-of-the-art method PROFbval[2] for the two-class prediction problem. MEDUSA successfully identifies potentially deformable protein regions for the proteins with known dynamical properties (such as calmodulin shown in figure from Vander Meersche et al. 2021). Furthermore, MEDUSA also detects presence of the locally rigid regions for the proteins without stable fold. Therefore, MEDUSA predictions complement the information provided by the disorder prediction tools.

References

1. Y. Vander Meersche, G. Cretin, A.G. de Brevern, J.-C. Gelly, T. Galochkina (2021) MEDUSA: Prediction of Protein Flexibility from Sequence, Journal of Molecular biology 433:11, 166882.

2. A. Schlessinger, G. Yachdav, B. Rost. PROFbval: predict flexible and rigid residues in proteins (2006). Bioinformatics 22:7, 891–893.

OC2
BitQT: A Graph-Based Approach to the Quality Threshold Clustering of Molecular Dynamics

Roy González-Alemán[a, b], Daniel Platero-Rochart[a], David Hernández-Castillo[c], Erix W. Hernández-Rodríguez[d], Julio Caballero[e], Fabrice Leclerc[b], and Luis Montero-Cabrera[a]

[a] Laboratorio de Química Computacional y Teórica (LQCT), Facultad de Química, Universidad de La Habana, La Habana, 10400, Cuba
[b] Institute for Integrative Biology of the Cell (I2BC), CEA, CNRS, Université Paris Saclay, Gif-sur- Yvette, F-91198, France.
[c] Institute of Theoretical Chemistry, University of Vienna, Währinger Str. 17, 1090 Vienna, Austria.
[d] Laboratorio de Bioinformática y Química Computacional, Escuela de Química y Farmacia, Facultad de Medicina, Universidad Católica del Maule, 3460000 Talca, Chile.
[e] Departamento de Bioinformática, Facultad de Ingeniería, Centro de Bioinformática, Simulación y Modelado (CBSM), Universidad de Talca, Talca, Chile.

Classical Molecular Dynamics is a standard computational approach to model time-dependent processes at the atomic level. The inherent sparsity of increasingly huge generated trajectories demands clustering algorithms to reduce other post-simulation analysis complexity. The quality threshold (QT) variant is an appealing one from the vast number of available clustering methods. It guarantees that all members of a particular cluster will maintain a collective similarity established by a user-defined threshold. Unfortunately, its high computational cost for processing big data limits its application in the molecular simulation field. In the present work, we propose a methodological parallel between QT clustering and another well-known algorithm in the field of Graph Theory, the Maximum Clique Problem. Molecular trajectories are represented as graphs whose nodes designate conformations, while unweighted edges indicate mutual similarity between nodes. The use of a binary-encoded RMSD matrix coupled to the exploitation of bitwise operations to extract clusters significantly contributes to reaching a very affordable algorithm compared to the few implementations of QT for Molecular Dynamics available in the literature. Our alternative provides results in good agreement with the exact one while strictly preserving the collective similarity of clusters. The source code and documentation of BitQT are free and publicly available on GitHub (https://github.com/LQCT/BitQT.git) and ReadTheDocs (https://bitqt.readthedocs.io/en/latest/) respectively.

OC3
SINAPs: A software tool for analysis and visualization of interaction networks of molecular dynamics simulations

Corentin BEDART [a,b], Nicolas RENAULT [b], Philippe CHAVATTE [b], Adeline PORCHERIE [a], Abderrahim LACHGAR [a], Monique CAPRON [a], Amaury FARCE [b]

[a] Par'Immune, Bio-incubateur Eurasanté, 70 rue du Dr. Yersin, 59120, Loos-Lez-Lille, France
[b] Univ. Lille, Inserm, CHU Lille, U1286 - Infinite - Institute for Translational Research in Inflammation, F-59000 Lille, France

Introduction. To study enzymatic mechanisms of complex proteins, we had to perform many molecular dynamics simulations generating a huge amount of data. The analysis of such an amount of data to identify the main differences is a bottleneck. The simple visualization of the results is also a complex point. However, since there was no appropriate tool to perform these steps together, we developed a software written in Python and we named it SINAPs (Structural Interaction Networks Analysis Protocols). This proprietary tool allows us to identify either similarities or differences of interactions and lead us to visualize the results through an extension to UCSF Chimera.

Methods. The analysis module takes as input molecular dynamics simulations or crystallographic structures, to calculate non-bonded interactions with precise control of the definition parameters for each type. The visualization module allows the observation of the results by displaying each type of interaction via a finely controlled mode of representation. The SINAPs software was validated on several biological targets described in the literature: GLUT-1 and A2AR. In addition, an exploratory study was done using classical molecular dynamics of Par'Immune's drug candidate, P28GST, to assist the study of its molecular mechanism of action.

Results. The molecular dynamics study of two conformations of the Glucose Transporter 1 allowed us to identify the main interaction networks previously described in the literature. We specifically pinpointed the isolation mechanism of the ligand-binding site from the extracellular environment, or the mechanism allowing the opening of the ligand-binding site towards the intracellular side. The study of crystallographic conformations of the Adenosine A2A receptor allowed us to show the main interactions governing the activation steps and ligand-binding mode, confirming the similarity of the interactions from different agonists, and identifying the differential interactions made by an inhibitor. The study of P28GST using SINAPs validated the molecular dynamics simulations performed by finding specific interactions already described in the literature, supporting assumptions regarding the enzymatic mechanism of glutathione activation, and allowing to highlight several amino acids impacting this activation.

Conclusion. SINAPs, which will be soon available to the scientific community, may accelerate the structural analysis of different protein conformations by highlighting the similarities and/or differences of non-covalent molecular interactions while suggesting a simplified and customizable visualization.

OC4
Comparative Evaluation of Shape Retrieval Methods on Macromolecular Surfaces: An Application of Computer Vision Methods in Structural Bioinformatics.

Mohamed Machat[a], Florent Langenfeld[a], Daniela Craciun[a], Léa Sirugue[a], Taou%k Labib[a], Nathalie Lagarde[a], Maxime Maria[a], b and Mathieu Montes[a].

[a] Laboratoire GBCM, EA 7528, Conservatoire National des Arts et Métiers, Hesam université, 2 rue Conté, 75003 Paris, France
[b] Laboratoire XLIM, UMR CNRS 7252, Université de Limoges, 123 avenue Albert Thomas, 87000 Limoges, France

Motivation: The investigation of the structure of biological systems at the molecular level gives insight about their functions and dynamics. Shape and surface of biomolecules are fundamental to molecular recognition events. Characterizing their geometry can lead to more adequate predictions of their interactions. In the present work, we assess the performance of reference shape retrieval methods from the computer vision community on protein shapes.

Methods: Using an established dataset[1], we assess the ability of 4 shape retrieval methods (VFH[2], PANORAMA[3], 3DSurfer[4] and ShapeDNA[5]) from the computer vision field to retrieve the protein classification from the SCOPe database[6]. We then compare their performance to 3 well established algorithms (CE[7], DeepAlign[8] and TM-Align[9]) from the structural bioinformatics field. The dataset is composed of 5298 protein conformations extracted from 211 NMR structures of the Protein Data Bank[10].

Results: Shape retrieval methods are efficient in identifying orthologous proteins (i.e. proteins having the same function in different organisms) and tracking large conformational changes where the structure alignment tools fail. Furthermore, the shape retrieval methods are able to detect surficial homologs (i.e. proteins with similar surfaces but different folds). This work illustrates the interest for the protein surface shape as a higher-level representation of the protein structure that 1) abstracts the underlying protein sequence, structure or fold, 2) allows the use of shape retrieval methods to screen large databases of protein structures to identify sur%cial homologs and possible interacting partners, 3) opens an extension of the protein structure-function paradigm towards a protein structure-surface(s)-function paradigm.

Availability: All data are available online at http://datasetmachat.drugdesign.fr

1 Langenfeld F, Axenopoulos A, Benhabiles H, Daras P, GiacheG A, Han X, Hammoudi K, Kihara D, Lai T M, Liu H, Melkemi M, Mylonas S K, Terashi G, Wang Y, Windal F, Montes M. Protein Shape Retrieval Contest. Eurographics Workshop on 3D Object Retrieval, 2019, 25-31.
2 Rusu RB, Bradski G, Thibaux R, Hsu J. Fast 3d recognition and pose using the viewpoint feature histogram. 2010 IEEE/RSJ Interna#onal Conference on Intelligent Robots and Systems, 2010, 2155-2162.
3 Papadakis P, Pratikakis I, Theoharis T and Perantonis S. PANORAMA: A 3D shape descriptor based on panoramic views for unsupervised 3D object retrieval. Int J Comput Vis, 2010, 177-192.
4 La D, Esquivel-Rodríguez J, Venkatraman V, Li B, Sael L, Ueng S, Ahrendt S, Kihara D.
3D-SURFER: soMware for high- throughput protein surface comparison and analysis. Bioinforma#cs, 2009, 2843-2844.
5 Reuter M, Wolter F-E, Peinecke N. Laplace-Beltrami spectra as 'Shape-DNA' of surfaces and solids. Comput Aided Des, 2006, 342-366.
6 Fox N K, Brenner S E, Chandonia J M. Nucleic Acids Res, 2014 , D304-309
7 Shindyalov I N, Bourne P E. Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. Protein Eng, 1998, 739-747.
8 Wang S, Ma J, Peng J, Xu J. Protein structure alignment beyond spatial proximity. Sci Rep. 2013, 1448.
9 Zhang Y, Skolnick J. TM-align: a protein structure alignment algorithm based on the TM-score. Nucleic Acids Res. 2005, 2302-2309.
10 Berman H M, Westbrook J, Feng Z, Gilliland G, Bhat T N, Weissig H, Shindyalov I N, Bourne P E. The Protein Data Bank. Nucleic Acids Res. 2000, 235-242.

OC5
Flexible protein structural alignment for non trivial comparisons

Gabriel Cretin[a,b], Charlotte Périn[a,b], Nicolas Zimmermann[a,b], Tatiana Galochkina[a,b], Jean-Christophe Gelly[a,b]

[a] Université de Paris, Inserm UMR_S 1134 BIGR, INTS, 6 rue Alexandre Cabanel, 75015 Paris, France
[b] Laboratoire d'Excellence GR-Ex, 75015 Paris, France

Background and objectives: Accurate protein structure superposition is crucial for the detection of evolutionary relationships between proteins and for the analysis of their impact on structure and function. Therefore, protein structural alignment remains one of the most essential techniques in structural bioinformatics. Classical computational approaches rely on rigid body superposition. However, they perform poorly for proteins with high intrinsic flexibility or related by evolutionary processes changing the organization of the protein architecture. Here we propose ICARUS, a new method for flexible structural alignment based on the Protein Peeling algorithm[1].

Methods: The ICARUS algorithm is based on iterative alignments of small compact regions called Protein Units (PUs) identified by the Protein Peeling1 algorithm. ICARUS builds a number of iterative rigid body alignments of PUs of the query on the target structure and then switches target and query. Once all the intermediate alignments are explored, the best global alignment is chosen on the basis of the TM-score normalized by the length of the smallest protein.

Results: ICARUS outperforms the traditional structural alignment tools for the most difficult structural alignment cases taken from the RIPC (Repetitions, large InDels, circular Permutations and Conformational variability) dataset[2]. ICARUS obtains an average TM-score of 0.74, while TM-align[3] and FATCAT[4] obtain 0,53 and 0,66 respectively. Furthermore, among other tested tools ICARUS is the only tool able to detect structural similarity even for the particularly complex targets of the RIPC dataset and the obtained alignment scores never fell below 0.5.

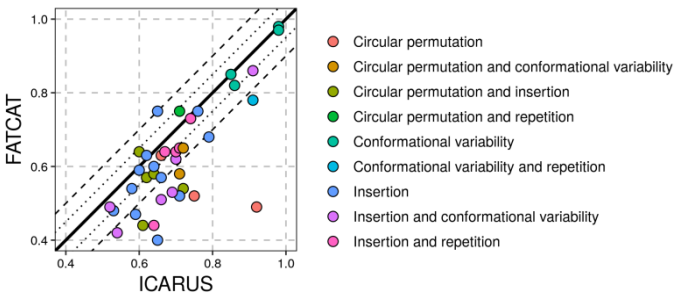


Fig. 1. Comparison of structural alignment TM-scores obtained by FATCAT and ICARUS on all pairs of proteins from the RIPC dataset.

1 Gelly, J.-C. et al: "Protein Peeling": an approach for splitting a 3D protein structure into compact fragments. Bioinformatics. 22, 129–133 (2006)
2 Mayr, G. et al: Comparative analysis of protein structure alignments. BMC Struct. Biol. 7, 50 (2007)
3 Zhang, Y. et al: TM-align: a protein structure alignment algorithm based on the TM-score. Nucleic Acids Res. 33, 2302–2309 (2005)
4 Ye, Y. et al: Flexible structure alignment by chaining aligned fragment pairs allowing twists. Bioinformatics. 19 Suppl 2, ii246–55 (2003)

OC6
Mechanism and energetics of proton-powered c-ring rotation in mitochondrial ATP synthase

Florian E. C. Blanc[a], Gerhard Hummer[a,b]

[a] Theoretical Biophysics Department, Max Planck Institute of Biophysics, Frankfurt am Main, Germany

[b] Institute of Biophysics, Goethe University Frankfurt, Frankfurt am Main, Germany

Background
F_o-F_1 ATP synthase is the enzyme complex responsible for ATP synthesis at the end of the mitochondrial respiratory chain, making it a central actor for cellular metabolism. It combines two rotary molecular motors which harness the spontaneous flow of protons across the inner mitochondrial membrane (IMM) to catalyse ATP synthesis. The transmembrane motor, or F_o region, consists of the c-ring and the a-subunit. An access half-channel allows protons to flow from the intermembrane space to the a-subunit/c-ring interface and bind an accepting Glutamate on the c-ring. The neutralised Glutamate can then insert into the membrane, promoting directional rotation of the c-ring against the a-subunit. APer a full rotation, the Glutamate enters the exit half-channel on the matrix side where higher pH promotes proton release. Torque from proton-driven rotation powers ATP synthesis by the second motor, F_1.

Objectives
The detailed structural mechanism of proton-powered rotation of the c-ring has long been inaccessible due to lack of high-resolution structures of the Fo region. Capitalising on recent cryo-EM structures, we aim at a structural and energetic description of the mechanism to understand how functional directional rotation is achieved.

Methods
We built all-atom models of the F_o region of *Polytomella sp*. mitochondrial ATP synthase, covering several protonation states for the accepting Glutamates and embedded into a realistic IMM including cardiolipin. To investigate the rotation mechanism, we performed steered MD simulations in which torque was applied on the ring. In addition, to explore how the differences in protonation affect the energetics of rotation, we ran multi-μs extended Adaptive Biasing Force (eABF) calculations to determine protonation-dependent free en- ergy profiles along the c-ring rotation angle.

Results and conclusions
Steered MD and eABF simulations reveal that rotation proceeds by dynamic sliding of the ring over the a- subunit surface, during which interactions with conserved polar residues are formed sequentially and sta- bilise intermediates along the rotation step. Free energy calculations show that rotation of the "rotation- primed" protonation state occurs spontaneously along the free energy gradient, and support the role of the conserved, essential a-subunit Arginine in stabilising the rotated configuration by forming a salt-bridge with the charged c-ring Glutamate. Overall, we describe the functional mechanism and energetics of an impor- tant proton-powered transmembrane molecular motor with an unprecedented level of details.

OC7
PHF6 aggregation process responsible for Alzheimer's disease investigated by molecular dynamics

Charline Fagnen[a], Johanna Giovannini[a], Anne Sophie Voisin-Chiret[a] , Jana Sopkova-de Oliveira Santos[a]

[a] Normandie University UNICAEN, EA 4258 CERMN (Centre d'Etudes et de Recherche sur le Médicament de Normandie) FR CNRS 3038 INC3M SF4206 ICORE, Boulevard Henri Becquerel 14032 Caen

Described in 1906 by Alois Alzheimer[1], Alzheimer's disease is a neurodegenerative disorder that represents 60 to 80% of dementia cases. Without treatment today, predictions reveal that its prevalence will be multiplied by four until 2050. This pathology is caused by the aggregations of Tau protein and amyloids that disrupt the neuronal signal. It is why we focus on Tau protein, more especially a short but crucial sequence called PHF6: 306VQIVYK311 responsible for its own aggregation. Our goal is to understand this aggregation mechanism at the atomic scale.

To do it, 2 microseconds molecular dynamics (MD) simulations were performed, including 62 PHF6 peptides. MD simulations showed that the PHF6 aggregation is initiated by a dimer aggregation, followed by a trimer, a tetramer… Next, the assembly of multimers took place. The PHF6 aggregation mechanism is not known to an atomic level, and our goal is to understand it. A recent publication[2] describing the PHF6 dimerization has given some clues about the beginning of its mechanism. This knowledge is essential to understand how to prevent these phenomena for the treatment of Alzheimer's disease.

This work focuses on the different processes: i) dimerization, ii) trimerization, iii) tetramerization and iv) complexation to gain a better comprehension of PHF6 aggregation. To evaluate each process, we led energetic and structural studies.
Two aggregation mechanisms are detected, showing the crucial role of the hydrogen bonds. Two main sets of different hydrogen bonds were detected: the first regrouping the hydrogen bonds involved in the parallel aggregation, the second regrouping the hydrogen bonds involved in the anti-parallel aggregation. We noticed that the aggregation initiation was realized with no preference mechanism, but more a multimer grows, more it promotes the parallel aggregation. The study of the complexation showed the essential role of the Tyr310 in the aggregation through a m-stacking interaction. The hydrogen bonds and m-stacking interaction are excellent targets to design disruptors to prevent the tau aggregation.

1 Berchtold N. C. Evolution in the Conceptualization of Dementia and Alzheimer's Disease: Greco-Roman Period to the 1960s, Neurology of Aging, 1998, 173-189

2 Liu H. Disclosing the Mechanism of Spontaneous Aggregation and Template-Induced Misfolding of the Key Hexapeptide (PHF6) of Tau Protein Based on Molecular Dynamics Simulation, ACS Chemical Neurosciences, 2020, 4810-4823

OC8
Identification of allosteric modulatory sites in the Glycine receptor by coarse- grained and atomistic Molecular Dynamics simulations

Alessio Bartocci [a] and Marco Cecchini [a]

[a] Laboratoire d'Ingénierie des Fonctions Moléculaires (IFM), Faculté de Chimie, Université de Strasbourg, 4 rue Blaise Pascal, 67000, Strasbourg

E-mail: abartocci@unistra.fr

Glycine receptors (GlyRs) are pentameric ligand-gated ion channels that play a critical role in motor coordination and essential sensory functions such as vision and audition [1]. They are transmembrane protein assemblies that have been recognized as pharmacological targets for pain [2]. Here, we present an original simulation strategy based on Coarse-Grained (CG) and all-atom Molecular Dynamics to aid the design of positive and negative allosteric modulators of GlyRs. First, "flooding" CG simulations are carried out to identify putative binding sites in the transmembrane domain of the receptor, giving us an idea of the interacting associative and dissociative mechanisms. Second, all-atom MD started from statistically relevant binding modes is carried out to refine these recognition events and sample the energetics of the protein- ligand interaction. Last, direct evaluations based on the MM/PBSA free energy method are used out to estimate the protein-ligand binding affinity, and prioritize the most relevant binding mode. In the current protocol, molecular parameters were taken from MARTINI [3] and the CHARMM-36 [4] force field, whereas simulations were carried with GROMACS-2021 [5]. To validate the approach, cholesterol was used as a test case, which is natively parameterized in MARTINI [6]. The simulation results indicate that the main interaction site for cholesterol is an inter-subunit pocket located between the transmembrane helices M1 and M3. This pocket largely overlaps with the crystallographic binding site of ivermectin (IVM) [7], i.e. a potent allosteric modulator of GlyR, and is consistent with recent CG simulations [8]. This analysis is being extended to a database of 218 modulatory ligands of GlyR, which was recently published by us [1]. These studies provide fundamental insights on the ligand-receptor interactions in the lipid membrane and are useful to guide the conception of novel allosteric modulators of brain receptors.

[1] A.H. Cerdan, M. Sisquellas, G. Pereira, D.E. Barreto Gomes, J.P. Changeux, and M. Cecchini, "Bioinformatics", 36, 2020, 3376–3384
[2] J.W. Lynch, Y. Zhang, S. Talwar, and A. Estrada-Mondragon, "Advances in Pharmacology", 79, 2017, 225—253
[3] S.J. Marrink, J. Phys. Chem. B., 111, 2007, 7812—7824
[4] J.B. Klauda, J. Phys. Chem. B., 114, 2010, 7830—7843
[5] https://doi.org/10.5281/zenodo.4457626
[6] MN. Melo, HI. Ingolfsson, and SJ. Marrink, "JCP", 143, 2015, 243152
[7] X. Huang, H. Chen, and P.L. Shaffer, Structure, 25, 2017, 945—950
[8] M.A. Damgen, and P.C. Biggin, PLOS, 17, 2021, e1007856

OC9
Molecular Dynamics Simulations reveal the conformational changes and the allosteric behavior in the human Insulin Degrading Enzyme

M. Ghoula[1], G. Postic[1], A-C Camproux[1], G. Moroy[1]

[1] Université de Paris, BFA, UMR 8251, CNRS, ERL U1133, Inserm,
F-75013 Paris, France

mariem.ghoula@inserm.fr

Abstract

Insulin Degrading Enzyme (IDE) is a metallopeptidase that degrades a large panel of amyloidogenic peptides and is thought to be a potential therapeutic target for type-2 diabetes and neurogenerative diseases like Alzheimer's disease [1]. Interestingly, IDE is a cryptidase. Its catalytic chamber, known as a crypt, is formed so that peptides can be enclosed and degraded [2]. However, the molecular mechanism of IDE remains elusive. It has been shown that IDE undergoes several conformational switches between closed and open states in order to regulate peptide degradation [3]. Thereby, it is essential to unfold IDE mechanism and provide more information on how conformational dynamics can modulate the catalytic cycle of IDE.

In this aim, a free-substrate IDE crystallographic structure (PDB ID: 2JG4) was used to build a complete structure of IDE with the MODELLER software [4]. IDE stability and flexibility were studied through Molecular Dynamics simulations with the GROMACS software [5] and the CHARMM36m force field [6]. In total, we ran 7 simulations of 1!s each to cover a wide range of the IDE conformational space. The crypt volume as well as the Solvent Accessible Surface Area (SASA) were calculated to witness IDE conformational dynamics switching from a closed to an open state. The Gibbs free energy landscapes were also investigated to indicate the different conformational states accessible to the protein during the simulations. The Molecular Mechanics/Poisson-Boltzmann Surface Area (MM/PBSA) method was used to identify key residues involved in IDE rearrangement.

References

1. Tang WJ. Targeting Insulin-Degrading Enzyme to Treat Type 2 Diabetes Mellitus. Trends Endocrinol Metab. 2016;27(1):24-34.
2. Shen Y, Joachimiak A, Rosner MR, Tang WJ. Structures of human insulin-degrading enzyme reveal a new substrate recognition mechanism. Nature. 2006;443(7113):870-874.
3. B. Webb, A. Sali. Comparative Protein Structure Modeling Using Modeller. Current Protocols in Bioinformatics 54, John Wiley & Sons, Inc., 5.6.1-5.6.37, 2016.
4. M.J. Abraham, D. van der Spoel, E. Lindahl, B. Hess, and the GROMACS development team, GROMACS User Manual version 2019
5. Huang, J., Rauscher, S., Nawrocki, G., Ran, T., Feig, M, de Groot, B.L., Grubmuller, H., and MacKerell, A.D., Jr., "CHARMM36m: An Improved Force Field for Folded and Intrinsically Disordered Proteins," Nature Methods, 14:71-73, 2016.

OC10
Does temperature contribute to enhance the aggregation risk of antibodies? A molecular dynamics study on a representative biodrug.

Adam Bellaïche[a], David Zanfack[b], Catherine Etchebest[c] .

[a,c]: Equipe DSIMB, UMR-S1134, Université de Paris, Inserm, 6 Rue Alexandre Cabanel, Paris, 75015
[a,b]: Equipe R&D, Audensiel Healthcare, 93 rue Nationale, Boulogne-Billancourt, 92100

[a]adam.bellaiche@etu.u-paris.fr
[b]d.zanfack@audensiel.fr
[c]catherine.etchebest@inserm.fr

Monoclonal antibodies (mAbs), especially IgG molecules, are the best-selling biopharmaceutical class. Percentage of yield products at the end of their manufacturing is impacted by environmental physical-chemical conditions. Indeed, along the production process, mAbs endure stress due to variations in environmental factors. This can lead to a local even a global denaturation of their structure, which can result in aggregate formation, reducing the yield product accordingly. Temperature and pH are reputed to have the biggest impact on antibodies stability but how these factors affect the structural and dynamics of antibodies remains unclear. Thus, a better mechanistic understanding and characterization of the temperature and pH impact at a molecular level is needed to guide antibodies engineering, formulation et storage. In the present work, we first focus on temperature impact on a well-known biodrug, the pembrolizumab.

We started with the protein X-ray atomistic structure and completed the missing parts. From this model, we launched 12 molecular simulations of 1 µs containing a monomer of pembrolizumab at temperatures from 253K to 330K, a temperature range that is retrieved along production and storage steps. Pembrolizumab was solvated in tip3p or tip4p (at 253K) water and 0.1M of NaCl box. Glycerol was also used as a solvent at 253 K. The simulations were run using Gromacs, following a classical MD protocol. Several structural parameters were examined, e.g evolution of native contacts, stability of sub-domains, orientations and motions of these sub-domains. Overall, the results show that the antibody is able to adopt new conformations even at low temperature. Indeed, we observe a particular movement at 273K in the FC region. Surprisingly, we also find a difference between in-vivo temperature (310K) and in-vitro temperature (300K) behaviour. However, no major unfolding was observed, which confirms the antibody fold stability to a large temperature variation.

Besides these global parameters, we paid attention to more local structural features in order to identify aggregation prone regions (APRs). We used the AGGRESCAN3D scoring function based on the accessibility, the physico-chemical properties and evaluated the APRs along the dynamics. Interestingly, the results highlight some important functional regions that seem to be impacted by the temperature.

In summary, this study presents for the first time, the impact of temperature on the structure and dynamics of a complete antibody. It offers a new view of the mechanical properties of this important family of proteins and revisits its aggregation properties at the light of the local dynamics.

OC11
Towards a theory-driven design of a DNA-based aptasensor

Mario Araujo-Rocha[a], Benoît Piro[a], Vincent Noël[a] and Florent Barbault[a]

[a] Laboratoire ITODYS, CNRS UMR 7086, Université de Paris, 15 rue de Baïf, 75013 Paris

Liquid Electrolyte-Gated Organic Field-Effect Transistors (EGOFETs) are organic transistors for which an electrolytic solution acts as the dielectric material. These devices, due to their simple structure, are easy to fabricate and miniaturize, and are sensitive to their ionic environment which makes them ideal biosensing devices. Their functioning relies on the capacitance change at the gate/ electrolyte interface in the presence of a target molecule, which can be achieved by functionalizing the gate surface with a biorecognition probe, such as an antibody or a DNA aptamer.

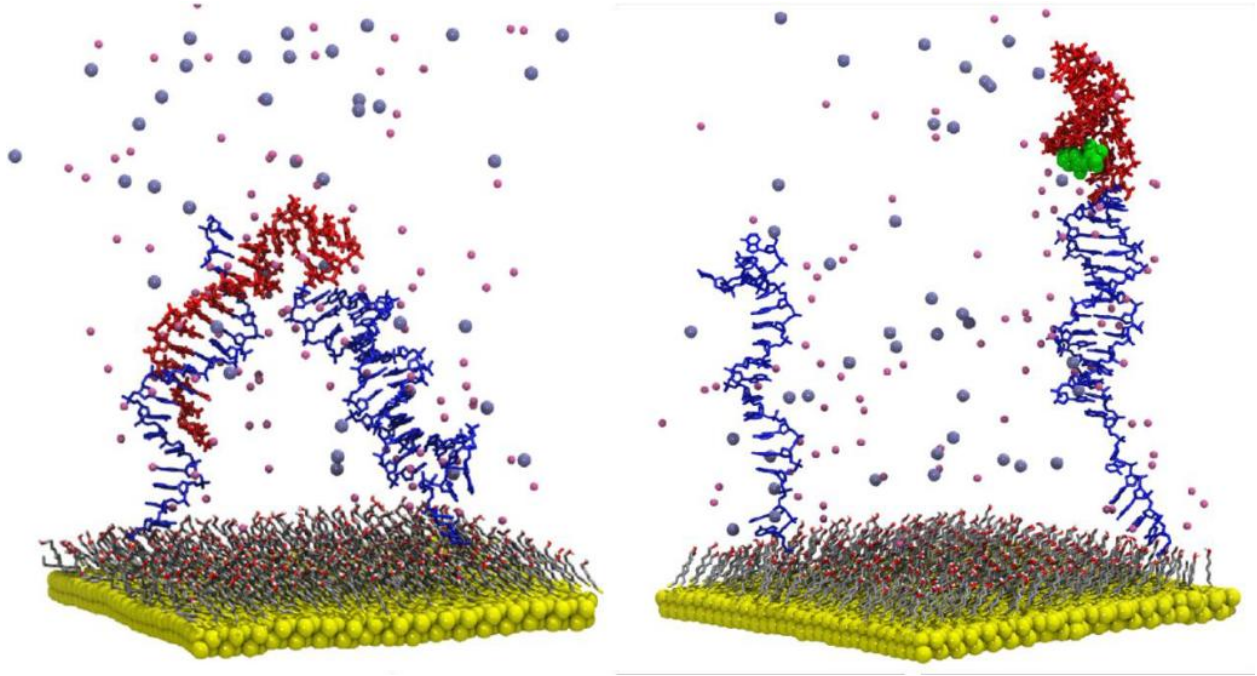


Figure 1: The complete off-state bridge system (Left) and open on-state system (Right) grafted to the Au/MCH surface, with Na^+ (pink) and Cl^- (purple) ions in solution and the AMP molecule in green. The aptamer sequence in both systems is marked in red.

We choose as a model system a three-strand DNA structure as a probe (cf. figure 1). The probe is composed of two DNA strands anchored on a gold surface and partially hybridized with an aptamer sequence sensitive to Ampicillin (AMP). The DNA structure has been designed to show strong structural change upon AMP binding to its aptamer.
Using computational techniques, mainly with molecular dynamics simulations, we firstly investigated the structure change upon the analyte binding (cf. figure 1). Original analyses of ions distributions along the trajectories unveil a distinct pattern between both states which can be related to changes in capacitance of the interface between these states. We have also tested the effects of surface charges on the DNA structure and in ions distributions, simulating a typical EGOFET usage. To our knowledge, this work demonstrates for the first time the ability of computational investigations to drive, in-silico, the design of aptasensors.

OC12
Revealing Short-Range Imbalances in the AMBER Lennard-Jones Potential for Sugar…Base Lone-pair…π Contacts in Nucleic Acids.

Pascal Auffinger[a], Klaudia Mrazikova[b,c], Jiri Sponer[b], Vojtech Mlynsky[b], Holger Kruse[b]

[a] Architecture et Réactivité de l'ARN, Université de Strasbourg, Institut de Biologie Moléculaire et Cellulaire du CNRS, Strasbourg, 67084, France
[b] Institute of Biophysics of the Czech Academy of Sciences, Královopolská 135, 612 65 Brno, Czech Republic
[c] National Centre for Biomolecular Research, Faculty of Science, Masaryk University, Kamenice 5, 625 00 Brno, Czech Republic

The lone-pair…π (lp…π) sugar…nucleobase stacking is a recurring structural motif in Z-DNA and RNAs.1 It is part of the structural signature of the CpG Z-steps in Z-DNA and r(UNCG) tetraloops, two nucleic acid systems that are notoriously poorly behaving in molecular dynamics (MD) simulations. Although the exact origin of the MD inaccuracies is unclear a significant part of the problem might be due to an imbalanced description of non-bonded interactions including the characteristic lp…π stacking. To provide insights into the links between lp…π stacking and MD issues, we present an in-depth comparison between accurate large-basis-set double-hybrid Kohn-Sham density functional theory calculations DSD-BLYP-D3/ma-def2-QZVPP (DHDF-D3) and data obtained with the non-bonded potential of the AMBER force field (AFF) for NpN Z-steps (N = G, A, C, U). We found that the AFF overestimates the DHDF-D3 lp…π distances by ~0.1-0.2 Å and that the deviation between the DHDF- D3 and AFF descriptions sharply increases in short-range region of the interaction. Based on the SAPT analysis and atom-in-molecule polarizabilities, we suggest that the DHDF-D3 vs. AFF differences originate in the inappropriate form of the repulsive part of the Lennard-Jones (LJ) potential as well as in the LJ radii of nucleobase carbon atoms which are assigned the same value in AFF while having different volumes due to the presence or absence of connected electron withdrawing groups. Finally, MD simulations of a Z-DNA helix and a r(UUCG) tetraloop revealed sampling of short lp…π distances in regions where the AFF overestimate short-range repulsion. It indicates that i) the restricted conformational space of the Z-steps supports compression of the lp…π contacts in MD simulations and ii) the over-repulsive AFF description of these contacts then further increases the strain in the Z-steps. We propose this as a major reason of the poor behavior of Z-step in nucleic acid simulations and suggest qualitative size-relations new LJ parameters for nucleobase carbons should follow.

(1) Kruse, H.; Mrazikova, K.; D'Ascenzo, L.; Sponer, J.; Auffinger, P. Short but Weak: The Z-DNA Lone-Pair · · · π Conundrum Challenges Standard Carbon Van Der Waals Radii. Angew. Chemie - Int. Ed. 2020, 59 (38), 16553–16560.

OC13
Computational Design of miniprotein binders

Younes Bouchiba[a], Manon Ruffini[a,b], Thomas Schiex[b], Sophie Barbe[a]

[a] Toulouse Biotechnology Institute, TBI, CNRS, INRAE, INSA, ANITI, Toulouse 31077 cedex 4, France
[b] Université de Toulouse, ANITI, INRAE, UR, MIAT, F-31320, Castanet-Tolosan, France

Miniprotein binders hold a great interest as a class of drugs that bridges the gap between monoclonal antibodies and small molecule drugs [1]. Like monoclonal antibodies, they can be designed to bind to therapeutic targets with high affinity, but they are more stable and easier to produce and to administer [2]. Thus, efficient and robust methods are expected to reduce the cost and time required for tailor-made inhibitor design. Here, we present a structure-based computational generic approach for de novo miniprotein inhibitor design. Specifically, we describe step-by-step the implementation of the approach for the design of miniprotein binders against the SARS-CoV-2 coronavirus, using available structural data on the SARS-CoV-2 spike receptor binding domain (RBD) in interaction with its native target, the human receptor ACE2 [3]. Structural data being increasingly accessible around many protein-protein interaction systems, this method might be applied to the design of miniprotein binders against numerous therapeutic targets. The computational pipeline exploits provable and deterministic artificial intelligence-based protein design methods [4,5], with some recent additions in terms of binding energy estimation, multistate design and diverse library generation [6-8].

Acknowledgements
The authors thank the financial support from French ANR through grant ANR-19-PI3A-0004.

References

[1] Vazquez-Lombardi R, Phan TG, Zimmermann C, et al (2015) Challenges and opportunities for non-antibody scaffold drugs. Drug Discov Today 20:1271–1283. https://doi.org/10.1016/j.drudis.2015.09.004
[2] Crook ZR, Nairn NW, Olson JM (2020) Miniproteins as a Powerful Modality in Drug Development. Trends Biochem Sci 45:332–346. https://doi.org/10.1016/j.tibs.2019.12.008
[3] Wang Q, Zhang Y, Wu L, et al (2020) Structural and Functional Basis of SARS-CoV-2 Entry by Using Human ACE2. Cell 181:894-904.e9. https://doi.org/10.1016/j.cell.2020.03.045
[4] Traoré S, Allouche D, André I, et al (2013) A new framework for computational protein design through cost function network optimization. Bioinformatics 29:2129–2136. https://doi.org/10.1093/bioinformatics/btt374
[5] Traoré S, Allouche D, André I, et al (2017) Deterministic Search Methods for Computational Protein Design. Methods Mol Biol Clifton NJ 1529:107–123. https://doi.org/10.1007/978-1-4939-6637-0_4
[6] Vucinic J, Simoncini D, Ruffini M, et al (2020) Positive multistate protein design. Bioinformatics, 36:122–130. https://doi.org/10.1093/bioinformatics/btz497
[7] Ruffini M, Vucinic J, Givry S de, et al (2021) Guaranteed Diversity Quality for the Weighted CSP. Network Based Computational Protein Design Methods. Algorithms 2021, 14, 168. https://doi.org/10.3390/a14060168
[8] Viricel C, de Givry S, Schiex T, Barbe S (2018) Cost function network-based design of protein-protein interactions: predicting changes in binding affinity. Bioinformatics 34:2581–2589. https://doi.org/10.1093/bioinformatics/bty092

OC14
Mapping amino acids at protein-membrane interfaces to update the current membrane binding model.

Thibault Tubiana[a,b], Nathalie Reuter[a,b]

[a]Department of Chemistry, University of Bergen, Norway
[b]Computational Biology Unit, University of Bergen, Norway

Peripheral Membrane Proteins (PMP) are essential components of multiple biological processes like lipid metabolism, membrane remodelling and other complex cellular pathways. They bind transiently to the surface of biological membranes through short helices or loops, and their Interfacial Binding Site (IBS) has been described as displaying basic and hydrophobic amino acids. However, recent studies indicate that this model is too simplistic to account for the fine-tuned lipid specificity achieved by many PMPs[1,2].

Terms like “hydrophobic spikes” or “protruding loops” have been used to describe PMP IBS. Recently we proposed a hydrophobic protrusion model for IBSs[3]. This model is based on the calculation of the protein convex hull and its vertices. When tested on a dataset of 300 protein families, it showed good performance in discriminating membrane-binding proteins from a reference dataset. This shows that it captures a structural pattern common to PMP families. Yet it does not have the resolution to capture fine differences between families, and these differences may be decisive for their membrane specificity. In order to characterize the IBS of the different PMP superfamilies at a finer resolution, we collected a curated dataset of structures from the CATH database[4]. We developed a statistical framework to map the frequency of charged amino acids and aromatic amino acids at protrusions and around protruding hydrophobes in our curated dataset. We will present and discuss the differences between the superfamilies and how they relate to their function, and the general trend we can observe among all the PMP we have in our dataset. Our results allow us to have a better understanding of the PMP membrane binding site.

1 C. Grauffel, B. Yang, T. He, M. F. Roberts, A. Gershenson, and N. Reuter, “Cation-m interactions as lipid-specific anchors for phosphatidylinositol-specific phospholipase C,” J. Am. Chem. Soc., vol. 135, no. 15, pp. 5740−5750, Apr. 2013.
2 N. L. Chon et al., “Membrane Docking of the Synaptotagmin 7 C2A Domain: Computation Reveals Interplay between Electrostatic and Hydrophobic Contributions,” Biochemistry, vol. 54, no. 37, pp. 5696−5711, Sep. 2015.
3 E. Fuglebakk and N. Reuter, “A model for hydrophobic protrusions on peripheral membrane proteins,” PLoS Comput. Biol., vol. 14, no. 7, p. e1006325, 2018.
4 I. Sillitoe et al., “CATH: Expanding the horizons of structure-based functional annotations for genome sequences,” Nucleic Acids Res., vol. 47, no. D1, pp. D280−D284, Jan. 2019.

OC15
Amphipathic helix folding in membranes: Markov State Models to decipher the mechanism

Paula Milán Rodríguez [a], Patrick Fuchs [a,b].

[a] Sorbonne Université, Ecole Normale Supérieure, PSL University, CNRS, Laboratoire des Biomolécules (LBM), 75005 Paris, France
[b] Université de Paris, UFR Sciences du Vivant, 75013 Paris, France

Amphipathic helices (AH) are short sequence peptides that interact with lipid membranes. They are unfolded in water and they fold into an α-helix while partitioning into the membrane. The helix has a polar side exposed to the solvent and a hydrophobic side inserted into the membrane. They have biological key roles such as targeting peripheral membrane proteins, translocating small molecules through the membrane or in signaling pathways. They also have interesting applications in pharmacology, notably drug delivery.

AH / membrane interactions can be studied by experimental methods such as NMR, Circular Dichroism (CD), or fluorescence, but these techniques generally lack atomistic details of the dynamics. Computational methods can help describe those details, but it remains challenging to characterize phenomenon on the millisecond time scale, such as helix folding within a membrane. To tackle this problem, one can use Markov State Models (MSM) which provides the formalism for assessing the kinetics and thermodynamics of the process from many short Molecular Dynamics (MD) trajectories started from different points of the energy landscape.

In this project, we use MSM modeling to study the partitioning and folding of an AH within a membrane. This AH is the well-known mastoparan from wasp venom (14 residues). All MD simulations are run with the GROMACS software with the CHARMM36m force field. MSM construction is performed using PYEMMA-Deeptime software.

So far, a first MSM of mastoparan in pure water has been constructed. The conformations observed in the simulations were divided in 3 sets of structures and we were able to characterize their probability and the rates of transition between them. The peptide remains essentially unfolded in water as assessed by CD. We are currently working on a second MSM characterizing the folding of mastoparan within the membrane. To generate various starting conformations, we used Replica-Exchange MD (REMD). Some structures were then selected along the folding pathway, which were used as input of regular MDs at room temperature. These MDs are then used to construct this second MSM.

Our strategy should prove useful to assess thoroughly AH / membrane interaction by computational means.

OC16
Integrative structural biology revealed how GAS M1 protein inhibits the IgG1 Fc-receptor binding site

Hamed Khakzad[1], Lotta Happonen[2], Yasaman Karami[3], Michael Nilges[3], Guy Tran Van Nhieu[1], Johan Malmström[2] and Lars Malmström[2]

[1] Équipe Signalisation Calcique et Infections Microbiennes, École Normale Supérieure Paris-Saclay, Gif-sur-Yvette, France
[2] Division of Infection Medicine, Department of Clinical Sciences, Lund University, Sweden
[3] Department of Structural Biology and Chemistry, C3BI, Institute Pasteur, Paris, France

Abstract

Group A Streptococci (GAS) is an important human pathogen capable of causing both mild infections in the upper respiratory tract as well as severe systemic diseases. To cause invasive infection, GAS produces a range of virulence factors to evade the human defenses. These virulence factors are predominantly secreted or surface- associated proteins that speci8cally target varied human proteins and protein complexes involved in the immune systems.

immunoglobulin G (IgG) antibodies are key elements of humoral immunity to detect bacteria. They bind to bacterial surface antigens via their Fab domains. The respective Fc domains bind to Fc-receptors on phagocytic cells, initializing phagocytic events. To prevent IgG-mediated phagocytosis, GAS employs two major virulence factors: M proteins and IgG degrading enzymes. The M proteins are the major surface-associated virulence factor of GAS. M proteins speci8cally bind to IgG-Fc-domains rendering it inaccessible for Fc-receptors. Here, we have characterized this sophisticated protein complex by combining Targeted cross-linking Mass Spectrometry (TX-MS[1]) and Molecular Dynamics (MD) simulations.

We determined the native interaction of M1 protein on the surface of bacteria with human IgGs by cross-linking the intact bacteria in human plasma. This revealed that M1-IgG interaction inhibits the Fc-receptor (FcγR) binding site on IgG. We further identi8ed that two domains of M1 have the highest a@nity to bind IgGs: hypervariable domain (A), and the C1-repeat. These are important domains of the M1 protein which have previously been shown to be highly conserved in almost all members of M protein family[2]. These two binding interfaces were further evaluated by MD simulations where the results indicated that the peptide from A domain has the highest a@nity to bind IgG. Interestingly, this peptide is a homologue of two important IgG FcγR binding proteins: staphylococcal protein A and streptococcal protein G.

In conclusion, using integrative structural biology, we identi8ed that M1 protein can capture human IgGs from their FcγR binding site. We characterized the molecular details of M1-IgGs interactions using TX-MS combined with MD simulations. Finally, these results indicate that the two speci8c domains of M1 play a crucial role in inhibiting IgG-Fc receptors that can be considered as potential vaccine candidates in future studies[3].

References

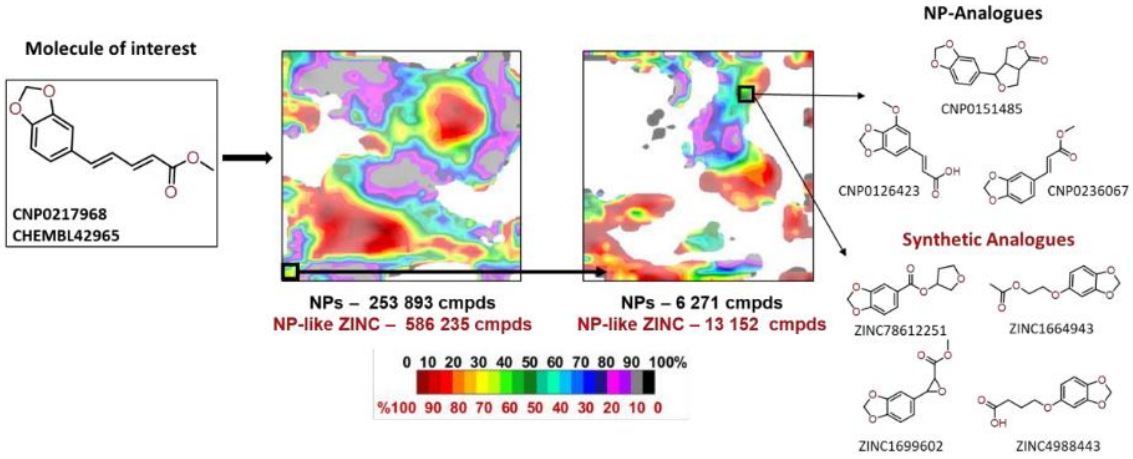
1 Hauri S, et. al., Rapid determination of quaternary protein structures in complex biological samples. Nature communications, 2019.
[2] Khakzad H, et. al., Greedy de novo motif discovery to construct motif repositories for bacterial proteomes. BMC Bioinformatics, 2019.
[3] Khakzad H, et. al., Structural determination of Streptococcus pyogenes M1 protein interactions with human immunoglobulin G using integrative structural biology. PLOS Computational Biology, 2021.

OC17
ChemSpace Atlas: empowering ultra-large library exploration

Y. Zabolotna[a], F. Bonachera[a], D. Horvath[a], O. Klimchuk[a], G. Marcou[a], A. Varnek[a]

[a] - Laboratoire de Chémoinformatique, UMR 7140, CNRS, Strasbourg, France

The Big Data era in medicinal chemistry is marked by a boom of novel chemical and biological information reported on a daily basis. New information is now produced at a higher speed than it can be analyzed and interpreted by human actors in the field. Therefore, there is an urgent need for efficient "big data"-compatible computational tools for the exploration of the chemical space of ultra-large libraries. This exploration should include interactive visualization, diversity, property and chemotype analysis, library comparison, in silico activity and ADMETox prediction, etc.
Here we present ChemSpace Atlas - an intuitive polyvalent tool for the efficient exploration of the ultra-large chemical space and its analysis with respect to medicinal chemistry problems. It is based on the hierarchical ensemble of tens of thousands of Generative Topographic Maps(GTM)[1], featuring biologically relevant chemical space. ChemSpace Atlas can be separated into several chapters depending on the chemical subspaces at focus: screening compounds[2] (fragment-like, lead-like, drug-like and PPI-like), natural products[3] (NPs) and soon DNA- encoded libraries and building blocks. The hierarchically organized GTMs allow a user to easily navigate through the hundreds of millions of compounds from a bird's eye view to structural pattern detection[4].
In order to facilitate navigation, a small set of "tracking" compounds can be provided by the user. These molecules will be projected onto the GTMs, appearing as dots on the selected landscapes. These dots will help to choose the zones of chemical space worth exploring in the context of user's needs. Apart from simple navigation, ChemSpace Atlas can be used for efficient analysis of underlying libraries - chemotype distribution, physicochemical properties, (reported and/or predicted) biological activity and commercial availability. Moreover, activity prediction is also available.



Search of the NPs and synthetic analogs of a compound of interest using NP Navigator chapter of ChemSpace Atlas
(https://infochm.chimie.unistra.fr/npnav/chematlas_userspace).

References

[1] C. M. Bishop, M. Svensén, C. K. I. Williams, Neural Comput. 1998, 10, 215-234.
[2] Y. Zabolotna, A. Lin, D. Horvath, G. Marcou, D. M. Volochnyuk, A. Varnek, J. Chem. Inf. Model. 2020.
[3] Y. Zabolotna, P. Ertl, D. Horvath, F. Bonachera, G. Marcou, A. Varnek, NP Navigator: a New Look at the Natural Product Chemical Space, 2021.
[4] A. Lin, B. Beck, D. Horvath, G. Marcou, A. Varnek, J. Comput. Aided Mol. Des. 2019.

OC18
Biasing RNA coarse-grained folding simulations with Small-Angle X-ray Scattering (SAXS) data

Lina ALFERKH[1], Liuba MAZZANTI[1], Elisa FREZZA[1] and Samuela PASQUALI[1]

[1] Cibles Thérapeutiques et Conception de Médicaments CiTCoM CNRS UMR 8038, 4 Avenue de l'Observatoire, 75006, Paris, France

RNA molecules can easily adopt alternative structures in response to different environmental conditions. As a result, a molecule's energy landscape is rough and can exhibits a multitude of deep basins. In the absence of a high-resolution structure, Small Angle X-ray Scattering data (SAXS) can narrow down the conformational space available to the molecule and be used in conjunction with physical modeling to obtain high-resolution putative structures to be further tested by experiments. Because of the low-resolution of this data, it is natural to implement the integration of SAXS data into simulations using a coarse-grained representation of the molecule, allowing for much wider searches and faster evaluation of SAXS theoretical intensity curves than with atomistic models. We present here the theoretical framework and the implementation of a simulation approach based on our coarse-grained model HiRE-RNA [1] combined with SAXS evaluations "on- the-fly" leading the simulation toward conformations agreeing with the scattering data, starting from partially folded structures as the ones that can easily be obtained from secondary structures predictions based tools. We show on three benchmark systems how our approach can successfully achieve high- resolution structures with remarkable similarity with the native structure recovering not only the overall shape, as imposed by SAXS data, but also the details of initially missing base pairs.

[1] T.Cragnoli, Y.Laurin, P.Derreumaux, S.Pasquali, J.Chem.Theory Comput., 11, 3510(2015).Barbara Gastel andRobert A Day. How to write and publish a scientific paper. ABC-CLIO, 2016.

OC19
Insights into the transport mechanism of BCRP through Molecular Dynamics with excited Normal Modes simulations

B. Dudas[a,b], X. Decleves[c], S. Cisternino[c], D. Perahia[b], M. A. Miteva[a]

[a]Inserm U1268 MCTR, CiTCoM UMR 8038 CNRS - University of Paris, France
[b]Laboratoire de Biologie et Pharmacologie Appliquée, Ecole Normale Supérieure Paris-Saclay, UMR 8113 CNRS, France
[c]Inserm UMRS-1144 Optimisation Thérapeutique en Neuropsychopharmacologie - University of Paris, France

Breast cancer resistance protein (BCRP/ABCG2) is member of the G subfamily of ATP-binding cassette (ABC) transporters and has a specific expressional distribution in the body. Indeed, BCRP has a pivotal role in detoxification by effluxing xenobiotics into the vascular compartment through the plasma membrane in tissues (e.g. blood-brain, blood-testis, blood-retinal, and maternal-fetal barriers), and also at the small intestine, the kidney, and the liver, among others to favor active body substrate excretion. BCRP transports a wide variety of substrates including cytotoxic anticancer agents and antibiotics and is identified to be a contributor to Multi Drug Resistance (MDR) phenotype in some tumor cells. In the transport process BCRP fluctuates between two conformational extrema, the inward-facing and the outward-facing states.

We performed Molecular Dynamics (MD) simulations starting from both the inward-facing and the outward-facing states, as well as employed the recently developed approach of MD with excited Normal Modes (MDeNM) to elucidate the path between the two states. MDeNM allowed to explore an extended conformational space of BCRP, which has not been achieved by using classical MD. Our results propose a pathway between the two states of BCRP. With the help of the identified intermediate structures, the identification of a wider range of substrates and inhibitors may become possible.

OC20
Investigating the interaction of enzymes with functionalized surfaces: Lessons from multiscale modeling approaches

Sophie Sacquin-Mora [1,2], Nicolas Bourassin[1,2], Marc Baaden[1,2]

1) CNRS, Université de Paris, UPR 9080, Laboratoire de Biochimie Théorique, 13 rue Pierre et Marie Curie, F-75005, Paris, France
2) Institut de Biologie Physico-Chimique-Fondation Edmond de Rothschild, PSL Research University, Paris, France

E-mail: sacquin@ibpc.fr

Redox enzymes represent promising tools for H_2-based technologies such as biofuel cells. However, many aspects of theses enzymes remain to be understood. In particular, designing efficient biofuel cells requires us to grasp details of the interaction between the enzymes and the electrode surfaces on the molecular level. Such information can be obtained using molecular modeling approaches on different scales, either with classical all-atom Molecular Dynamics simulations, or with coarse-grain calculations based on Elastic Network Models. Applications on [NiFe]-hydrogenases (which catalyze hydrogen oxidation), copper-billirubin oxidase (which catalyzes oxygen reduction) and Beta-Glucosidase show how simulations give us insight on factors determining enzymes orientation on the electrode surfaces, and how the adsorption on a solid surface can impact proteins structure, dynamics and mechanical properties, and therefore their catalytic activity.
Keywords: Redox proteins, biofuel cells, protein mechanics, protein dynamics, Molecular Dynamics, coarse-grain simulations

References:

[1] Oteri et al., PCCP 2014, 16, 11318-11322
[2] Oteri et al., J. Phys. Chem. B 2014, 118, 13800-13811
[3] Bourassin et al. J. Phys. Chem. B 2020, 124, 8516-8523

OC21
A MACHINE LEARNING MODEL OF CYP3A4 LIGAND SELECTIVITY BASED ON GEOMETRIC MODELING OF ACTIVE SITE ACCESS CHANNELS

Lydia Benkaidali[a,OC]‡, François André[b], Gautier Moroy[c], Anne-Claude Camproux[c], Michel Petitjean[c].

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

Cytochromes P450 (CYPs) constitute a large family of ubiquitous hemoproteins involved in drug metabolism in mammals and many biosynthesis pathways in all living organisms where monoxygenase activity is required. Remarkably, they recognize a broad spectrum of molecules sharing no common molecular motif. This multispecificity is not yet understood, although it seems clear that ligand recognition occurs not only in the active site, but also during its route through a particular ingress channel. A channel is defined as a ligand accessible pathway leading from the protein surface to the buried heminic cavity. Most cytochromes P450 exhibit several access channels with different structural features.

This study focuses on the 3A4 isoform, the main CYP involved in drug metabolism in human liver (about 35% of xenobiotics). We investigated the relationships between the nature of CYP3A4 ligands and the potential pathway they follow to reach the active site. This is done with the help of the CCCPP computational channels software [1].

By comprehensive scrutiny of several dozen crystal structures of CYP3A4 complexed with various CYP3A4 substrates and inhibitors, we found that the enzyme exhibits 3 main conformations 2 allowing the opening of 4 main channels 3. In our interpretation, channel plasticity allows geometric and electronic selection for different classes of drugs. A diagram predicting these access openings was proposed [2].

CCCPP has the originality to define the channels accurately by atoms (position and chemical properties) delimiting the voids in the protein by explicit walls allowing quantitative comparisons (volume and surface). It thus takes into account the structural features (shape and size) of the compound and the physico-chemical characteristics of its various interactions with the channel along its route to the active site, causing channels deformations.

In a subsequent study we used these data to perform a machine learning validation. The refined analysis of the rigid structures allowed us to study the flexibility of the isoenzyme. With a set of substrates, we perform a statistical model of the 3A4 channeling selectivity, predicting how the 3A4 opens specific channels to given ligands. The resulting 3A4 predictive conformation model may be enriched with in-vitro and in-vivo trials to get a relevant model of CYP3A4 metabolism. We further aim at extending this model to other major CYP isoforms involved in human liver metabolism.

[1] Benkaidali, L.; André, F.; Maouche, B.; Siregar, P.; Benyettou, M.; Maurel, F.; Petitjean, M. Computing cavities, channels, pores and pockets in proteins from non spherical ligands models. Bioinformatics 2014, 30 [6], 792-800. doi10.1093/bioinformatics/btt644
[2] Benkaidali L, André F, Moroy G, Tangour B, Maurel F, Petitjean M. The Cytochrome P450 3A4 has three Major Conformations: New Clues to Drug Recognition by this Promiscuous Enzyme. Mol Inform. 2017 Oct;36(10). doi: 10.1002/minf.201700044.
[3] Benkaidali L, André F, Moroy G, Tangour B, Maurel F, Petitjean M. Four Major Channels Detected in the Cytochrome P450 3A4: A Step toward Understanding Its Multispecificity. Int J Mol Sci. 2019;20(4):987. doi: 10.3390/ijms20040987.

OC22
Focused Library Design via Fragment-Bound Subpocket Alignment and Deep Generative Linking: A Proof- of-Concept for CDK8 Inhibitors

Merveille Eguida[a], Didier Rognan[a]

[a] Laboratoire d'Innovation Thérapeutique, UMR 7200 CNRS-Université de Strasbourg, 67400 Illkirch, France.

Virtual screening is widely used in drug discovery to predict the affinity of molecules prior to biological testing in a time and cost effective manner. However, it is a challenge to find new hits, which not only depends on the performances of the methods used, but even more depends on an efficient exploration of the largest possible chemical space. Although recent studies reported that the chances to discover new hits is higher when exploring very large on-demand libraries of billions of compounds, such an ultra-large scale screening is only possible with huge computing resources[1]. On the contrary, drug discovery would benefit from screening focused libraries for a specific target of protein family[2], with the idea that such libraries are customized to be enriched in hits for that protein target.

Herein, we introduce a novel approach for designing a focused compound library for virtual screening on a specific target exploiting available pocket information. The protein pocket is described as a cloud of points with key shape and pharmacophoric properties. In a first step, 3D fragmented ligands derived from other protein structures are aligned with ProCare (protein cavity alignment method)[3] at specific positions in the target binding site on the basis of the fragments subpockets/target pocket similarity. In a second step, pairs of fragment hits are linked by a deep generative method (DeLinker)[4] to yield fully connected molecules.

The above-described approach was applied to design a focused library for CDK8 inhibition and assessed in two scenarios. First, we evaluated the ability of the approach to design known CDK8 inhibitors. We were indeed able to reconstruct a few molecules similar to CDK8 inhibitors from fragments originally bound to both kinase and non-kinase protein structures. Second, we generated diverse and synthetically accessible molecules, chemically dissimilar to known CDK8 inhibitors, that we prioritize for in vitro CDK8 inhibition assays.

In a nutshell, the herein described approach combines a subpocket knowledge-driven positioning of fragments into a protein structure to fragments 3D linking, to design a focused library of molecules with a particular advantage that it is applicable to any druggable protein target, irrespective of prior ligand information.

1 Lyu, J.; Wang, S.; Balius, T. E.; Singh, I.; Levit, A.; Moroz, Y. S.; O'Meara, M. J.; Che, T.; Algaa, E.; Tolmachova, K.; Tolmachev, A. A.; Shoichet, B. K.; Roth, B. L.; Irwin, J. J. Ultra-Large Library Docking for Discovering New Chemotypes. Nature 2019, 566, 224−229.
2 Sydow, D.; Schmiel, P.; Mortier, J.; Volkamer, A. KinFragLib: Exploring the Kinase Inhibitor Space Using Subpocket-Focused Fragmentation and Recombination. J. Chem. Inf. Model. 2020, 60, 6081−6094.
3 Eguida, M.; Rognan, D. A Computer Vision Approach to Align and Compare Protein Cavities: Application to Fragment-Based Drug Design. J. Med. Chem. 2020, 63, 7127−7142.
4 Imrie, F.; Bradley, A. R.; van der Schaar, M.; Deane, C. M. Deep Generative Models for 3D Linker Design. J. Chem. Inf. Model. 2020, 60, 1983− 1995.

OC23
Investigating conformational changes between perturbed molecular dynamics : a network approach

Gheeraert Aria[a,b,c], Ivan Rivalta[b,d], Laurent Vuillona,c, Claire Lesieur[c,e], Bernard Maigret[f]

[a] LAMA, Université Savoie Mont-Blanc, CNRS, Bourget-du-Lac, France
[b] Dipartimento di Chimica Industriale "Toso Montanari", Università di Bologna, Viale Risorgimento 4, I-40136 Bologna, Italy
[c] Institut Rhônalpin des systèmes complexes, IXXI-ENS-Lyon, 69007 Lyon, France
[d] Univ Lyon, Ens de Lyon, CNRS UMR 5182, Université Claude Bernard Lyon 1, Laboratoire de Chimie, F69342 Lyon, France
[e] AMPERE, CNRS, Univ. Lyon, 69622 Lyon, France
[f] LORIA Institute, Capsid team, CNRS - INRIA – University of Lorraine, France

The use of atomistic Molecular Dynamics (MD) simulations to elucidate changes in conformation of a proteic system within different environments is standard. The most challenging part consists in analyzing these MD simulations. Previous works used amino acid contact networks (AAN) to analyze the differences between the crystal structure of mutants[1,2] and between the molecular dynamics of mutants[3]. Mutations can be viewed as a type of perturbation that changes both atomic numbers and positions. Hence we can propose a generalization of the AAN approach beyond the mutation context to study differences between a set of initial MD simulations and a set of perturbed MD simulations by building their AAN and then subtracting them to highlight dynamical changes resulting from these perturbations. The application of this technique was successfully used to explore allosteric pathways of an enzyme[4] and showed good agreement with previous theoretical and experimental results, even detecting new allosteric spots proposed for mutagenesis experiments. Overall this tool demonstrated to be an elegant and fast way to analyze perturbed sets of MD simulations and is largely scalable with the number of atoms and length of simulation.

(1) Vuillon, L.; Lesieur, C. From Local to Global Changes in Proteins: A Network View. Current Opinion in Structural Biology 2015, 31, 1–8. https://doi.org/10.1016/j.sbi.2015.02.015.
(2) Dorantes-Gilardi, R.; Bourgeat, L.; Pacini, L.; Vuillon, L.; Lesieur, C. In Proteins, the Structural Responses of a Position to Mutation Rely on the Goldilocks Principle: Not Too Many Links, Not Too Few. Physical Chemistry Chemical Physics 2018, 20 (39), 25399–25410. https://doi.org/10.1039/c8cp04530e.
(3) Doshi, U.; Holliday, M. J.; Eisenmesser, E. Z.; Hamelberg, D. Dynamical Network of Residue–Residue Contacts Reveals Coupled Allosteric Effects in Recognition, Catalysis, and Mutation. Proceedings of the National Academy of Sciences 2016, 113 (17), 4735–4740. https://doi.org/10.1073/pnas.1523573113.
(4) Gheeraert, A.; Pacini, L.; Batista, V. S.; Vuillon, L.; Lesieur, C.; Rivalta, I. Exploring Allosteric Pathways of a V-Type Enzyme with Dynamical Perturbation Networks. The Journal of Physical Chemistry B 2019, 123 (16), 3452–3461. https://doi.org/10.1021/acs.jpcb.9b01294.

OC24
Structure guided de novo drug design using deep generative modeling, a case study

Maud Jusot[1], Nicolas Devaux[1], Anna Kriukova[1], Christopher Housseman[1], Yann Lamotte[2], Brice Hoffmann[1]

1 Iktos, 65 rue de Prony 75017, Paris, France
2 Oncodesign, 25-27 Avenue du Quebec, 91140 Villebon-sur-Yvette, France

In early stage drug discovery projects, Structure-based virtual screening has demonstrated in numerous cases its ability to identify new hits[1]. Although the datasets of virtual molecules continue to grow to reach billions[2], everyone uses the same datasets and are therefore likely to find the same hits on a given target. Generative AI brings new solutions for hit discovery, having the possibility to explore the chemical space beyond the molecules described in existing databases[3].

In this study, we have developed a pipeline of generative AI guided by a structure-based assessment of molecules. We applied this protocol to search for new and original inhibitors of PIM-1, a kinase protein involved in multiple human cancers, including prostate cancer[4].

We trained a recurrent neural network (RNN) with long short-term memory architecture on the Chembl dataset. We evaluated the generated compounds using docking, interaction analysis and physical properties. The weights of the model are adjusted to maximize the probability of generating molecules having better scores using a policy gradient algorithm. Molecules with the same scaffold as one of the known PIM inhibitors got a score of zero in order to force the generator to find new solutions. The most promising molecules have been synthesized and evaluated in biochemical and early ADME assays.

Twelve molecules have been synthesized from five different scaffolds. These molecules are either the exact molecules proposed by the generator or close analogs from common intermediates. Two molecules were measured with activity $\leq 1\mu M$ from two different scaffolds and good preliminary ADME properties (logD, solubility, stability).

We demonstrated the capacity of generative algorithms to find molecules that are both innovative and active towards a protein of therapeutic interest, using structural assessment as a guide for the improvement of the model. This shows a new way to navigate the chemical space by overcoming the limitations of existing databases.

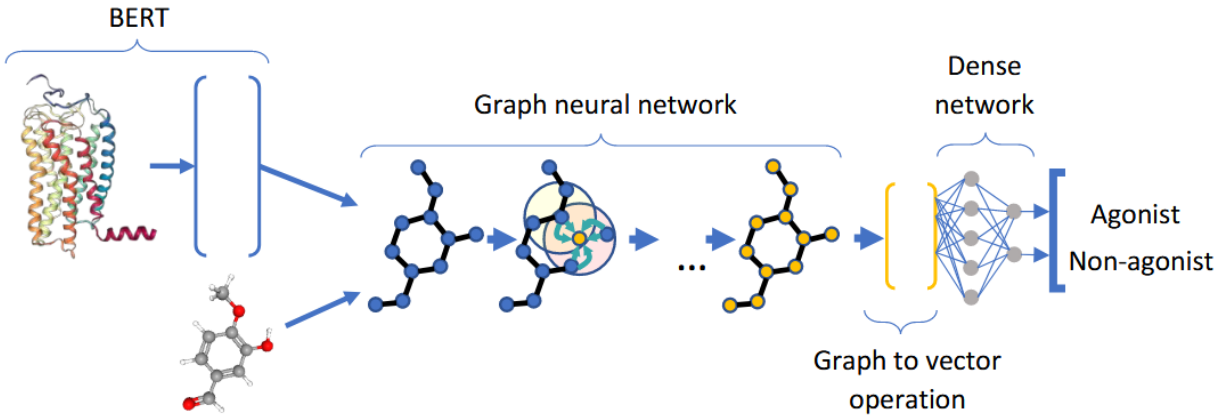
1 Brown, D. G., & Boström, J. (2018). Where do recent small molecule clinical development candidates come from?. Journal of medicinal chemistry Journal of medicinal chemistry, 61(21), 9442-9468.
2 Irwin, J. J. et al. (2020). ZINC20—A Free Ultralarge-Scale Chemical Database for Ligand Discovery. Journal of Chemical Information and Modeling.
3 Gómez-Bombarelli, R., Wei, J. N., Duvenaud, D., Hernández-Lobato, J. M., Sánchez-Lengeling, B., Sheberla, D., ... & Aspuru-Guzik, A. (2018). Automatic chemical design using a data-driven continuous representation of molecules. ACS central science, 4(2), 268-276.
4 Mumenthaler, S. M. et al. (2009). Pharmacologic inhibition of Pim kinases alters prostate cancer cell growth and resensitizes chemoresistant cells to taxanes. Molecular cancer therapeutics, 8(10), 2882-2893.

OC25
Representation learning to overcome scarce data in machine learning. Application to chemosensory receptors

Matej Hladiš[a], Sébastien Fiorucci[a], Jérémie Topin[a], Jérôme Golebiowski[a]

[a] Institute of Chemistry - Nice, UMR 7272 CNRS – Université Côte d'Azur, 06108 Nice cedex, France

Our sense of smell relies on the use of around 400 genes expressing functional odorant receptors (OR), endowing us with the power to discriminate more than 1 trillion stimuli. ORs are transmembrane proteins which belong to the family of class A G protein-coupled receptors (GPCR). Establishing a relationship between the structure of a molecule and the smell is a long-standing difficulty. The first step to break the combinatorial code of olfaction relies on identification of OR-ligand pairs.
Nowadays, the data linking a molecule to a set of ORs remain scarce and only 131 ORs have an identified ligand. Thus, building a machine learning protocol taking ORs' sequence explicitly remains challenging. To tackle this issue, we leverage recent advances in representation learning and combine them with graph neural network (GNN) to build a receptor-ligand interaction prediction model.
Several methods inspired by success of representation learning in the natural language processing (NLP) have been proposed to represent protein sequences. These methods aim at finding an informative vector representation of amino acids in the protein sequence and are trained in an unsupervised manner enabling them to use large datasets with millions of protein sequences. Here we use architecture based on BERT model[1] to represent ORs which was previously trained on more than 200M protein sequences. We then use this representation as a starting point for receptor-ligand prediction. We treat ligands as graphs and process ORs and ligands simultaneously using GNN.
We evaluate our model on a set of more than 7500 OR-ligand pairs. Our model is achieving a Matthews correlation coefficient of 0.40 in the case that all receptors are included in the training set (i.e. random split). The performance on a much more difficult deorphanization task (i.e. discarding all pairs of a given receptor) remains acceptable with a value of 0.27. As a comparison, an exhaustive in vitro search would lead to a success rate of ~3% and MCC equal to 0. We also apply the same protocol on other subfamilies of class A GPCR.



1 Elnaggar A, Heinzinger M, Dallago C, Rehawi G, Wang Y, Jones L, Gibbs T, Feher T, Angerer C, Steinegger M, Bhowmik D, Rost B. Prottrans: Towards cracking the language of life's code through self-supervised learning. bioRxiv, 2021

OC26
Application of Frags2Drugs for the fragment-based drug design of macrocyclic kinase inhibitors

Gautier PEYRAT[a], Pascal KREZEL[a], Stéphane BOURG[a], Samia ACI-SECHE[a], Pascal BONNET[a]

[a] Institut de Chimie Organique et Analytique (ICOA UMR 7311), Université d'Orléans - Pôle de chimie, rue de Chartres - BP 6759, 45067 Orléans Cedex 2, France

Macrocycles are molecules composed of at least 12 atoms in a ring architecture. Macrocycles are able to inhibit challenging targets and they sometimes present a better affinity and selectivity than linear kinase inhibitors[1]. The dysfunction of protein Anaplastic Lymphoma Kinase (ALK) can lead to several cancers. In 2018, the FDA approved the macrocyclic lorlatinib to treat patients with anaplastic lymphoma kinase (ALK)-positive metastatic non-small cell lung cancer. This drug is also a potent brain-penetrant, third-generation ALK/ROS1 inhibitor. We focused the study on the discovery of novel ALK macrocyclic inhibitors to validate Frags2Drugs (F2D) as a tool for the design of novel macrocycles.

F2D is an in silico Fragment Based Drug Design program building protein kinase inhibitors in the ATP binding sites. F2D uses a fragment library stored in a graph-oriented database. It relies on data obtained by the fragmentation of X-ray crystal structures of protein kinase-ligand complexes. In a minute scale, F2D builds every possible molecule fitting in the target cavity. Molecules presenting the best predicted affinity are then filtered and kept. Among several specific molecular filters, we apply an in-house Protein Kinase Inhibitors-like filter obtained from the Protein Kinase Inhibitor Database (PKIDB)[2,3].

Starting from the 2-aminopyridine seed from co-crystallized lorlatinib, we obtained 592 macrocycles. After applying several filters based on PAINS (pan assay interference compounds substructure removal)[4], synthetic accessibility estimation5 and molecular docking for binding mode confirmation, we selected 153 ALK macrocycles. None of these inhibitors are similar (Tc > 0.7 using FCFP4 fingerprints) to compounds present in the ChEMBL, ZINC, PKIDB or Ambinter databases. Among the 153 new macrocyclic inhibitors, 9 have a quantitative estimation of drug-likeness score[6] > 0.7 suggesting potential lead compounds after few optimization steps. The suggested macrocycles will be synthesized and evaluated on ALK and on a panel of protein kinases for selectivity.

This work demonstrates the ability of F2D to discover ALK protein kinase macrocyclic inhibitors. This approach shows that F2D could be used in future to design macrocyclic inhibitors for novel protein kinases. The filtering steps greatly help to reduce the number of macrocycles, but it still lacks today a bioactivity estimation, which is currently under development.

(1) Driggers, E. M.; Hale, S. P.; Lee, J.; Terrett, N. K. The Exploration of Macrocycles for Drug Discovery — an Underexploited Structural Class. Nat. Rev. Drug Discov. 2008, 7 (7), 608–624. https://doi.org/10.1038/nrd2590.
(2) Carles, F.; Bourg, S.; Meyer, C.; Bonnet, P. PKIDB: A Curated, Annotated and Updated Database of Protein Kinase Inhibitors in Clinical Trials. Molecules 2018, 23 (4), 908. https://doi.org/10.3390/molecules23040908.
(3) Bournez, C.; Carles, F.; Peyrat, G.; Aci-Sèche, S.; Bourg, S.; Meyer, C.; Bonnet, P. Comparative Assessment of Protein Kinase Inhibitors in Public Databases and in PKIDB. Molecules 2020, 25 (14), 3226. https://doi.org/10.3390/molecules25143226.
(4) Baell, J. B.; Holloway, G. A. New Substructure Filters for Removal of Pan Assay Interference Compounds (PAINS) from Screening Libraries and for Their Exclusion in Bioassays. J. Med. Chem. 2010, 53 (7), 2719– 2740. https://doi.org/10.1021/jm901137j.
(5) Ertl, P.; Schuffenhauer, A. Estimation of Synthetic Accessibility Score of Drug-like Molecules Based on Molecular Complexity and Fragment Contributions. J. Cheminformatics 2009, 1 (1), 8. https://doi.org/10.1186/1758-2946-1-8.
(6) Bickerton, G. R.; Paolini, G. V.; Besnard, J.; Muresan, S.; Hopkins, A. L. Quantifying the Chemical Beauty of Drugs. Nat. Chem. 2012, 4 (2), 90–98. https://doi.org/10.1038/nchem.1243.

OC27
Improving SAR analysis via pharmacophoric feature selection and feature transformation

Hajar Rehioui[a], Abdelkader Ouali[a], Christophe Couronne[a], Jean-Luc Lamotte[b], Alban Lepailleur[b], Jean-Luc Manguin[a], Ronan Bureau[b], Albrecht Zimmermann[a], Bertrand Cuissart[a]

[a] GREYC, Normandie Univ., UNICAEN, CNRS – UMR 6072, 14000 Caen, France
[b] Centre d'Etudes et de Recherche sur le Médicament de Normandie, Normandie Univ, UNICAEN, CERMN, 14000 Caen, France

Recently, the analysis of Structure-Activity Relationships has been confronted with the high dimensionality of chemical representations of molecular datasets, making analysis for drug discovery more complicated. To address this, different machine learning (ML) approaches have been exploited and proved their effectiveness by extracting relevant information. Generally, before applying ML methods, the raw data must be preprocessed to obtain better results.
In our work, we start with a dataset described by pharmacophores obtained by the Norns tool1. Norns considers a dataset of molecules for which both structure and activity are given, and extracts a set of pharmacophores whose occurrences in the data set fulfill specified properties. For example, Norns allows the automatic extraction of 112047 pharmacophores from ligands tested on BCR-ABL, without any prior supervised selection. However, the size of this set is too large to perform efficient analysis and offer explicable results based on ML algorithms. This is why, in a first step, we select a subset of pharmacophores via grouping them into equivalence classes: pharmacophores occurring in the same set of molecules. This first unsupervised selection step allows us to retain 22127 pharmacophores.

As the latter step removes redundant pharmacophores without losing statistical information, it allows to cope with the high dimensionality exhibited by chemogenomics datasets. We can then perform more sophisticated operations using the new representation: we do so by passing the data through two neural network-based transformations. The first neural network (NN) performs an unsupervised transformation2 by reducing a loss function based on similarity computations. Its goal is to reduce the distance between similar data and increase it between dissimilar ones. The second NN exploits (a part of) the activity information with the aim of obtaining a more structured data space. While we use labels contained in the data in our work, the necessary information could also be the result of an expert interacting with the process, a first step towards interactive mining of pharmacophores. We obtain good clustering performances which lead to an easier and more efficient analysis. In addition, feature weights derived from the NN-based transformations could help to explain the results of the clustering step.

The oral communication will lay out the methodological details of our proposal, which will be practically illustrated and assessed thanks to an experiment conducted on BCR-ABL ligands. Then, we will provide our research perspectives in the framework of two projects: the FEDER-funded project SCHISM* and the ANR project InvolvD**.

*Our work is part of the European project "SCHISM", funded by the European Union within the framework of the Operational Programme ERDF/ESF 2014-2020.
**ANR-20-CE23-0023.

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 Yan D., Zhou X., Wang X., & Wang R., An off-center technique: Learning a feature transformation to improve the performance of clustering and classification. Information Sciences, 2019,503, 635-651

OC28
Predicting potential endocrine disrupting chemicals binding to Estrogen receptor a using a combination of structure based and ligand based in silico methods

Asma Sellami [a], Matthieu Montes [a] and Nathalie Lagarde [a]

[a] Laboratoire GBCM, EA 7528, Conservatoire National des Arts et Métiers, HESAM Université, F-75003 Paris, France

Background. Estrogen receptors α (ERα) are transcription factors belonging to the nuclear receptors (NRs) protein family involved in several key physiological processes to which several xenobiotics can bind. Indeed, endocrine disrupting chemicals (EDCs) can bind to several NRs, including ERα, triggering toxicological pathways. Experimental campaigns are conducted to identify potential EDCs. An intermediate step of prioritization can reduce the number of compounds to be tested and thus facilitate the process. In the present work, we designed a pipeline for the prediction of compounds binding to ERα. These flagged compounds can be further explored using experimental techniques to assess their potential to be EDCs. The proposed pipeline combines structure-based (docking and pharmacophore models) and ligand based (pharmacophore models) methods.
Methods. The prediction models have been constructed using the EPA data [1] encompassing many structurally diverse compounds experimentally tested in ERα binding assays. Two external databases were also used for the validation step: the NR-DBIND [2] and the EADB [3]. For the docking approach, single and ensemble structure docking protocols were performed using three different softwares with free academic license i.e smina [4], Surflex-dock [5] and PLANTS [6]. For the pharmacophore modeling approach, both structure-based (SB) and ligand-based (LB) pharmacophores were generated using the LigandScout software. Two different combinations of docking and pharmacophore model approaches were investigated, i.e. consensus and hierarchical protocols.
Results. For the docking approach, the best results (AUC higher than 0.7) were obtained with the single structure protocol and smina-dkoes. Furthermore, an original strategy using the predictiveness curve
[7] was implemented to define a scoring threshold able to efficiently discriminate ERα binding from non-binding compounds. For the pharmacophore approach, 15 SB and 11 LB optimized pharmacophores were retained and combined into a unique SBLB collection. Although, each method yielded good performances at distinguishing between ERα binding and non-binding compounds, the combination of single structure docking and SBLB pharmacophores approaches were associated with better results. The consensus protocol was the best suited for toxicological studies with sensitivity and specificity values reaching 0.81 and 0.54, respectively. The hierarchical protocol, associated with higher specificities values, could be used for potential drug discovery projects.

References:
[1] Williams AJ, Grulke CM, Edwards J, McEachran AD, Mansouri K, Baker NC, et al. The CompTox Chemistry Dashboard: a community data resource for environmental chemistry. Journal of Cheminformatics 2017;9:61.
[2] Réau M, Lagarde N, Zagury J-F, Montes M. Nuclear Receptors Database Including Negative Data (NR-DBIND): A Database Dedicated to Nuclear Receptors Binding Data Including Negative Data and Pharmacological Profile. J Med Chem 2019;62:2894−904.
[3] Shen J, Xu L, Fang H, Richard AM, Bray JD, Judson RS, et al. EADB: an estrogenic activity database for assessing potential endocrine activity. Toxicol Sci 2013;135:277−91.
[4] Koes DR, Baumgartner MP, Camacho CJ. Lessons learned in empirical scoring with smina from the CSAR 2011 benchmarking exercise. J Chem Inf Model 2013;53:1893−904. h
[5] Jain AN. Surflex: fully automatic flexible molecular docking using a molecular similarity-based search engine. J Med Chem 2003;46:499−511.
[6] Korb O, Stützle T, Exner TE. PLANTS: Application of Ant Colony Optimization to Structure-Based Drug Design. In: Dorigo M, Gambardella LM, Birattari M, Martinoli A, Poli R, Stützle T, editors. Ant Colony Optimization and Swarm Intelligence, vol. 4150, Berlin, Heidelberg: Springer Berlin Heidelberg; 2006, p. 247−58.
[7] Empereur-mot C, Guillemain H, Latouche A, Zagury J-F, Viallon V, Montes M. Predictiveness curves in virtual screening. Journal of Cheminformatics 2015;7:52.

OC29
Drug Design with Martini 3 Coarse-Grained Model

Paulo Cesar Telles de Souza[a]

[a]Molecular Microbiology and Structural Biochemistry, UMR 5086 CNRS and University of Lyon, Lyon, France

The Martini model[1,2] is among the most popular coarse-grained (CG) models in the field of biomolecular simulation, due to its easy to use building block principle. Martini relies on a mapping of two to four non- hydrogen atoms and has been parameterized using a top-down approach with thermodynamic data as the main target. With its growing use, however, a number of shortcomings have recently become apparent[3]. These limitations pushed us to fully reparametrize the CG beads - the fundamental building blocks of Martini - until now largely untouched since 2007[1]. In this talk, I will present the refined model, with an improved interaction balance, new bead types and expanded ability to include specific interactions based hydrogen bonding and electron polarizability[2]. The new model, coined Martini 3, allows more accurate predictions of molecular packing and interactions in general, allowing new applications for coarse-grained models involving protein-small-molecule interactions. Remarkably, we achieve high accuracy without the need of any a priori knowledge of binding pockets or pathways[4]. Given its excellent computational efficiency, this opens the way to high-throughput drug screening based on dynamic docking pipelines[5].

1. Marrink, S. J., Risselada, H. J., Yefimov, S., Tieleman, D. P. & de Vries, A. H. The MARTINI force field: coarse grained model for biomolecular simulations. J. Phys. Chem. B 111, 7812–7824 (2007).
2. Souza, P. C. T. et al. Martini 3: a general purpose force field for coarse-grained molecular dynamics. Nat. Methods 18, 382–388 (2021).
3. Alessandri, R. et al. Pitfalls of the Martini Model. J. Chem. Theory Comput. 15, 5448–5460 (2019).
4. Souza, P. C. T. et al. Protein–ligand binding with the coarse-grained Martini model. Nat. Commun. 11, 1–11 (2020).
5. Souza, P. C. T., Limongelli, V., Wu, S., Marrink, S. J. & Monticelli, L. Perspectives on High-Throughput Ligand/Protein Docking With Martini MD Simulations. Front. Mol. Biosci. 8, 199 (2021).

OC30
Understanding Structure-Activity Relationship With Interpretability Methods for Molecular Activity and Chemical Property Prediction

Doron Israel[1], Joseph-André Turk*[1], Christopher Housseman[1], Stéphanie Labouille[1], Vinicius Barros[1], Nicolas Drizard[1], Quentin Perron[1]

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

* E-mail: joseph-andre.turk@iktos.com

While the predictive performance of supervised learning models has improved over the years, the interpretability of modern machine learning algorithms remains poorly understood, especially in computational chemistry[1]. In this work, we obtain explanations of models by primarily using a local interpretation method called SHAP[2] (for SHapley Additive exPlanations). This method can be applied to any type of machine learning model and is based on solid mathematical foundations. The explanation of a prediction is given by a set of values that represent the fair contribution of each feature to the prediction.

We first show how this method can help the chemist to study the structure–activity relationships (SARs). Next, we introduce an interactive web application, providing an iterative workflow between the model explanations and the chemist to improve the robustness of the model. Indeed, the user chooses a set of interpretable features (such as Circular Fingerprints[3] or molecular descriptors) and then trains a model to predict a target from which expert knowledge is available. SHAP algorithm is then used to obtain the contributions of each feature to the prediction of each molecule. The expert user then analyzes those explanations and is able to detect spurious correlations if some features appear to be important for the model but do not fit his expertise. Those features are not causally linked to the target and are likely correlated with activity because of a selection bias in the dataset leading to overfitting, which prevents generalization to new chemical spaces. The model is then re-trained without these problematic features and the whole process is repeated until model explanations better fit the expert's knowledge, and that at the same time the predictive performance in cross-validation has not been too negatively impacted. This workflow enables the user to "fix" the model until it focuses on the true causal relationships between the features and the target, in order to improve its generalisability and robustness.

Bibliography:

[1] Jiménez-Luna, José, Francesca Grisoni, and Gisbert Schneider. "Drug discovery with explainable artificial intelligence." Nature Machine Intelligence 2.10 (2020): 573-584.
[2] Lundberg, Scott M., et al. "From local explanations to global understanding with explainable AI for trees." Nature machine intelligence 2.1 (2020): 56-67.
[3] Rogers, David, and Mathew Hahn. "Extended-connectivity fingerprints." Journal of chemical information and modeling 50.5 (2010): 742-754.

POSTERS

PP1
Leveraging Zernike polynomials for peptide conformer identification

Amaya-Ramirez Diego [a], Maigret Bernard[a], Corman Etienne[a], Taupin Jean-Luc[b] and Devignes Marie-Dominique [a]

a Université de Lorraine, CNRS, Inria, LORIA, Nancy, F-54000, France
b INSERM, Hôpital Saint Louis, F-75010 Paris

Background : Peptides play a key role in multiple biological processes, e.g. peptide hormones or peptide-mediated immune response in signaling pathways1,2 . For this reason, peptides have gained relevance in recent years as pharmacological agents3. However, peptides are highly flexible leading to diverse three-dimensional (3D) conformers that can modify their specificity and efficacy in binding to their targets. Identifying and understanding such 3D-conformers is therefore fundamental. This can be achieved by analysis of molecular dynamics (MD) runs. However, classical methods such as RMSD maps are difficult to process automatically and manual interpretation can be subjective.
Objectives : To implement an automatic strategy for identifying 3D conformers of the [Pyr1]apelin-13 peptide in a MD run.
Methods : The peptide [Pyr1]apelin-134 (with sequence pGLU-ARG-PRO-ARG-LEU-SER-HIS-LYS-GLY-PRO-MET-PRO-PHE) is selected as a case study. A 100ns MD simulation is performed using NAMD and CHARMM force-field, from which a 1000-frame trajectory is extracted to represent the conformational space. For each frame, a 3D Zernike descriptor5,6 is used to represent the 3D shape of the peptide as a vector of invariants representing a series of orthogonal polynomials. The 3D shape is the solvent-accessible surface, constructed from the Van der Waals radius sphere associated with each atom type and obtained with Chimera7 in VTK format. The MindBoggle8 python library is then adapted to calculate the 3D Zernike descriptor relying on 121 invariants..
Conformer identification is performed by clustering the previously calculated 3D Zernike descriptors, using the K-means and silhouette analysis algorithms in the scikit-learn9 implementation to identify the optimal number of clusters. For each cluster, the conformation closest to its centroid is chosen as a representative.
Results : A 3D Zernike descriptor vector is calculated for each frame in the trajectory. According to silhouette analysis, two main clusters are discernable. The representative conformers correspond to frames 748 and 840.
Conclusions : The 3D Zernike descriptors have already been used for protein tertiary structure comparison but to our knowledge, not to identify 3D-conformers from MD runs. The Zernike vector format allows using a variety of clustering methods to process MD frames and to automatically determine the optimal number of clusters/conformers. This strategy could be promising for other biomolecules such as proteins or nucleic acids.

(1) Wang, H.; Feng, Z.; Xu, B. Assemblies of Peptides in a Complex Environment and Their Applications. Angew. Chem. Int. Ed. 2019, 58 (31), 10423–10432. https://doi.org/10.1002/anie.201814552.
(2) Kálmán, A.; Thunecke, F.; Schmidt, R.; Schiller, P. W.; Horváth, C. Isolation and Identification of Peptide Conformers by Reversed-Phase High-Performance Liquid Chromatography and NMR at Low Temperature. J. Chromatogr. A 1996, 729 (1–2), 155–171. https://doi.org/10.1016/0021-9673(95)01059-9.
(3) Salmaso, V.; Sturlese, M.; Cuzzolin, A.; Moro, S. Exploring Protein-Peptide Recognition Pathways Using a Supervised Molecular Dynamics Approach. Structure 2017, 25 (4), 655-662.e2. https://doi.org/10.1016/j.str.2017.02.009.
(4) Yang, P.; Kuc, R. E.; Brame, A. L.; Dyson, A.; Singer, M.; Glen, R. C.; Cheriyan, J.; Wilkinson, I. B.; Davenport, A. P.; Maguire, J. J. [Pyr1]Apelin-13(1–12) Is a Biologically Active ACE2 Metabolite of the Endogenous Cardiovascular Peptide [Pyr1]Apelin-13. Front. Neurosci. 2017, 11. https://doi.org/10.3389/fnins.2017.00092.
(5) Novotni, M. & Klein, R. Shape retrieval using 3D Zernike descriptors. Computer-Aided Design. 2004, 36, 1047–1062, doi: 10.1016/j.cad.2004.01.005.
(6) Di Rienzo, L.; Milanetti, E.; Lepore, R.; Olimpieri, P. P.; Tramontano, A. Superposition-Free Comparison and Clustering of Antibody Binding Sites: Implications for the Prediction of the Nature of Their Antigen. Sci. Rep. 2017, 7 (1), 45053. https://doi.org/10.1038/srep45053.
(7) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera—a Visualization System for Exploratory Research and Analysis. J. Comput. Chem. 2004, 25 (13), 1605–1612.
(8) Klein, A.; Ghosh, S. S.; Bao, F. S.; Giard, J.; Häme, Y.; Stavsky, E.; Lee, N.; Rossa, B.; Reuter, M.; Neto, E. C.; Keshavan, A. Mindboggling Morphometry of Human Brains. PLOS Comput. Biol. 2017, 13 (2), e1005350. https://doi.org/10.1371/journal.pcbi.1005350.
(9) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V. Scikit-Learn: Machine Learning in Python. J. Mach. Learn. Res. 2011, 12, 2825–2830.

PP2
The structure of the universal bacterial DNA repair protein Mfd dictates the pathogenicity of Bacillus cereus strains

Samantha Samson[1], Delphine Cormontagne[2], Seav-Ly Tran[2], Solène Albert[1,2], Nalini Ramarao[2] and Gwenaëlle André-Leroux[1]

1Maiage, INRAE, AgroParisTech, Université Paris-Saclay, 78350 Jouy-en-Josas, France.
2Micalis Institute, INRAE, AgroParisTech, Université Paris-Saclay, 78350 Jouy-en-Josas, France.

Background: The bacterial protein Mfd (mutation frequency decline) is highly conserved among bacteria and has been shown to be involved in bacterial DNA repair. We have recently shown that this protein confers bacterial resistance to the host nitrogen response produced during an infection by the immune system1, 2. Indeed, Mfd helps preserving DNA integrity and is involved in the repair following NO-induced DNA damage. Also, Mfd has been identified as a highly conserved "evolvability factor" that increases mutagenesis and the capacity of bacteria to evolve antibiotic existence3. Mfd is a nanomachine of 1200 amino acid residues organized into structural and functional modules4 that need to be timely remodeled along a functional cycle that we intend to decipher.

Objective: In this work, we aim to address the following issue: is the combination of sequence, structure of Mfd responsible for the clinic vs non-pathogenic phenotype of Bacillus cereus?

In silico Methods: We used a large collection of B. cereus strains with various pathogenicity signature, and we challenged them with very recent cryo-EM structures Mfd of E. coli, apo and holo as Mfd is in complex with RNA polymerase and/or DNA. In silico/In vivo Results: We could further establish a link between Mfd sequences, 3D structures remodeling and virulence. Protein sequence comparison, homology modeling and 3D structure/function analysis were used to determine specific domains and identify the dynamic of crucial interactions involved in the virulence of B. cereus. The residues identified were subjected to rational mutations. In parallel, using an in vivo insect model of infection, we showed that the mfd gene of a pathogenic strain could complement the avirulent phenotype of a non-pathogenic strain. These data strongly suggest that the 3D structure of Mfd plays an essential role in its function and might be a new and interesting way to discriminate pathogenic from harmless B. cereus strains. Further comprehension of its inhibition at a molecular level through anti evolution drug is now in progress.

Conclusions: As Mfd is widely conserved within bacteria, those findings could improve our understanding of the pathogenicity of a large spectrum of bacteria, especially the ESKAPE ones, considered as priority pathogens by the WHO5.

1 C. Darrigo, et al “The Bacterial Mfd Protein Prevents DNA Damage Induced by the Host Nitrogen Immune Response in a NER-Independent but RecBC-Dependent Pathway,” pp. 1–16, 2016, doi: 10.1371/journal.pone.0163321.
2 E. Guillemet, et al, “The bacterial DNA repair protein Mfd confers resistance to the host nitrogen immune response,” Nat. Publ. Gr., pp. 2–13, doi: 10.1038/srep29349.
3 M. N. Ragheb, et al, "Inhibiting the Evolution of Antibiotic Resistance", pp157-165, 2019, doi:10.1016/j.molcel.2018.10.015
4 J. Y. Kang, et al, "Strucural Basisi for transcription complex disruption by the Mfd translocase", p1-28, 2021, doi:10.7554/eLife.62117
5 https://www.who.int/news/item/27-02-2017-who-publishes-list-of-bacteria-for-which-new-antibiotics-are-urgently-needed

PP3
Identification of New Small Molecules Antiviral Agents as Protein-Protein Inhibitors of the ACE2 : Coronavirus Spike Interactions to Prevent the Entry of SARS-CoV-2 to the host Cell

Adnane Aouidate [a], Vincent Roy [a], Luigi Agrofoglio [a], Hugues de Rocquigny [b], Samia Aci-sèche [a], Pascal Bonnet [a]

a Institut de Chimie Organique et Analytique (ICOA UMR 7311), Université d'Orléans - Pôle de chimie, rue de Chartres - BP 6759, 45067 Orléans Cedex 2, France

b Inserm - U1259 MAVIVH Morphogenèse et Antigénicité du VIH et des Virus des Hépatites 10 boulevard Tonnellé - BP 3223 37032 Tours Cedex 1 - France)

Background: Angiotensin-converting enzyme 2 (ACE2) is the host cellular receptor that binds onto the surface spike protein of the novel, highly infectious and deadly SARS-CoV-2, responsible of COVID-19 pandemic1,2 the most challenging pandemic in the current century which having already caused more than a one hundred and sixty million infections and three million deaths worldwide3.
Objectives: One strategy to avoid the virus infection is to identify small molecules inhibitors targeting the protein- protein interface (PPI), preventing the virus entry into the host cells. These inhibitors of PPI (iPPI), that can impact the binding of the spike protein to its receptor, are of great interest as potential antiviral drugs against COVID-19. The SARS-CoV-2 spike/ACE2 complex represents thus a promising therapeutic target for identifying such compounds. iPPIs might provide alternative therapeutic solutions to peptides or antibodies targeting PPI that are suitable for oral administration, metabolically stable4, less-prone to mutation and strain-sensitive and less-immunogenic5.
Methods: In the search of inhibitors against COVID-19, we developed a computationally driven approach aiming to identify SARS-CoV-2 spike/ACE2 inhibitors using consensus structure based virtual screening (SBVS), various scoring methods and molecular dynamics simulations. For this purpose, we used 3 different docking software, namely Smina6, rDock7 and Plants8. The obtained poses from each software were rescored with two methods (SPLIF-Score 9 and a combination of Tanimoto coefficient and docking software scores).The protocol was first developed by using 16 molecules designed by our chemist collaborators at ICOA and then applied to screen an iPPI dedicated chemical library of ~ 11000 compounds.
Results: Based on this rescoring, the three top poses, one by docking software, were selected for each molecule. Compounds that exhibited similar predicted binding modes with the three docking software were kept. This strategy highlighted few promising compounds that are currently experimentally evaluated on in vitro assays for their antiviral activity.
Conclusions: These rigorous in silico studies led to the identification of promising small molecules inhibitors of ACE2: SARS-CoV-2 spike that are currently assessed experimentally by in vitro assays for their antiviral activity.

(1) Matheson, N. J.; Lehner, P. J. How Does SARS-CoV-2 Cause COVID-19? Science (80-.). 2020, 369 (6503), 510–511. https://doi.org/10.1126/science.abc6156.
(2) Moore, J. B.; June, C. H. Cytokine Release Syndrome in Severe COVID-19. Science (80-.). 2020, 368 (6490), 473–474. https://doi.org/10.1126/science.abb8925.
(3) World Health Organization. COVID-19 Weekly Epidemiological Update 22. World Heal. Organ. 2021, No. December, 1–3.
(4) Zhao, Y.; Bai, L.; Liu, L.; McEachern, D.; Stuckey, J. A.; Meagher, J. L.; Yang, C.-Y.; Ran, X.; Zhou, B.; Hu, Y.; Li, X.; Wen, B.; Zhao, T.; Li, S.; Sun, D.; Wang, S. Structure-Based Discovery of 4-(6- Methoxy-2-Methyl-4-(Quinolin-4-Yl)-9 H -Pyrimido[4,5- b]Indol-7-Yl)-3,5-Dimethylisoxazole (CD161) as a Potent and Orally Bioavailable BET Bromodomain Inhibitor. J. Med. Chem. 2017, 60 (9), 3887–3901. https://doi.org/10.1021/acs.jmedchem.7b00193.
(5) Bojadzic, D.; Buchwald, P. Toward Small-Molecule Inhibition of Protein–Protein Interactions: General Aspects and Recent Progress in Targeting Costimulatory and Coinhibitory (Immune Checkpoint) Interactions. Curr. Top. Med. Chem. 2018, 18 (8), 674–699. https://doi.org/10.2174/1568026618666180531092503.
(6) Koes, D. R.; Baumgartner, M. P.; Camacho, C. J. Lessons Learned in Empirical Scoring with Smina from the CSAR 2011 Benchmarking Exercise. J. Chem. Inf. Model. 2013, 53 (8), 1893–1904. https://doi.org/10.1021/ci300604z.
(7) Ruiz-Carmona, S.; Alvarez-Garcia, D.; Foloppe, N.; Garmendia-Doval, A. B.; Juhos, S.; Schmidtke, P.; Barril, X.; Hubbard, R. E.; Morley, S. D. RDock: A Fast, Versatile and Open Source Program for Docking Ligands to Proteins and Nucleic Acids. PLoS Comput. Biol. 2014, 10 (4), e1003571. https://doi.org/10.1371/journal.pcbi.1003571.
(8) Korb, O.; Stützle, T.; Exner, T. E. PLANTS: Application of Ant Colony Optimization to Structure- Based Drug Design; 2006; pp 247–258. https://doi.org/10.1007/11839088_22.
(9) Da, C.; Kireev, D. Structural Protein–Ligand Interaction Fingerprints (SPLIF) for Structure-Based Virtual Screening: Method and Benchmark Study. J. Chem. Inf. Model. 2014, 54 (9), 2555–2561. https://doi.org/10.1021/ci500319f.

PP4
Allosteric modulation of the nicotinic acetylcholine receptor α7 by ivermectin

Mariia Avstrikova[a,b] , Ryan E. Hibbs[c,d] , Jean-Pierre Changeux[e], and Marco Cecchini[b]

a Channel-Receptors Unit, UMR 3571, CNRS, Institut Pasteur, Paris 75015, France
b Institut de Chimie de Strasbourg, UMR 7177, CNRS, Université de Strasbourg, Strasbourg Cedex 67083, France
c Department of Neuroscience, Peter O'Donnell Jr. Brain Institute, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA
d Department of Biophysics, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA
e Department of Neuroscience, CNRS UMR 3571 Institut Pasteur & Collège de France, 75005 Paris, France

Nicotinic acetylcholine receptors (nAChRs) play a critical role in nervous signal transduction, being involved in fundamental cognitive functions such as memory, learning, attention1. Moreover, recent studies established a connection between nAChRs and the pathophysiology of SARS-CoV-22,3 infection, making them interesting pharmacological targets in the context of the COVID-19 pandemic.

Apart from the orthosteric binding sites for neurotransmitters, nAChRs possess several topographically distinct binding sites dedicated to allosteric modulators that may control the functional conformational transitions of the nAChR 1. In the last years, several high-resolution structures of nAChRs have been published. They represent different receptor's subtypes and physiological states, and it becomes of particular interest to identify these allosteric modulatory sites4.

This project aims at investigating the allosteric binding site of nAChRs dedicated to ivermectin, an anthelmintic drug that is known to potentiate allosterically the α7 nAChR5 and to show an anti-COVID-19 action on Hamsters infected by SARS-CoV-2 virus6.

Comparative analysis of glycine receptor structures solved in complex with IVM7, and recently solved structures of α7 nAChR8 highlights significant differences in the pocket, corresponding to the ivermectin binding site. Thus, there is a need to model such site in nAChRs. To reach this goal, we performed all-atom Molecular Dynamics (MD) simulations of the protein embedded in a lipid bilayer and extracted the most populated cluster centers, which provide relaxed and statistically relevant conformations. Using these models the ivermectin binding sites was explored by blind docking9 with free-energy rescoring of the docking predictions via the Molecular Mechanics-Generalized Born Surface Area (MM/GBSA) end-point free energy method10.

Our studies reveal two possible binding sites for ivermectin located in the transmembrane domain of α7 nAChR. Interestingly, the most favorable binding mode differs in the active versus the desensitized states of the receptor. The structural characterization of the allosteric transmembrane site in α7 nAChR with atomic resolution is useful for subsequent virtual screening campaigns on the quest for novel modulators of nAChRs, in particular as possible anti-covid agents.

1 Cecchini M, Changeux JP. The nicotinic acetylcholine receptor and its prokaryotic homologues: Structure, conformational transitions & allosteric modulation, Neuropharmacology, 2015, 137-149.
2 Changeux JP, Amoura Z, Rey FA, Miyara M. A nicotinic hypothesis for Covid-19 with preventive and therapeutic implications, Comptes Rendus - Biol., 2020, 33-39.
3 Oliveira ASF, Ibarra AA, Bermudez I, et al. A potential interaction between the SARS-CoV-2 spike protein and nicotinic acetylcholine receptors. Biophys J., 2021, 983-993.
4 Zhao Y, Liu S, Zhou Y, et al. Structural basis of human α7 nicotinic acetylcholine receptor activation, Cell Res.,2021, 1-4.
5 Krause RM, Buisson B, Bertrand S, et al. Ivermectin: A positive allosteric effector of the α7 neuronal nicotinic acetylcholine receptor, Mol Pharmacol, 1998, 283-294.
6 Dias de Melo G, Sonthonnax F, Lepousez G, et al. A combination of two human monoclonal antibodies cures symptomatic rabies, EMBO Mol Med., 2020, e12628.
7 Huang X, Chen H, Shaffer PL. Crystal Structures of Human GlyRα3 Bound to Ivermectin, Structure, 2017, 945-950.
8 Noviello CM, Gharpure A, Mukhtasimova N, et al. Structure and gating mechanism of the α7 nicotinic acetylcholine receptor, Cell , 2021, 1-14.
9 Hassan NM, Alhossary AA, Mu Y, Kwoh CK. Protein-Ligand Blind Docking Using QuickVina-W with Inter-Process Spatio- Temporal Integration, Sci Rep., 2017, 1-13.
10 Massova I, Kollman PA. Combined molecular mechanical and continuum solvent approach (MM- PBSA/GBSA) to predict ligand binding, Perspect Drug Discov Des., 2000, 113-135.

PP5
DockSurf: a molecular modelling software for predicting protein/surface association

Florent BARBAULT[a], Eric BREMOND[a], Julien REY[b], Pierre TUFFERY[b] and François MAUREL[a]

a Laboratoire ITODYS, CNRS UMR 7086, Université de Paris, 15 rue de Baïf, 75013 Paris
b Laboratoire BFA, CNRS UMR 8251, Université de Paris, 35 rue Hélène Brion, 75013 Paris

One key aspect in the development of bionanotechnologies is the elucidation of the structural interfaces between proteins and inorganic surfaces1. Despite its wide importance, the interfacial structures between proteins and metallic surfaces remain elusive and the lack of experimental investigation refrains the development of many devices2.

In this context, molecular simulations represent an affordable alternative to understand the underlying physical processes which occur at the protein/surface interface. According to the size of proteins, molecular dynamics simulations are the reference technique but required an initial structural organization for studying the interaction with a surface. To overcome this drawback, we propose to consider the generation of protein/surface structure as a molecular docking problem for which is target is an homogenous plan.

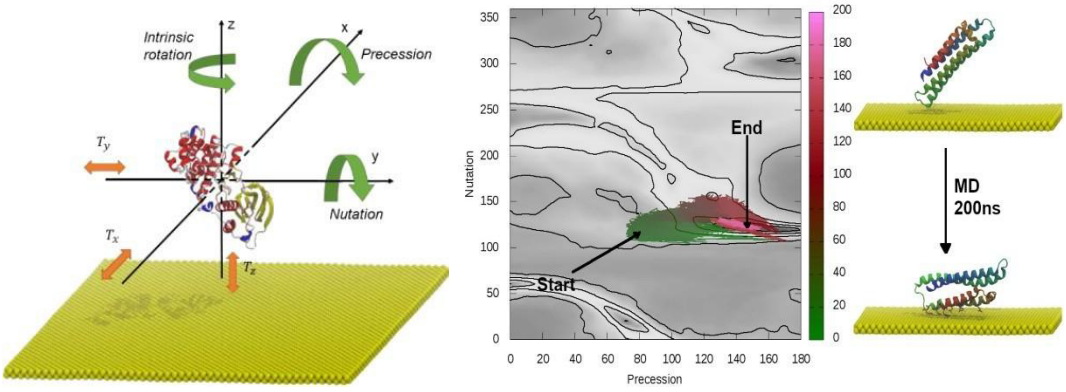


Figure 1: left, Euler's angles used for exploration. Right, example of investigation for the apoliprotein.

A molecular docking software can be viewed as a conformation exploration investigation and an interaction energy computation. Our approach and originality is to consider the conformational exploration with Euler's angles (cf. figure 1), like a plane above a surface, providing this way a cartography instead an unique structure. Secondly, the interaction energies were derived from QM computations for a set of small molecules that describe protein atom-types, and implemented in a DLVO potential3 for considering large systems such as proteins. Validation of DockSurf software has been made for corona proteins with Au{111} surface (cf. figure 1) and provide enthusiastic results. This software will be soon implemented in the RPBS platform to provide a wide access to the scientific community.

1 Ozboyaci, M.; Kokh, D. B.; Corni, S.; Wade, R. C. Modeling and Simulation of Protein–Surface Interactions: Achievements and Challenges. Quart. Rev. Biophys. 2016, 49, e4.
2 Dalsin, J. L.; Messersmith, P. B. Bioinspired Antifouling Polymers. Materials Today 2005, 8 (9), 38–46.
3 Overbeek, J. T. G. Recent Developments in the Understanding of Colloid Stability. Journal of Colloid and Interface Science 1977, 58 (2), 15.

PP6
SINAPs: A software tool for analysis and visualization of interaction networks of molecular dynamics simulations

Corentin BEDART [a,b], Nicolas RENAULT [b], Philippe CHAVATTE [b], Adeline PORCHERIE [a], Abderrahim LACHGAR [a], Monique CAPRON [a], Amaury FARCE [b]

a Par'Immune, Bio-incubateur Eurasanté, 70 rue du Dr. Yersin, 59120, Loos-Lez-Lille, France
b Univ. Lille, Inserm, CHU Lille, U1286 - Infinite - Institute for Translational Research in Inflammation, F-59000 Lille, France

Introduction. To study enzymatic mechanisms of complex proteins, we had to perform many molecular dynamics simulations generating a huge amount of data. The analysis of such an amount of data to identify the main differences is a bottleneck. The simple visualization of the results is also a complex point. However, since there was no appropriate tool to perform these steps together, we developed a software written in Python and we named it SINAPs (Structural Interaction Networks Analysis Protocols). This proprietary tool allows us to identify either similarities or differences of interactions and lead us to visualize the results through an extension to UCSF Chimera.

Methods. The analysis module takes as input molecular dynamics simulations or crystallographic structures, to calculate non-bonded interactions with precise control of the definition parameters for each type. The visualization module allows the observation of the results by displaying each type of interaction via a finely controlled mode of representation. The SINAPs software was validated on several biological targets described in the literature: GLUT-1 and A2AR. In addition, an exploratory study was done using classical molecular dynamics of Par'Immune's drug candidate, P28GST, to assist the study of its molecular mechanism of action.

Results. The molecular dynamics study of two conformations of the Glucose Transporter 1 allowed us to identify the main interaction networks previously described in the literature. We specifically pinpointed the isolation mechanism of the ligand-binding site from the extracellular environment, or the mechanism allowing the opening of the ligand-binding site towards the intracellular side. The study of crystallographic conformations of the Adenosine A2A receptor allowed us to show the main interactions governing the activation steps and ligand-binding mode, confirming the similarity of the interactions from different agonists, and identifying the differential interactions made by an inhibitor. The study of P28GST using SINAPs validated the molecular dynamics simulations performed by finding specific interactions already described in the literature, supporting assumptions regarding the enzymatic mechanism of glutathione activation, and allowing to highlight several amino acids impacting this activation.

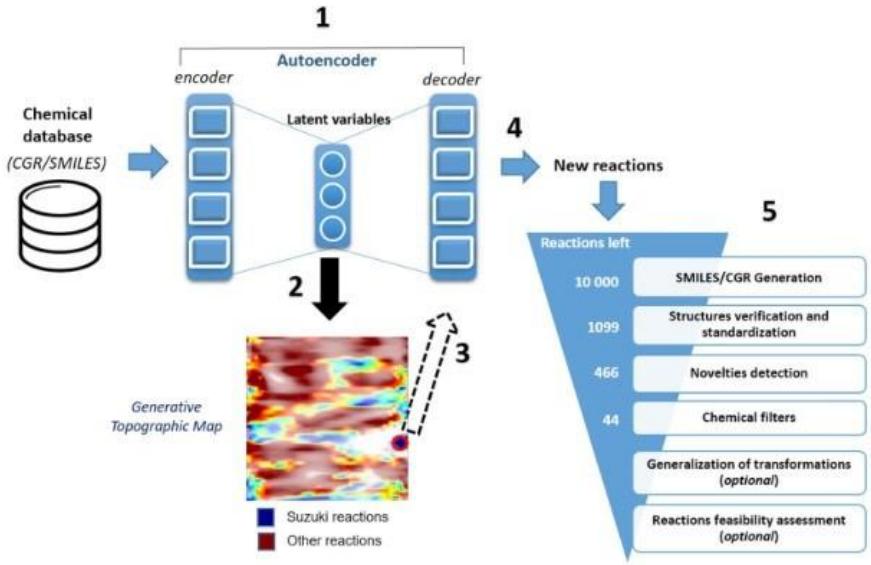
Conclusion. SINAPs, which will be soon available to the scientific community, may accelerate the structural analysis of different protein conformations by highlighting the similarities and/or differences of non-covalent molecular interactions while suggesting a simplified and customizable visualization.

PP7
Discovery of novel chemical reactions by Deep Generative Recurrent Neural Network

William Bort[1], Igor I. Baskin[1,4], Timur Gimadiev[3], Artem Mukanov[2], Ramil Nugmanov[2], Pavel Sidorov[3], Gilles Marcou[1], Dragos Horvath[1], Olga Klimchuk[1], Timur Madzhidov[2] & Alexandre Varnek[1,3]

[1] Laboratory of Chemoinformatics, UMR 7140 CNRS, University of Strasbourg, France.
[2] Kazan Federal University, Russia
[3] Institute for Chemical Reaction Design and Discovery (WPI-ICReDD), Hokkaido University, Sapporo, Japan.
[4] Department of Materials Science and Engineering, Technion – Israel Institute of Technology, 3200003 Haifa, Israel.*

In this work, we demonstrate how Artificial Intelligence tools can successfully propose novel stoichiometrically coherent chemical reactions. Being coupled to reaction space cartography, de novo reaction design may be focused on the desired reaction class. Here, a sequence-to-sequence autoencoder with bidirectional Long Short-Term Memory layers was trained on on-purpose developed "SMILES/CGR" strings, encoding reactions of the USPTO database. The autoencoder latent space was visualized on a generative topographic map. Novel latent space points were sampled around an area populated by Suzuki reactions and decoded to corresponding reactions. The feasibility of generated reactions has been assessed in quantum mechanics DFT calculations. Among generated reactions we discovered transformations with 13 new reaction centers which did not occur in the training set. Five out of 13 transformations were then found in the reaction databases (not used in the model training), thus showing the reliability of our approach to generate new synthetically feasible reactions.



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

PP8
Integrative structural modeling reveals functional molecular switches of human G protein- coupled bitter-taste receptors

Cédric Bouysset[a], Jérémie Topin[a]*, Yiseul Kim[b], Jody Pacalon[a], MeeRa Rhyu[b], Sébastien Fiorucci[a]*, Jérôme Golebiowski[a,c]

a Institut de Chimie de Nice UMR7272, Université Côte d'Azur, CNRS, France
b Korea Food Research Institute, 245 Iseo-myeon, Wanu-gun, Jeollabuk-do 55365, Republic of Korea
c Department of Brain & Cognitive Sciences, DGIST, 333, Techno JungAng, Daero, HyeongPoong Myeon, Daegu, 711-873, Republic of Korea

*jeremie.topin@univ-cotedazur.fr, sebastien.fiorucci@univ-cotedazur.fr

Among the multiple perceptions triggered by the gustatory system, bitterness is usually associated with the avoidance of food and is believed to have evolved to alert us against the consumption of toxic plants.1 The human genome possesses 25 functional TAS2R genes encoding bitter taste receptors that are differentially activated by a broad range of chemically and structurally diverse bitter compounds. TAS2Rs belong to the G protein-coupled receptor (GPCR) family. While several GPCR structures have been experimentally solved, none of them are chemosensory receptors. Without such key structural information, predicting the activity and mechanism of action of bitter molecules on TAS2Rs mostly relies on molecular modeling. Here, we present an optimized computational protocol to build relevant 3D models of bitter taste receptors (Figure 1). Homology modeling allows us to model TAS2Rs based on their sequence similarity with a template for which an experimental structure is available. By combining with site-directed mutagenesis followed by in vitro functional assays, we show how our optimized sequence alignment between TAS2Rs, olfactory receptors (ORs) and non-OR GPCRs accurately recapitulates and predicts experimental data.

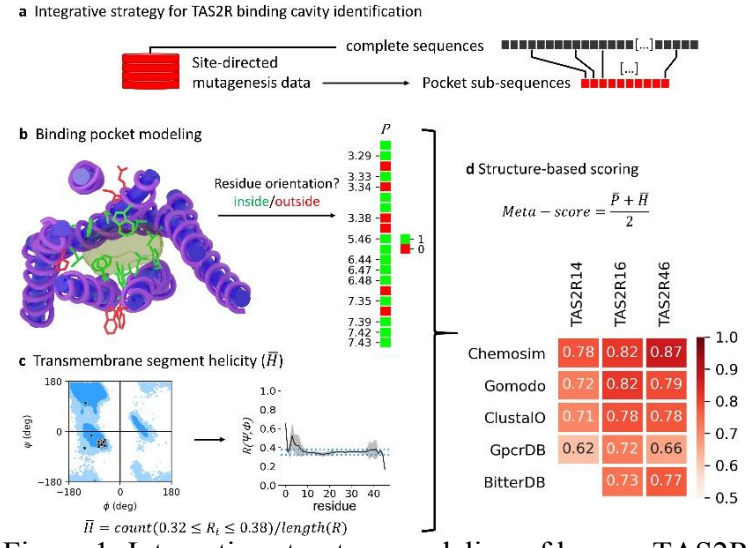


Figure 1: Integrative structure modeling of human TAS2R.

1 Lindemann B. Receptors and transduction in taste. Nature, 2001, 413, 219-225.

PP9
Molecular modelling of hepatitis B virus Core protein C-terminal domain: Implications for nucleocapsid trafficking
Modélisation moléculaire du domaine C-terminal de la protéine Core du virus de l'hépatite B: Implications pour le trafic intracellulaire de la nucléocapside

Jean-Charles Carvaillo[1*], Stéphane Bressanelli[1], Yves Boulard[1]

1 Université Paris-Saclay, CEA, CNRS, Institute for Integrative Biology of the Cell (I2BC), 91198, Gif-sur-Yvette, France.

* jean-charles.carvaillo@i2bc.paris-saclay.fr

The Core protein of the hepatitis B virus is composed of an N-terminal structured assembly domain (NTD, residues 1-149) followed by an intrinsically disordered C-terminal domain (CTD, residues 150-183). Crystal structures of recombinant (E. coli-produced) NTD show that in the absence of genetic material, residues 1-149 are organized into a helical domain. The NTD forms dimers with a protruding spike composed of a four-helix bundle, while lateral contacts between other helices and loops allow 120 dimers to assemble into an icosahedral capsid [1]. The positively-charged CTD is located inside this capsid, where it interacts with the pre- genomic RNA upon immature nucleocapsid assembly in the cytosol of the infected cell. However, the CTD also harbours the nuclear localisation signals (NLS) that direct mature nucleocapsids to the nucleus. Thus, some at least of the 240 CTD in a mature nucleocapsid must be displayed through the pores that fenestrate the capsid. This structural feature is not clearly visible, even in recent high-resolution cryo-EM structures of full-length recombinant nucleocapsid-like particles [2], so that there is some controversy even as to which type of pore allows display of the CTD [3,4] and how far the CTD exits the pore [5,6].
Using molecular dynamics simulations of Core dimers, we show that the CTD tends to interact with its NTD in such a way that it would either sit below 3-fold and quasi-3-fold pores, or be threaded though the same pores up to the tip of the Core spike. Further molecular modelling in the context of the capsid, including biased simulations, shows atomic-level features of CTD egress through 3-fold-type pores. We discuss the agreement of our results with published functional and structural properties of CTD display by the mature HBV nucleocapsid.

[1] S. A. Wynne, R. A. Crowther, A. G. Leslie, Mol. Cell 1999, 3, 771.
[2] B. Böttcher, M. Nassal, J. Mol. Biol. 2018, 430, 4941.
[3] C. Chen, J. C.-Y. Wang, E. E. Pierson, D. Z. Keifer, M. Delaleau, L. Gallucci, C. Cazenave, M. Kann, M. F. Jarrold, A. Zlotnick, PLoS Pathog. 2016, 12, e1005802.
[4] X. Yu, L. Jin, J. Jih, C. Shih, Z. H. Zhou, PLoS ONE 2013, 8, e69729.
[5] L. Selzer, R. Kant, J. C.-Y. Wang, B. Bothner, A. Zlotnick, J. Biol. Chem. 2015, 290, 28584.
[6] J. Heger-Stevic, P. Zimmermann, L. Lecoq, B. Böttcher, M. Nassal, PLoS Pathog. 2018, 14, e1007488.

PP10
Small-molecule inhibitors of SARS-CoV-2 interfering with Spike / ACE2 interactions

Laurent Chaloin[a], Aria Gheeraert[b,c], Olivier Moncorgé[a], Laurent Vuillon[b], Ivan Rivalta[c,d], Marie-
Dominique Devignes[e] and Bernard Maigret[e]

a Infectious disease Research Institute of Montpellier (IRIM), CNRS, University of Montpellier, France
b Laboratory of Mathematics (LAMA), CNRS, University of Savoie Mont Blanc, France
c Dipartimento di Chimica Industriale "Toso Montanari", Università di Bologna, Bologna, Italy
d Laboratory of Theoretical Chemistry, ENS Lyon, CNRS, Lyon, France
e LORIA Institute, Capsid team, CNRS - INRIA – University of Lorraine, France

Background
The Covid-19 pandemic started in December 2019 has provoked an unprecedent sanitary crisis leading to 109, 000 deaths in France and more than 3.5 million worldwide, to date. The causing agent is the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) that belongs to the betacoronavirus genus and infects mainly respiratory cells. Besides the recent progress in delivering a vaccine, the impossibility to protect immunodeficient persons in addition to the emergence of new variants represents a major concern and the development of new antiviral treatments are urgently needed. In order to propagate, the virus must first infect pulmonary cells and this critical step is achieved through the interaction between the transmembrane viral spike glycoprotein and human angiotensin-converting enzyme-2 receptors (hACE2)1,2. This attachment to cells is mediated through specific recognition between the receptor binding domain (RBD) of spike and the N-terminal helix of hACE23.

Objectives
Based on available cryo-EM structures of the Spike RBD/hACE2 complex, we explored by in silico screening the discovery of small-molecules able to interfere with this crucial step of virus binding to cell. To increase the success rate of finding active hits, we designed a specific chemical library including all criteria described for protein-protein interaction inhibitors and used sequential docking steps to improve accuracy predictions.

Results and conclusions
From a virtual screening of 400, 000 compounds targeting the major groove of spike RBD, 300 best-ranked molecules were re-docked using more stringent parameters of the search algorithm and scoring function. This refinement led to the selection of 23 hit compounds which were purchased for in vitro evaluation on SARS-CoV-2 infected cells. Two hits (VS012 / VS014) were able to inhibit viral replication in the low micromolar range both in simian or in primary human epithelia cells. Leads are currently being optimized by chemoinformatic approaches and new structural analogs will be evaluated for their antiviral activity. In parallel, molecular dynamics simulations have been performed on the full-length spike (wild-type and variants of concern) to decipher the mechanism of receptor binding at the atomic level, a critical stage that is likely responsible of the increase in viral infectivity for some variants.

1. Wang, Q. et al. Structural and Functional Basis of SARS-CoV-2 Entry by Using Human ACE2. Cell 181, 894-904.e9 (2020).
2. Walls, A. C. et al. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. Cell 181, 281-292.e6 (2020).
3. Lan, J. et al. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. Nature 581, 215– 220 (2020).

PP11
Sugar stay or sugar go: role of the sugar moiety on the biophysical properties of mycobacterial mycocerosate containing lipids

Clément Birbes[a], Adrien Schahl[b], Evert Haanappel[a], Valérie Réat[a], Georges Czaplicki[a], Franck Jolibois[c], Catherine Astarie-Dequeker[d], Christophe Guilhot[d], Olivier Saurel[a], Matthieu Chavent[b]

a Integrative Biological NMR, IPBS, Université de Toulouse, CNRS, Université Paul Sabatier, 31400 Toulouse, France.
b Multiscale modeling of host-pathogens interactions and immune response, IPBS, Université de Toulouse, CNRS, Université Paul Sabatier, 31400 Toulouse, France.
c Laboratoire de Physique et Chimie des Nano-objets, Institut National des Sciences Appliquées, Université de Toulouse, CNRS, Université Paul Sabatier, F-31077 Toulouse, France.
d Molecular mycobacterial pathogenesis, IPBS, Université de Toulouse, CNRS, Université Paul Sabatier, 31400 Toulouse, France.

There is growing evidence showing that lipid molecules can modulate membrane protein functions by directly interacting with them or by modifying the biophysical properties of the cellular membranes. This property relies on lipid chemical structures.

While the lipid structure-function relationship has been thoroughly studied for lipids constituting eukaryotic membranes, such a relationship remains elusive for numerous bacterial lipids. This is especially true for the lipid virulence factors synthetized by mycobacteria such as Mycobacterium tuberculosis (Mtb), the causative agent of the human disease Tuberculosis. These Mtb lipids have specific structural characteristics such as very long and complex acyl- chains that differentiate them from eukaryotic phospholipids. They have also a broad spectrum of actions on host membranes.

Among these lipids the mycocerosate containing lipids, Phthiocerol DIMycocerosates (PDIM) and Phenolic GLycolipids (PGL), are key molecules for Mtb pathogenicity[1]. Combining Molecular Dynamics (MD) simulations with NMR and cell biology experiments, we have recently shown how the peculiar shape of PDIM lipids can modify the biophysical properties of macrophage membrane and modulate important immune cell functions such as phagocytosis[2].

Here, we are extending this work by studying PGL lipids. These lipids differ from PDIM molecules by the addition of one (PGL of M. Bovis) or three (PGL of Mtb) methylated sugar molecules at the extremity of the phtiocerol (acyl) chain. By using multiscale MD simulations and NMR assays, we will show how the addition of the different sugar moieties can change the structure and the biophysical properties of these different molecules. From a biophysical perspective, this work is shedding new light on how structural changes, such as the addition of sugar moieties, can affect the lipid structure-function relationship. From a biological point of view, this work can contextualize the evolution of mycobacterial strains expressing different types of PGL lipids.

Arbués, A., Lugo-Villarino, G., Neyrolles, O., Guilhot, C. & Astarie-Dequeker, C. Playing hide-and- seek with host macrophages through the use of mycobacterial cell envelope phthiocerol dimycocerosates and phenolic glycolipids. Frontiers in cellular and infection microbiology 4, 173 (2014).

Augenstreich, J. et al. The conical shape of DIM lipids promotes Mycobacterium tuberculosis infection of macrophages. Proc Natl Acad Sci USA 116, 25649–25658 (2019).

evotec #RESEARCHNEVERSTOPS

Evotec SAS

PP12
Application of Bayesian Optimization in Drug Discovery

Colliandre Lionel[a], Rabal Gracia Obdulia[a], Diaz Gonzalez Constantino[a], Hristozov Dimitar[b], Bodkin Michael[b]

a Evotec SAS, Campus Curie, 195 route d'Espagne, 31100 Toulouse, France
b Evotec, 114 Innovation Dr, Park Dr, Milton, Abingdon OX14 4RZ, Royaume-Uni

Drug discovery requires the exploration of a vast chemical space (1) to find the right compound with desired properties. The main question in this exploration is "which compound should I make and test next?". The answer for any particular project depends on the knowledge that has been acquired so far and on the allocated resources.

Bayesian optimization (BO) is an efficient method to perform an active search for the parameters which optimize an objective function (2). At each step, it selects a new set of parameters that should provide the maximum amount of information over the parameters' space. The objective function is then evaluated using these parameters and the results are fed back. Hence, it can chose the next set of parameters in a better informed manner allowing a faster exploration of the space toward the global optimum.

BO has been applied to various drug discovery problems: to optimize molecular conformations (3), for the robotic automation of chemical synthesis (4), for hit and lead optimization (5, 6). The molecular optimization is generally performed in the molecular descriptors space (e.g. fingerprints). These molecular descriptors support only an active learning approach where the next compound to make is selected from a pre-defined list. To fully explore the BO capabilities, the ability to convert an optimal set of parameters (descriptors) into a molecule is required. To do so, BO can be combined with deep learning methods that generate molecular descriptors that can be decoded back into a chemical structure (7).

In analogy with the work of Griffiths et al. (8), we implemented a BO pipeline based on the CDDD descriptors (7). These descriptor vectors are used in the construction of a Gaussian Process model. The model can be used to score a pre-defined list of molecules (Active Learning), or to generate a new descriptor vector that will be decoded into a chemical structure.

In this presentation, we will show the retrospective application of our BO implementation on external and internal datasets. We will demonstrate the power of the method to explore efficiently the available chemical space in order to improve one or several properties of compounds.

1- Polishchuk P. G. et al., Estimation of the size of drug-like chemical space based on GDB-17 data, J. Comput. Aided Mol. Des., 2013, 27, 675–679
2- Shahriari B. et al. Taking the Human Out of the Loop: A Review of Bayesian Optimization, Proceeding of the IEE, 2016, 104, 148-175
3- Chan L., Bayesian optimization for conformer generation, 2019, J. Cheminform., 11, 32
4- Roch L. et al., ChemOS: An orchestration software to democratize autonomous discovery, 2020, PLOS ONE, 15(4)
5- Pyzer-Knapp E.O. Bayesian optimization for accelerated drug discovery, IBM J. RES. & DEV., 2018, 62, 6 6- Wang Y. et al., Bayesian Active Drug Discovery, ICML, 2020, workshop paper
6- Winter R. et al. Learning continuous and data-driven molecular descriptors by translating equivalent chemical representations, 2019, Chem. Sci., 10, 1692-1701
7- Griffiths R. et Hernandez-Lobato J.M., Constrained Bayesian optimization for automatic chemical design using variational autoencoders, 2020, Chem. Sci., 11, 577-586

PP13
PYTHIA: Deep Learning Approach For Local Protein Conformation Prediction

Gabriel CRETIN[1,2], Tatiana GALOCHKINA[1,2], Alexandre G. DE BREVERN[1,2] and Jean-Christophe GELLY[1,2]

1 Université de Paris, Biologie Intégrée du Globule Rouge, UMR_S1134, BIGR, INSERM, F-75015, Paris, France
2 Laboratoire d'Excellence GR-Ex, Paris, France

Corresponding Author: jean-christophe.gelly@u-paris.fr

Background and objectives
Protein Blocks (PBs) is a structural alphabet describing the local protein conformation with higher precision than classical secondary structures1. PBs correspond to 16 structural conformational states, which can be adopted by five consecutive amino acids. Encoding of complex protein structures (3D) in a PB sequence (1D) has already been successfully applied to protein structure alignment and protein structure prediction[2,3]. In the current study we developed a deep learning model for prediction of the protein local conformations in terms of PB directly from amino acid sequence.

Methods
Each amino acid is encoded by 58 physico-chemical properties4 and a position-specific substitution matrix (PSSM) generated by PSI-BLAST. We performed a 10 fold cross-validation on a non-redundant dataset of 9638 protein chains from PDB. Prediction was performed using a deep inception-inside-inception convolutional neural architecture.

Results
The developed model named PYTHIA (Predicting Any Conformation at High Accuracy) clearly outperforms the reference method for PB prediction LOCUSTRA5. The mean accuracy (Q16) equals 70% for PYTHIA and 61% for LOCUSTRA. Furthermore, PYTHIA outperforms LOCUSTRA on every PB class even for the smallest ones such as 'g' (MCC equal 0.209 for PYTHIA vs. 0.154 for LOCUSTRA) and 'j' (MCC equal 0.315 for PYTHIA vs. 0.223 for LOCUSTRA).

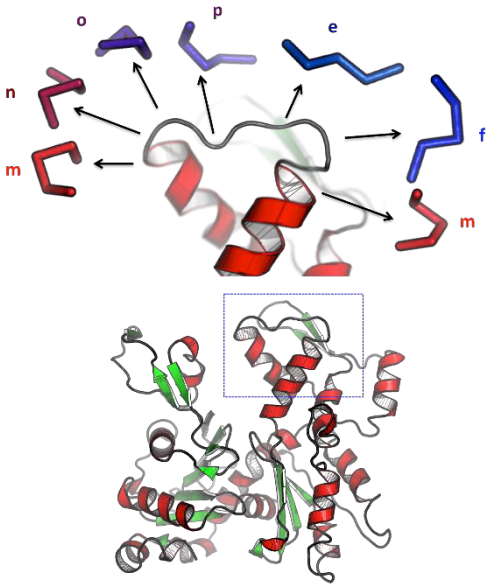


Figure: Description of local protein conformation using Protein Blocks[1].

1 de Brevern et al. (2000), Bayesian probabilistic approach for predicting backbone structures in terms of protein blocks. Proteins, 41: 271-287.
2 Gelly J-C et al. iPBA: a tool for protein structure comparison using sequence alignment strategies. Nucleic Acids Res. 2011;39: W18–W23.
3 Ghouzam Y et al. ORION: a web server for protein fold recognition and structure prediction using evolutionary hybrid profiles. Sci Rep. 2016;6: 1–10.
4 van Westen GJP et al. Benchmarking of protein descriptor sets in proteochemometric modeling (part 1): comparative study of 13 amino acid descriptor sets. J Cheminform. 2013;5: 41.
5 Olav Z. et al. LOCUSTRA: Accurate Prediction of Local Protein Structure Using a Two-Layer Support Vector Machine Approach. Journal of Chemical Information and Modeling 2008 48 (9), 1903-1908

PP14
LIMONADA: a database dedicated to the simulation of complex lipid membranes

Jean-Marc Crowet[a], Nicolas Belloy[a], Sébastien Buchoux[b], Catherine Sarazin[b], Manuel Dauchez[a], Laurence Lins[c]

a MEDyC, Université de Reims Champagne-Ardenne, Reims, France
b GEC, Université de Picardie Jules Verne, Amiens, France
c LBMI, Gembloux Agro-Bio Tech, University of Liège, Gembloux, Belgium

Biological membranes can be composed of hundreds of different lipids distributed heterogeneously between and within lipid leaflets. Their ratios are organism and organelle specific and several studies have yielded detailed membrane compositions of different cell types. Moreover, lipidomic studies have established that there could be over 100,000 different lipids. Besides, the conjunction of a greater computing power and methodological developments has led to an ever-increasing size and complexity of the lipid systems simulated by molecular dynamics (MD). Tens of model membranes with several lipid species designed to represent specific biological membranes have already been simulated. Modeling studies start then to encompass the wide diversity of biological membranes. Determination of complex lipid compositions is not trivial and simulating these membranes to equilibration can be quite computationally- demanding while only being a prerequisite for further simulations. Moreover, there are several forcefields (FF) available for membrane MD simulations and one major issue is that the number of lipid topologies is strongly FF-dependent. In this work, we present LIMONADA (Lipid Membrane Open Network And Database; https://limonada.univ-reims.fr/) which has been developed as an open database (https://github.com/limonadaMD/) allowing to handle the various aspects of lipid membrane simulation from a lipidomic background. LIMONADA presents published membrane patches with their simulation files and the cellular membrane it models in a simulation-ready fashion. Limonada uses the lipid classification established by the LipidMaps consortium and every lipid is unequivocally identified by a four-digit ID (i.e. compliant with e.g. pdb files). From this basis, registered users can add new membranes, lipids, topologies and/or forcefields to the database with the only limitation that each addition must be sustained by a reference paper and that the primary source of the files stored by Limonada are also provided.

PP15
Development of a chemogenomics library for phenotypic screening

Bryan Dafniet[a], Natacha Cerisier[a], Arnaud Gohier[b], Anaelle Clary[b], Pierre Ducrot[c], Thierry Dorval[b], David Brown[b], Karine Audouze[d], Olivier Taboureau[a]

a Université de Paris, INSERM U1133, CNRS UMR 8251, 75006 Paris, France
b Institut de recherche Servier, 125 Chemin de Ronde, 78290 Croissy-Sur-Seine, France
c AstraZeneca Cambridge, 1 Francis Crick Avenue, Cambridge, Royaume-Uni
D Université de Paris, INSERM UMR S-1124, 75006 Paris, France

With the development of advanced technologies in cell-based phenotypic screening, phenotypic drug discovery (PDD) strategies have re-emerged as promising approaches in the identification and development of novel and safe drugs[1]. However, phenotypic screening does not rely on knowledge of specific drug targets and needs to be combined with chemical biology approaches in order to identify therapeutic targets and mechanism of actions induced by drugs and associated with an observable phenotype[2].
In this study, we developed a system pharmacology network integrating drug-target-pathway-disease relationships as well as morphological profile from an existing high content imaging-based high-throughput phenotypic profiling assay known as "Cell Painting"[3]. Furthermore, from this network, a chemogenomic library of 5000 small molecules that represent a large and diverse panel of drug targets involved in diverse biological effects and diseases has been developed. Such a platform and a chemogenomic library could assist in the target identification and mechanism deconvolution of some phenotypic assays. The usefulness of the platform is illustrated through examples.

1. Childers WE, Elokely KM, Abou-Gharbia M (2020) The Resurrection of Phenotypic Drug Discovery. ACS Med Chem Lett 11(10): 1820-1828. doi: 10.1021/acsmedchemlett.0c00006
2. Kringelum J, Kjaerulff SK, Brunak S, Lund O, Oprea TI et al. (2016) ChemProt-3.0: a global chemical biology diseases mapping. Database 2016:bav123. doi:10.1093/database/bav123
3. Bray MA, Singh S, Han H, Davis CT, Borgeson B et al. (2016) Cell Painting, a high-content image- based assay for morphological profiling using multiplexed fluorescent dyes. Nat Protoc 11:1757– 1774. doi: 10.1038/nprot.2016.10

PP16
Investigate the impact of PTMs on the protein backbone conformation

Joseph Rebehmed[a], Tarun Jairaj Narwani[b], Pierrick Craveur[b,c], Alexandre G.de Brevern[b]

a Department of Computer Science and Mathematics, Lebanese American University, 1h401 2010 Byblos, Lebanon.

b DSIMB- Bioinformatics team, Université de Paris, INSERM UMR_S 1134, DSIMB, Université de la Réunion, labex GR-Ex, 6 rue Alexandre Cabanel 75015 Paris, France.
c Synsight, Genopole Entreprises, 4 rue Pierre Fontaine, 91058 Evry, France.

Protein residues can undergo covalent chemical modifications, which are usually called posttranslational modifications (PTMs). They comprise different types, ranging from the simple addition of a group of atoms such as phosphorylation, to the binding of large groups, e.g., the retinal in the bacteriorhodopsin. PTMs play important roles in modulating various biological functions by altering the physical and chemical properties, the localization and activity of proteins. They are also implicated in multiple human diseases. PTMs impact on protein structures and their association to flexible and disorder regions have already been spotted in the past decade[1]. Nonetheless, the high diversity of PTM types and their multiple modification schemes (several PTM sites and types per protein that can be modified in different combinations and can change over time, etc.) make the direct confrontation of PTM annotations and protein structure data difficult[2]. Few years back, we have designed and developed a dedicated database to deal with this issue, namely PTM-SD (http://www.dsimb.inserm.fr/dsimb_tools/PTM-SD/).

Subsequently, two main PTM types (N-glycosylation and phosphorylation) have been considered in non-redundant datasets downloaded/collected from the our previously established database to analyze the impact of these modifications on the protein structures at the local and global levels using the structural alphabet protein blocks (PBs)3. Four different protein structures were selected to illustrate our findings. We observed that PTMs could either stabilize or destabilize the backbone structure, at a local and global scale, and that these effects depend on the PTM types4. Striking non-published recent results will also be presented.

1 Nussinov R, Tsai CJ, Xin F, Radivojac P. Allosteric post-translational modification codes, Trends Biochem. Sci., 2012, 37, 447 – 455.
2 Craveur P, Rebehmed J, de Brevern AG. PTM-SD: a database of structurally resolved and annotated post-translational modifications in proteins, Database, 2014, 2014, bau041.
3 de Brevern AG, Etchebest C, Hazout S. Bayesian probabilistic approach for predicting backbone structures in terms of protein blocks, Proteins, 2000 41, 271-287
4 Craveur P, Narwani T, Rebehmed J, de Brevern AG. Investigate the impact of PTMs on the protein backbone conformation, Amino Acids, 2019, 51,1065-1079.

PP17
Des3PI: a Computational Fragment-based Approach to Design Peptides Targeting Protein-Protein Interactions

Maxence Delaunay[a] and Tâp Ha Duong[a]

a Université Paris Saclay, 5 rue Jean-Baptiste Clément, 92290 Châtenay-Malabry

Protein-protein interactions (PPI) play crucial roles in many cellular processes and their deregulation often leads to cellular dysfunctions[1]. One promising way to modulate PPI is to use peptide derivatives that bind their protein target with high affinity and high specificity[2].

Peptide modulators are often designed using secondary structure mimics. However fragment-based design is an alternative emergent approach in the PPI field. Most of the reported computational fragment-based libraries targeting PPI are composed of small molecules or already approved drugs,[3,4], but according to our knowledge, no amino acid based library has been reported yet.

In this context, we developed a new fragment-based approach called Des3PI (Design of Peptides targeting Protein-Protein Interactions) with a library composed of single amino-acids. Our goal is find the optimal sequence of cyclic peptides that will bind a given protein surface with high affinity.

Each amino acid of the library is docked into the target surface using Autodock Vina. The resulting binding modes are geometrically clustered, and in each cluster, the most populated amino acids are determined and form the hot spots that will compose the optimal cyclic peptide.

This approach has been applied on three proteins: Mcl1, Ras, and Abeta. For each target, the five best peptides determined by Des3PI have been tested in silico. First, the peptides were blindly redocked on their target. Between two and four of the five peptides have binding modes close to the DES3PI predictions among the top 2 % of the all the redocked conformations. Secondly, the stability of the "good" redocked complexes has been verified using 200 ns MD simulations.

DES3PI shows encouraging results with at least one peptide for each protein target that succeeds in passing the two in silico validation steps.

1 Ryan, D. P. et al., Protein-protein interactions in human disease, current Opinion in Structural Biology, 2005 441–446.
2 Nevola, L. et al., Modulating protein–protein interactions: the potential of peptides, chemical Communications, 2005, 3302–3315.
3 Zhou, H. et al., S. Design of Bcl-2 and Bcl-xL Inhibitors with Subnanomolar Binding Affinities Based upon a New Scaffold., J. Med. Chem.2012, 4664–4682
4 Over, B. et al., Natural-product-derived fragments for fragment-based ligand disco-very, Nat. Chem., 2013, 21–28

PP18
Insights into the human 5-HT4 receptor binding site: exploration of activation conformational changes

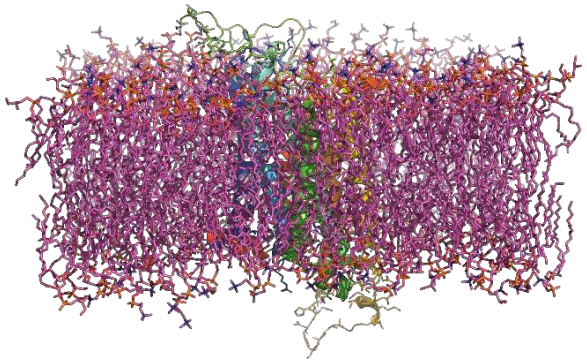
Jana Sopkova-de Oliveira Santos[a], Joffrey Garreaud[a], Valerie Collot[a], Christophe Rochais[a], Patrick Dallemagne[a],

a CERMN, EA 4258, Boulevard Becquerel 14032 Caen Cedex, France.

Are you feeling happy or hungry? Do you get migraines? All of these behaviors are controlled in part by the neurotransmitter serotonin. Serotonin is a small molecule synthesized from the tryptophan amino acid. Most of the serotonin in the body is found in the digestive system where it helps to control the motions needed for digestion, but the most important effects of serotonin are produced in the brain. One in a million neurons use it to communicate.

Serotonin is released from vesicles in nerve cells and bound by receptors on the target cell surface. There are 15 different forms of the serotonin receptors in our bodies and most of which are G protein-coupled receptors. Serotonin binds to the outside side of the receptor. This induces subtle changes in the receptor conformation and sends a signal to G proteins bound on receptor inside part. In some cases, this leads to an excitatory response in the cells, and in other cases it is inhibitory, all depending on the particular receptor and its individual cellular context.

In this study we are interested in the modeling of interaction between the serotonin receptor 5-HT4 and different ligands (partial agonist, antagonist …). The ligands[1,2] were docked into the activate and inactivate model of 5-HT4 receptor[3] and various representative poses were selected for each ligand. The built complexes were submitted to all atoms dynamics simulations (120 ns) and analyze with the aim of selecting a representative pose for each ligand (MMPBSA methods). The receptor conformational change was also inspected to highlight a correlation with the ligand nature.



1. Rochais C., Lecoutey C., Gaven F., Giannoni P., Hamidouche K., Hedou D., Dubost E., Genest D., Yahiaoui S., Freret T., Bouet V., Dauphin F., Sopkova-de Oliveira Santos J., Ballandonne C., Corvaisier S., Malzert-Fréon A., Legay R., Boulouard M., Claeysen S. Dallemagne P. Novel Multitarget-Directed Ligands (MTDLs) with Acetylcholinesterase (AChE) Inhibitory and Serotonergic Subtype 4 Receptor (5-HT4R) Agonist Activities As Potential Agents against Alzheimer's Disease: The Design of Donecopride. J. Med. Chem., 2015, 58 (7), 3172–3187.
2. Dubost E., Dumas N., Fossey C., Magnelli, R., Butt S., Ballandonne C., Caignard, D., Dulin F., Sopkova-de Oliveira Santos J., Millet P., Charnay Y., Rault S., Cailly, T., Fabis F. Synthesis and Structure-Affinity Relationships of Selective High-Affinity 5- HT4 Receptor Antagonists: Application to the Design of New Potential Single Photon Emission Computed Tomography (SPECT) Tracers. J. Med. Chem., 2012, 55(22), 9693-9707.
3. https://gpcrdb.org/

PP19
Simulating large extracellular matrix molecules as dynamic chains of rigid bodies

Camille Depenveiller[a], Hua Wong[a], Jean-Marc Crowet[a], Stéphanie Baud[a], Manuel Dauchez[a], Nicolas Belloy[a]

a UMR CNRS/URCA 7369 MEDyC / MAgICS, University of Reims Champagne-Ardenne, France

Extracellular matrix (ECM) is a tridimensional network composed of large biological macromolecules, divided into several families: glycosaminoglycans, proteoglycans, structure proteins and fibrous proteins. The ECM plays a key role in supporting tissues and organs. Many diseases are caused by dysfunction of ECM components, due to numerous changes in their structure and/or generation of fragments with negative effects. The precise organization of the ECM remains poorly known, because experimental data are rare and/or difficult to obtain. Indeed, the study of ECM proteins is challenging, as they are too small for being studied in detail by microscopy and also too large for classical methods of crystallography and NMR.

Classical molecular dynamics simulations are of interest to study domains of ECM proteins, but are still inappropriate for large multidomain macromolecules. Working at the mesoscopic scale allows to handle with this kind of systems. A dedicated numerical environment allowing to model and simulate large biological systems as dynamic chains of rigid bodies has been developed in the team[1]. Rigid bodies are modules that represent the domains composing each molecule and they allow to describe the representative conformations of molecules by simple shapes called primitives. The project aims to extend this approach in the case of one elastic protein of the ECM, elastin, and its monomer tropoelastin.

Thus, performing classical molecular dynamics on characteristic peptide motifs of tropoelastin, further simulations will then be adapted to rigid bodies, corresponding to the different domains composing the protein, in order to work at the mesoscopic scale. Each molecule will be built by an assembly of rigid bodies, linked together with molecular joints, on which constraints determined by characteristic angles between motifs will be applied. After classical molecular dynamics simulations of peptide motifs, the use of different clustering algorithms[2] and dihedral principal component analysis allows for sampling of the conformational space along molecular dynamics trajectories, in order to get the main characteristic conformations adopted by these proteins in the ECM. Properties of the proteins are also analyzed, particularly through the study of hydrophobicity[3] and electrostatics. These results enable to understand the behavior of the proteins and the structure-function-dynamics relationships. Elastin polypeptides, represented by their main conformations, as well as their properties, will be integrated in the ECM virtual model. These new insights will allow for a better understanding of the elasticity, the structural properties and thus the complexity of the ECM.

1 H. Wong, J. Prévoteau-Jonquet, S. Baud, M. Dauchez, and N. Belloy, “Mesoscopic Rigid Body Modelling of the Extracellular Matrix Self-Assembly”, J. Integr. Bioinforma., vol. 15, no 2, juin 2018, doi: 10.1515/jib-2018-0009.
2 T. Tubiana, J.-C. Carvaillo, Y. Boulard, and S. Bressanelli, “TTClust: A Versatile Molecular Simulation Trajectory Clustering Program with Graphical Summaries”, J. Chem. Inf. Model., vol. 58, no 11, p. 2178-2182, nov. 2018, doi: 10.1021/acs.jcim.8b00512.
3 R. Brasseur, “Differentiation of lipid-associating helices by use of three-dimensional molecular hydrophobicity potential calculations”, J. Biol. Chem., vol. 266, no 24, p. 16120-16127, August 1991.

PP20
Computational design of enzymes for cell-free biocatalysis

Delphine Dessaux[a], Sam Mallinson[b], Yannick Bomble[b], Sophie Barbe[a]

a TBI, Université de Toulouse, CNRS, INRAE, INSA, ANITI, Toulouse, France.
b Biosciences Center, National Renewable Energy Laboratory, Golden, CO, USA.

The use of enzymes for the sustainable and environmentally friendly production of biochemicals is continuously expanding and allows to bypass some of the drawbacks of the chemical production. However, these bioproduction processes currently rely mainly on microbial fermentation, whose effectiveness may be limited, mostly because of competitive metabolic pathways and/or the toxicity of the compounds produced. An appealing alternative is the use of in vitro multi-enzymatic pathways for cell-free bioproduction that may enable higher productivity and yields while reducing production costs. However, in order for this technology to generate the expected outcomes, it is essential to be able to endow enzymes with new functionalities and, in particular, new cofactor specificities. Indeed, most natural enzymes cofactors are expensive and labile and, even if they are reused, they can greatly limit the yield of the production.

Non-phosphorylating glyceraldehyde-3-phosphate dehydrogenase (GAPN) is a crucial enzyme involved in the regulation of ATP concentration in cell-free biocatalysis processes. However, GAPN is strictly dependent on the NADPH cofactor1, which prevents its use for NADH cofactor-dependent pathways or those based on the use of more stable and less expensive biomimetic cofactors.

In order to alter the cofactor specificity of GAPN, we developed and implemented a new rational enzyme engineering strategy based on computational protein design (CPD) methods[2-4]. These CPD approaches rely on artificial intelligence algorithms to predict mutant sequences for the design target while simultaneously considering several conformational states of the enzyme[5] representing different steps along the catalysis reaction. We combined these CPD methods with all-atom molecular dynamics simulations and binding free energy calculations, and ranked the GAPN mutants according to cofactor binding free energy and geometric features required for catalysis. This overall rational design strategy led to the prediction and the selection of a small number of mutants for experimental testing.

1 Boyd DA, Cvitkovitch DG, Hamilton IR. Sequence, expression, and function of the gene for the nonphosphorylating, NADP- dependent glyceraldehyde-3-phosphate dehydrogenase of Streptococcus mutans. J Bacteriol, 1995, 177(10):2622–2627.
2 Traoré, S, Allouche, D, André, I, de Givry, S, Katsirelos, G, Schiex, T, Barbe, S. A new framework for computational protein design through cost function network optimization. Bioinformatics, 2013, 2129–2136.
3 Traoré S, Allouche D, André I, Schiex T, Barbe S. Deterministic Search Methods for Computational Protein Design. In: Samish I. (eds) Computational Protein Design. Methods in Molecular Biology, 2017, vol 1529, Humana Press, New York, NY.
4 Vucinic J, Simoncini D, Ruffini M, Barbe S, Schiex T. Positive multistate protein design, Bioinformatics, 2020, 36(1):122-130.
5 Rahuel-Clermont S, Bchini R, Barbe S, Boutserint, S, André I, Talfournier F. Enzyme Active Site Loop Revealed as a Gatekeeper for Cofactor Flip by Targeted Molecular Dynamics Simulations and FRET-Based Kinetics. ACS Catal, 2019, 9, 1337–1346.

PP21
Molecular mechanisms governing oligomerization of Translocator Protein (TSPO) and their role in cholesterol translocation highlighted by coarse-grained approaches.

Julien Diharce[a,b], Rajas M. Rao[a,b], Frédéric Cadet[a,b,c], Catherine Etchebest[a,b].

a) Dynamique des Structures et Interactions des Macromolécules Biologiques, BIGR, UMR S_1134 Université de Paris – Inserm, Institut National de Transfusion Sanguine, 6 rue Alexandre Cabanel, 75015 Paris.
b) Laboratoire d'Excellence GR-Ex, 75015 Paris, France
c) Université de la Réunion, Biologie Intégrée du Globule Rouge, UMR_S1134, BIGR, Faculté des Sciences & Technologies Saint-Denis, F-97715 St. Denis, France

Translocator protein (TSPO), a mitochondrial membrane protein, has been extensively studied and its role is still debated and continues to be enigmatic. It is supposed to be involved in several biological processes, such as steroids biosynthesis or porphyrin transport but its primarily role is assumed to be the transport of cholesterol in the mitochondria outer membrane. Importantly, since the cholesterol is the precursor of neurosteroids, TSPO is considered as an interesting therapeutic target for Alzheimer's disease. From a structural perspective, several structures have been already solved. Two proteins from bacteria (R. Sphaeroides and B. Cereus) have been solved by means of X-ray crystallography. The mouse structure, the only solved mammalian form, has been resolved by NMR methodology. Moreover, the protein is highly dynamic in nature and has been proved to exist in several oligomeric states. However, the structure of these different oligomeric states remains elusive. In the present study, we addressed this important question and studied the dynamics of TSPO along the dimerization process. In this aim, we examined if and how mouse TSPO monomers could assemble to form dimer. Accordingly, we performed several coarse-grained molecular dynamics simulations considering two different initial configurations, one with pair of TSPO monomers distantly placed in a model of bilayer composed of DMPC/cholesterol mixture (same conditions as in the NMR experiment) and the other with preformed dimer models with different interfaces. We identified stable TSPO dimers with diverse interfaces, which confirms the versatility of the structure. Interestingly, some the symmetric interfaces were consistent with earlier experimental observations but stable asymmetric interfaces were also found. This last observation constitutes an interesting clue for larger oligomeric assemblies. Finally, we examined the dynamic behavior of the cholesterol nearby the protein. We found that the protein might contribute to cholesterol translocation through a hook mechanism. Ultimately, refinement of the stable interfaces was realized via all-atoms MD simulations that confirmed the high stability of the dimers formed. These calculations also emphasized the role of specific protein residues contributing to the cholesterol translocation and put into light some allosteric mechanisms. All these findings highlight the close relation between molecular properties and functional aspects, which are of major importance for understanding how ligands may impact TSPO function. This opens the way to design new ligands that can alter or improve cholesterol translocation by TSPO.

PP22
SenSaaS: Shape-based Alignment by Registration of Colored Point-based Surfaces

Dominique Douguet[a], Frédéric Payan[b], Lucas Grandmougin[b]

a Université Côte d'Azur, Inserm, CNRS, IPMC, 660 route des lucioles, 06560 Valbonne, France

b Université Côte d'Azur, CNRS, I3S, Les Algorithmes - Euclide B, 2000, route des lucioles 06900 Sophia Antipolis, France

3D point clouds or 3D meshes are data structures used in many fields well known by large audience (robotics, 3D reconstruction, games, autonomous navigation...) to model surfaces or volumes. Such 3D representations can also be relevant in chemistry to describe molecules although they were not the most used so far.
SENSAAS (SENsitive Surface As A Shape) is a shape-based alignment program using 3D point-based representation of the van der Waals surface[1]. SENSAAS is an original tool that combines recent methods dedicated to 3D registration[2], initially developed for the fusion of 3D point clouds, collected by devices such as depth cameras or LiDAR scanners.
Let's consider two molecules that we want to superimpose to search for similarities. SENSAAS gives a transformation matrix as output, leading to the 3D alignment of the Source molecule on the Target molecule which is kept fixed. The general idea is to align surface representations of the two molecules, to then compute a similarity score. SENSAAS follows four major steps: a) generation of a point cloud from the molecular surface of each input molecule; b) coarse alignment of the two point clouds thanks to a geometry-aware global registration3; c) labelling of each point of the two clouds according to user-defined classes; d) refinement of this alignment by applying a color and geometry-aware local registration[4]. This step results into a final transformation matrix (rotation + translation), that is applied to the Source molecule to get the final alignment. In parallel, fitness scores similar to Tversky coefficients are proposed to evaluate the embedding of a point cloud in another one.
To assess the efficiency of this approach, we tested its ability to reproduce the superimposition of X-ray structures of the benchmarking AstraZeneca (AZ) data set and, compared its results with those generated by the two shape-alignment approaches ShaEP[5] and SHAFTS[6]. These comparisons showed that SENSAAS provides accuracy performance equivalent to that of the reference methods, but also that it generates more accurate alignments in the first precision interval (RMSD $\leq$ 0.5 Å).
In addition, we showed that SENSAAS is able to also identify local similarities. SENSAAS possesses sub-matching properties which allow to align substructures or small fragments on a large molecule. Hence, SENSAAS provides a relevant contribution to shape-based alignment methods, especially in the field of lead optimization where scaffold hopping and bioisosteric replacement properties of a method are out of importance to identify promising compounds.

- Demo version at https://chemoinfo.ipmc.cnrs.fr/SENSAAS

- The code and documentation are available on GitHub https://github.com/SENSAAS/sensaas

1 Douguet D. and Payan F, SENSAAS: Shape-based Alignment by Registration of Colored Point-based Surfaces, Molecular Informatics, 2020, 8, 2000081. doi:10.1002/minf.202000081
2 Q.-Y. Zhou, J. Park, V. Koltun, arXiv 2018, 1801.09847.
3 Rusu, R. B.; Blodow, N.; Beetz, M., Fast Point Feature Histograms (FPFH) for 3D Registration. International Conference on Robotics and Automation (ICRA) 2009.
4 Park, J.; Zhou, Q.-Y.; Koltun, V., Colored Point Cloud Registration Revisited. IEEE International Conference on Computer Vision (ICCV) 2017, 143-152.
5 M. J. Vainio, J. S. Puranen, M. S. Johnson, J Chem Inf Model 2009, 49, 492-502.
6 X. Liu, H. Jiang, H. Li, J Chem Inf Model 2011, 51, 2372-2385.

PP23
A Molecular Bio-Assays Simulator to Unravel Predictors Hacking during Optimisation

Authors: Nicolas Drizard, Joseph-Andre Turk, Philippe Gendreau
nicolas.drizard@iktos.com
joseph-andre.turk@iktos.com
philippe.gendreau@iktos.com

Generative models are being increasingly used in drug discovery campaigns, very often coupled with ADME or bio-assays QSAR models to optimize a given set of properties. The molecules selected by these approaches are often revealed to be false positives, i.e. outside the true target product profile (TPP), as the predictors are being "hacked" during the optimisation. This observation can be seen as evidence of the GoodHart's law problem - where "when a measure becomes a target, it ceases to be a good measure"[1]. The impact is even larger in a multiobjective setting where the models need to extrapolate outside the training set distribution. To analyse and mitigate this issue, progress has been made in the evaluation of these model-based pipelines with the development of various kinds of "oracles"[2] — real-valued function used as a proxy for a molecular assay. Nonetheless, these oracles are often too easy to model in comparison to biological assays and are usually limited to single objective cases.

In this work, we introduce a simulator for multi-target bio-assays using a smartly initialized neural network (NN) which returns a continuous value for any input molecule. We use this oracle to replicate a real world prospective lead optimization scenario. First, we train predictive models on a small sample of molecules aimed at predicting their oracle values. Second, we generate new optimized molecules using the open-source GuacaMol[3] package. Finally, we select compounds which match the TPP according to the predicted target values. We observed that even when the predictive models have excellent estimated performance metrics, the final selection still contains many false positives.

We then evaluated the following different approaches to mitigate this issue using our NN-based oracle, and observed an increase of the enrichment factor of the final selection with respect to molecules truly in the TPP.
1/ Use of quality scores constraints on molecules during the generation;
2/ Rescoring and filtering in post-processing with predictor of the same target but not optimised during the generation;
3/ Selection of a diverse set of compounds to limit the false positives regions among the final selection of molecules;
4/ Use of ensemble predictive models;

1 Marilyn Strathern Improving ratings': audit in the British University system https://doi.org/10.1002/(sici)1234-981x(199707)5:3<305::aid-euro184>3.0.co;2-4
2 https://tdcommons.ai/functions/oracles/
3 Brown, N., Fiscato, M., Segler, M. H., and Vaucher, A. C. (Mar. 25, 2019). "GuacaMol: Benchmarking Models for de Novo Molecular Design". In: J Chem Inf Model 59.3, pp. 1096–1108.

PP24
Virtual Screening of Large Scale Libraries Guided with Docking and Deep Learning

Authors: Loïc Chadoutaud, Nicolas Drizard, Yasmine Nahal, Joseph-André Turk, Brice Hoffmann nicolas.drizard@iktos.com
loic.chadoutaud@iktos.com
yasmine.nahal@iktos.com
joseph-andre.turk@iktos.com

With the development of very large scale libraries of compounds, virtual screening campaigns can now explore a very diverse chemical space and help to identify more potent hits. Physics-based simulations - like docking - provide a reasonable proxy to rank compounds given a target.
Unfortunately, applying them at a large scale is not feasible. Hence, recent approaches[1] rely on predictive models and active learning to select most promising compounds without having to dock the entire library. Usually, those methods focus on a single objective function - e.g. the docking score - and do not evaluate the quality of the most active molecules obtained.

In this work, we first provide a thorough comparison of different frameworks and sampling methods to identify the most potent compounds in a given library. We evaluate different strategies to model the score to optimise either with a regression, with a classification using a fixed threshold or one using an adaptive threshold. We investigate methods that rely on greedy, uncertainty or diversity based sampling and balance between exploration and exploitation. We experiment with three datasets, two extracted from the ZINC database with documented docking-score against AMPC and D4 targets[2] and one from the LIT-PCBA dataset we docked internally. We provide an analysis of the efficiency of each approach looking at both the docking score and the diversity of the final selection and we observe that the diversity of the most potent compounds in the library strongly influences the performances of the different methods.

To improve the chemical quality of the final selection, we then extend these methods to multi-objective settings where we simultaneously optimise multiple docking scores and diversity,. We anticipate these methods to reveal molecules which would not only be promising hit compounds but also would have druggable properties.
Overall, we provide a better understanding of bayesian-optimization methods for virtual screening of large databases of candidate molecules for drug design

1. Francesco Gentile, Vibudh Agrawal, Michael Hsing, Anh-Tien Ton, Fuqiang Ban, Ulf Norinder, Martin E. Gleave, and Artem Cherkasov Deep Docking: A Deep Learning Platform for Augmentation of Structure Based Drug Discovery ACS Central Science 2020 6 (6), 939-949
2. Jiankun Lyu, Sheng Wang, Trent E. Balius, Isha Singh, Anat Levit, Yurii S. Moroz, Matthew J. O'Meara, Tao Che, Enkhjargal Algaa, Kateryna Tolmachova, Andrey A. Tolmachev, Brian K. Shoichet, Bryan L.Roth & John J. Irwin Ultra-large library docking for discovering new chemotypes Nature 566, 224–229 (2019)

PP25
Interface tangible modulaire, articulée et sans marqueur dédiée à la pédagogie et à la recherche en biologie moléculaire

Nicolas Férey[a], Bastien Vincke[b], Mohamed Anis Ghaoui[c]

a LISN—CNRS UMR , Univ. Paris-Saclay, Campus universitaire, bât 507, 91405 Orsay CEDEX, France
b,c SATIE—CNRS UMR 8029, Univ. Paris-Saclay, Digiteo Labs, BAT 660, 91405 Orsay CEDEX, France

La recherche moderne biologique moléculaire a toujours été associée à l'usage de modèles physiques, c'est à dire d'objets réels à une échelle adaptée pour que nous puissions les étudier et les manipuler. Ces objets facilitent notre compréhension des structure 3D et des mécanismes biomoléculaires. Avec l'ère de l'informatique, les modèles physiques ont peu à peu été supplantés par les modèles numériques, permettant de visualiser ces modèles en quelques minutes, là où les modèles physiques complexes étaient le résultats de travaux de tout une vie.

Si la modélisation in silico comporte évidemment de nombreux avantages et, en particulier la flexibilité en matière de contenus et de rendus visuels, la manipulation à travers des dispositifs d'interaction- humain-machine classiques restent peu aisées. La difficulté de manipulation et d'édition de modèles de plus en plus complexes, hétérogènes, avec des outils donnant accès de plus en plus fréquemment à leur caractéristiques dynamiques de manière interactive, reste un verrou important en interaction- humain-machine appliquée à la biologie moléculaire.

En parallèle au développement d'outils informatiques toujours plus puissants, la démocratisation des imprimantes 3D et les travaux sur les avantages de la manipulation d'objets tangibles ont redynamisé les recherches sur les modèles physiques et leurs usages, avec des propositions de modèles physiques toujours plus avancés. Le modèle Peppytide[1] est une illustration convaincante par son réalisme, permettant de construire de manière modulaire de petits peptides, de reproduire les structures secondaires usuelles et de ressentir les puits de potentiel des angles dihèdres, modèle idéal pour faciliter notre représentation mentale de la dimension spatiale et biomécanique caractérisant ces structures.

Dans ce contexte, inspirés par les travaux du pionnier dans le domaine des interfaces tangibles moléculaires Arthur J. Olson[2], associés à des recherches démontrant les avantages d'interfaces tangibles pour manipuler des objets numériques 3D articulées[3] ou des modèles physiques en facilitant l'apprentissage de la biologie moléculaire4, nous avons conçu une interface tangible modulaire, articulée, et surtout sans marqueur, élément qui restait rédibitoire pour un usage quotidien. La suppression de ces marqueurs grâce à la miniaturisation et aux technologies issues de l'internet des objets, a nécessité de mettre ensemble des compétences en biologie moléculaire, en informatique, et en électronique[5]. Cette interface a fait l'objet d'expérimentation pour mesurer sa pertinence face aux interfaces plus classiques, et nous entrons dans une nouvelle phase pour cerner les perspectives en de cette approche en matière pédagogique et de recherche, notamment pour supporter les approches de type Rational Drug Design.

1 Chakraborty, P.; Zuckermann, R.N. Coarse-grained, foldable, physical model of the polypeptide chain. Proc. Natl. Acad. Sci. USA 2013, 110, 13368–13373.
2 Gillet, A.; Sanner, M.; Stoffler, D.; Olson, A. Tangible interfaces for structural molecular biology. Structure 2005, 13, 483–491.
3 Jacobson, A.; Panozzo, D.; Glauser, O.; Pradalier, C.; Hilliges, O.; Sorkine-Hornung, O. Tangible and modular input device for character articulation. ACM Trans. Graph. (TOG) 2014, 33, 82.
4 Newman, D. L.; Stefkovich, M.; Clasen, C.; Franzen, M.A.; Wright; L. K. Physical models can provide superior learning opportunities beyond the benefits of active engagements, Biochem Mol Biol Educ. 2018,46(5): 435–444.
5 Vincke, B.; Ghaoui, M. A.; Férey, N.; Martinez, X. Modular and Articulated Interface for Interactive Molecular Manipulation, Sensors 2020, 20(18), 5415;

PP26
Evolution, structure and dynamics of IL-3 and IL-3R alpha interaction

Jade Fogha [a,b,c], Julien Diharce [a,b,c], Alexandre G. de Brevern [a,b,c,d]

a Université de Paris, Inserm, Biologie Intégrée du Globule Rouge, UMR_S 1134, DSIMB, F-75739 Paris, France.
b Institut National De La Transfusion Sanguine (INTS), F-75739 Paris, France.
c Laboratoire D'excellence Gr-ex, F-75739 Paris, France.
d Université De La Réunion, Inserm, Biologie Intégrée du Globule Rouge, UMR_S 1134, DSIMB, F- 97744 Saint-Denis, La Réunion, France.

Interleukin-3 (IL-3) is a cytokine belonging to the cytokine family of common β (βc) and is involved in the cell activation, proliferation, differentiation and survival in various biological systems. Its activity is mediated by the interaction with its receptor (IL-3R), a transmembrane glycoprotein member of the βc receptor family. The latter is a homodimer which has two distinct subunits: IL-3Rα and βc.[1] IL-3 and its receptor, especially IL-3Rα, play a crucial role in various pathologies such as hematological malignancies, sepsis, atopic/allergic and inflammatory diseases.[2] Thus, IL-3 and IL-3Rα are interesting therapeutic targets and we propose a thorough investigation of those proteins and their interaction based on structural, evolutionary analysis and molecular dynamics simulations. To achieve this goal, analysis of the different partners alone was carried out, followed by the protein-protein interaction assessment. Moreover, a comparative analysis was applied with other members of the βc family members. These results highlighted potential key residues for IL-3 and IL-3Rα interaction and could be an interesting stating point in drug discovery process.

1 S.E. Broughton, T.L. Nero, U. Dhagat, W.L. Kan, T.R. Hercus, D. Tvorogov, A.F. Lopez, M.W. Parker, The βc receptor family – Structural insights and their functional implications, Cytokine. 74 (2015) 247–258.
2 Testa, Pelosi, Castelli, CD123 as a Therapeutic Target in the Treatment of Hematological Malignancies, Cancers (Basel). 11 (2019) 1358.

PP27
Rational design of allosteric modulators of biomolecular motors

Katia Galentino[a] , Marco Cecchini[b]

a katia.galentino@etu.unistra.fr
b mcecchini@unistra.fr

Molecular motors are cytoplasmic proteins responsible of cellular movements through the conversion of chemical energy into mechanical work.
Among them, we find beta cardiac myosin (β-CM) that is of primary interest for its implication in heart-failure disease, hypertrophic cardiomyopathy and cardiomyopathies in general [1].

Here, we focus on a peculiar state of β-CM called pre-powerstroke (PPS), which is populated immediately after ATP hydrolysis [3].
In 2010 Morgan et al. reported on a series of selective activators of β-CM, whose potentiation of myosin's ATPase activity was given in terms of $AC40$ values [4]. This study culminated in the discovery of Omecamtiv Mecarbil (OMec), whose allosteric binding mode was later illuminated by X-ray christallography [3].
Here, we aim at a fundamental understanding of what makes a small molecule a myosin's activator and wonder whether its binding affinity (kd) or its rate dissociation constant ($koff$) would correlate with the biological activity.
For this purpose, we performed a docking analysis of 17 compounds from Morgan et al. [4] and evaluated their binding affinity by using the Molecular Mechanics Generalized-Born Surface Area method (MM-GBSA) [5].
The interdependence between the predicted affinities and the experimental activities ($AC40$) shows no correlation (R2 = 0.18).
Then, we evaluated the drug-target residence time of the same compounds using the efficient τ-Random Acceleration Molecular Dynamics (RAMD) simulations method [6].
Once again, we observed no correlation (R2 = 0.15) between the biological activity (AC40) and the drug rate dissociation constant.
Intriguingly, a correlation (R2 = 0.56) between calculated rate dissociation constants and calculated binding affinities was found.

What makes a small molecule a myosin's activator? The question remains open. To address it, we are currently setting up more rigorous free energy calculations, based on Free Energy Perturbation (FEP) [7] to explore the correlation with biological activity.

1- Malik FI, Hartman JJ, Elias KA, et al. Cardiac myosin activation: a potential therapeutic approach for systolic heart failure. Science. 2011;331(6023):1439-1443. doi:10.1126/science.1200113
2- Cecchini M, Alexeev Y, Karplus M. Pi release from myosin: a simulation analysis of possible pathways. Structure.2010;18(4):458-470. doi: 10.1016/j.str.2010.01.014
3- Planelles-Herrero VJ, Hartman JJ, Robert-Paganin J, Malik FI, Houdusse A. Mechanistic and structural basis for activation of cardiac myosin force production by omecamtiv mecarbil. Nat Commun. 2017;8(1):190. Published 2017 Aug 4. doi:10.1038/s41467-017-00176-5
4- Morgan BP, Muci A, Lu PP, Qian X, Tochimoto T, et al. Discovery of omecamtiv mecarbil the first, selective, small molecule activator of cardiac Myosin. ACS Med Chem Lett. 2010 Aug 20;1(9):472-7. doi: 10.1021/ml100138q.
5- Genheden S, Ryde U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. Expert Opin Drug Discov. 2015;10(5):449-461. doi:10.1517/17460441.2015.1032936
6- Kokh DB, Amaral M, Bomke J, et al. Estimation of Drug-Target Residence Times by τ-Random Acceleration Molecular Dynamics Simulations. J Chem Theory Comput. 2018;14(7):3859-3869. doi: 10.1021/acs.jctc.8b00230
7- Zoe Cournia, Bryce Allen, and Woody Sherman, Relative Binding Free Energy Calculations in Drug Discovery: Recent Advances and Practical Considerations, Journal of Chemical Information and Modeling 2017 57 (12), 2911-2937 DOI: 10.1021/acs.jcim.7b00564

PP28
New Tool for the Comparison of Molecular Structures Generations

Gantzer Philippe [a], Creton Benoit [a], Nieto-Draghi Carlos [a]

a IFP Energies nouvelles, 1 et 4 avenue de Bois-Préau, 92852 Rueil-Malmaison, France

The virtual generation of molecular structures allows to deeply explore the chemical space. This approach has gained a lot of interest the last decade and numerous algorithms are nowadays available to virtually generate new molecules.[1] The inversion of Quantitative Structure-Property Relationships models (i-QSPR) represents an attractive way to guide generation toward molecules with desired properties.[2] The selection of a relevant generation algorithm for a given application remains, however, challenging due to the initial dataset specificities, such as the number of available molecules, their chemistry, or their description. Indeed, although numerous metrics have been developed to compare QSPR performances, it is not obvious to compare i-QSPR approaches.

In this communication, we propose indices designed to compare relative performances of generation algorithms.[3] For this purpose, we assume a discretized representation of the chemical space considering the initial pool of molecular structures. New generated molecules are projected in that chemical space. Indices are then computed providing information about: (i) the coverage of the space, (ii) the coverage uniformity, (iii) the relative representativeness of the space by molecules generated by each algorithm ; (iv) and the closeness of generated molecule's property values to a target.

As an illustration, we are comparing two well-known generation approaches considering a database of flash point values (FP), i.e. by fragment assemblies (FA) and by evolutionary algorithms (EA), for the generation of molecules having their FP value within a desired interval. We observed a better ability for EA to explore the chemical space than for FA, according to the proposed indices. EA has a wider range of possible operations to create new molecules than FA which can only add fragments. When targeting a flash point interval, we showed that the choice of the number of molecules to be modified by EA at each iteration depends on the hypothetical number of feasible structures which can meet the property constrains.

1. Varnek, A.; Baskin, I. Machine Learning Methods for Property Prediction in Chemoinformatics: Quo Vadis? J. Chem. Inf. Model. 2012, 52, 1413–1437.
2. Gantzer, P.; Creton, B.; Nieto-Draghi, C. Inverse-QSPR for De Novo Design:A Review. Mol. Inf. 2020, 39, 1900087.
3. Gantzer, P.; Creton, B.; Nieto-Draghi, C. Generation of molecular structures based on fragments assemblies and genetic programming. Manuscript submitted, J. Chem. Inf. Model.

PP29
Visualization of the cluster-based partitioning of active and inactive pharmacophores. Application to the BCR-ABL pharmacophore network.

Damien Geslin[a,b], Jean-Luc Manguin[b], Vinh Vo Nhat [a], Alban Lepailleur[a], Bertrand Cuissart[b], Ronan Bureau[a]

a Centre d'Etudes et de Recherche sur le Médicament de Normandie, Normandie Univ, UNICAEN, CERMN, 14000 Caen, France
b Groupe de Recherche en Informatique, Image, Automatique et Instrumentation de Caen, Normandie Univ, UNICAEN, ENSICAEN, CNRS, GREYC, 14000 Caen, France (2) Groupe de Recherche en Informatique, Image, Automatique et Instrumentation de Caen, Normandie Univ, UNICAEN, ENSICAEN, CNRS, GREYC, 14000 Caen, France

A pharmacophore describes a combination of chemical features shared by several active molecules and responsible for favourable interactions with the active site of a target. We have recently described a new approach called Norns [1] for the automated detection of pharmacophores starting from a large dataset. With Norns, the main parameters for the extraction of the pharmacophores are their size (number of nodes for the pharmacophoric graphs, also known as order of the pharmacophores), their support (number of molecules associated to the pharmacophore) and the cut-off value for the growth-rate (GR, measure of their capability to discriminate between the two groups of molecules).

In this work, a dataset dealing with BCR-ABL tyrosine kinase (1492 compounds) was considered, like in the initial publication. With a biological cut-off at 1000 nM, the dataset was divided into two groups : active and inactive compounds (774 vs 718 compounds). The objective of this work is to analyse the potential to discriminate pharmacophores associated to active compounds towards pharmacophores associated to inactive ones. For this analysis, the clustering of the sets of pharmacophores was based on the graph edit distance[2].

Finally, the cluster-based partitioning of active and inactive pharmacophores was visualized on a BCR-ABL pharmacophore network.

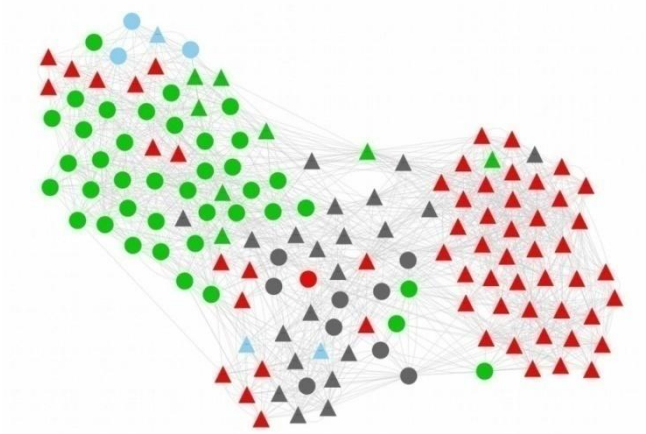


Figure 1 : Representation of a BCR-ABL pharmacophore network of order 7 (Triangle: active pharmacophore; Circle: inactive pharmacophore). The cluster-based partitioning of the pharmacophores is superimposed on the pharmacophore network (Red: active cluster; Green: inactive cluster; Grey: neutral clusters; Blue: outliers).

1 Métivier, J.-P.; Cuissart, B.; Bureau, R.; Lepailleur, A. The Pharmacophore Network: A Computational Method for Exploring Structure–Activity Relationships from a Large Chemical Data Set. J. Med. Chem. 2018, 61 (8), 3551–3564. https://doi.org/10.1021/acs.jmedchem.7b01890.
2 Riesen, K. Graph Edit Distance. In Structural Pattern Recognition with Graph Edit Distance: Approximation Algorithms and Applications; Springer International Publishing: Cham, 2015; pp 29–44. https://doi.org/10.1007/978-3-319-27252-8_2.

PP30
MCSS-Based Predictions of Binding Mode and Selectivity of Nucleotide Ligands

Roy González-Alemán[a, b], Taher Yacoub[b], Nicolas Chevrollier[b], Manuel Simoes[c], Luis Montero-
Cabrera[a], and Fabrice Leclerc[b]

a Laboratorio de Química Computacional y Teórica (LQCT), Facultad de Química, Universidad de La Habana, La Habana,10400, Cuba
b Institute for Integrative Biology of the Cell (I2BC), CEA, CNRS, Université Paris Saclay, Gif-sur- Yvette, F-91198, France.
c CPC Manufacturing Analytics, 67000 Strasbourg, France

Computational fragment-based approaches are widely used in drug design and discovery. One of their limitations is the lack of performance of docking methods, mainly the scoring functions. MCSS (multiple-copy simultaneous search) performs local and iterative docking calculations based on an efficient sampling method implemented in the CHARMM program.

With the emergence of fragment-based approaches for single-stranded RNA ligands, we analyze the performance in the docking power (success rate for identifying native poses) and the screening power (success rate for identifying the true native nucleotide) of an MCSS-based approach. The performance is evaluated on a benchmark of 121 protein−nucleotide complexes where the four RNA residues are used as fragments. Hybrid solvent models based on some partial explicit representations improve both the docking and screening powers. A detailed analysis of molecular features suggests various ways to optimize the performance further.

These results support the application of MCSS to the in silico fragment-based design of oligonucleotides targeting the catalytic site of proteins. We present a case study of the de novo design of pseudoaptamers inhibitors against BACE-1, a crucial protein involved in Alzheimer's disease.

PP31
Identification of biological targets for argon

Islem HAMMAMI [a,b], Geraldine FARJOT [b], Mikael NAVEAU [c], Thierry PRANGÉ [d], Ira KATZ [b], Nathalie COLLOC'H [a]

a ISTCT UMR 6030 CNRS Université de Caen -Normandie CEA, équipe CERVOxy, Centre Cyceron, 14032 Caen, France.

b Medical R &D, Healthcare World Business Line , Air Liquide Santé International , Paris Innovation Campus, 78354 Jouy-en-Josas, France.

c UMS 3408, CNRS Université de Caen-Normandie, Centre Cyceron, 14032 Caen, France.
d CiTCoM UMR 8038 CNRS Université Paris Descartes, 75006 Paris, France.

Argon (Ar) belongs to the group of chemically inert noble gases, which display a remarkable spectrum of clinically useful biological properties. Noble gases by binding to physiologically relevant sites in proteins involved, for example in cell signaling would induce functional modifications that could lead to interesting protective properties. However, the literature on this is relatively small and the fascinating biology of these atoms is still being uncovered.
In an attempt to better understand noble gases, notably argon's mechanism of action, we mined a massive molecular modelling database[1] which lists all possible noble gas binding sites in the proteins listed in the Protein Data Bank. We developed a new method of analysis to identify amongst all predicted noble gas binding sites, the potentially relevant ones within protein families, which are likely to be modulated by Ar. Our method consists mainly in determining within structurally aligned proteins, the conserved binding sites whose shape, localization, hydrophobicity and binding energies are further examined. This method was applied to the analysis of two protein families where crystallographic noble gas binding sites are known.

Our findings indicate that amongst the most conserved binding sites, either the most hydrophobic one or the site which has the best binding energy correspond to the crystallographic noble gas binding sites with the best occupancies, therefore the best affinity for the gas.
This method will allow us to predict relevant noble gas (Ar) binding sites that would be endowed with potential pharmacological interest and thus Ar targets that will be prioritized for further studies including in vitro validation.

1. Winkler DA, Katz I, Farjot G, Warden AC, Thornton AW. Decoding the Rich Biological Properties of Noble Gases: How Well Can We Predict Noble Gas Binding to Diverse Proteins? ChemMedChem. 2018;13(18):1931-1938. doi:10.1002/cmdc.201800434

PP32
ATOLL: A visualization tool to compare transmembrane domains structures

Authors: Célien Jacquemard[a], Esther Kellenberger[a]

a Faculté de Pharmacie, 74 Route du Rhin, 67400 Illkirch-Graffenstaden, France – Université de Strasbourg

The 3D structure of a transmembrane protein can be represented in a 2D frame in various ways, with more or less structural information. The top-viewed helix box diagram, as in GPCRdb[1], simply shows which helices are in contact and the putative contact residues. The screenshot of a crystal structure or a modelled conformation can provide atomic details, yet it can also be difficult to interpret. If the protein is described by multiple conformations (e.g. molecular dynamics), the representation has to be simplified. Usually dimensionality reduction is applied. For example, a distance-based heatmap that shows distance pair frequencies, or, the root mean square deviation of atomic coordinates over time. In these cases, global view of the receptor is lost.

Here we propose ATOLL (Aligned Transmembrane dOmains Layout fLattening), a tool to visualize the multiple layouts of the transmembrane domains of a protein. It is based on the 3D-aligment of the domains followed by a projection of atomic coordinates onto the plane of the membrane.
We describe here two possible applications of ATOLL: the characterization of ADRB2 deactivation as simulated by molecular dynamics[2] and the comparison of 265 PDB structures describing 53 GPCRs in active, intermediate or inactive states[3]. The pictures well show that the difference of structural characteristics between active and inactive states, within the same receptor or shared by the class A GPCRs.

ATOLL is applicable not only to GPCRs but to any transmembrane proteins. It produces self-explanatory, user-customizable and high quality plots (github.com/LIT-CCM-lab/ATOLL).
References

1 Munk, C. et al. (2019) An online resource for GPCR structure determination and analysis. Nat. Methods 16:151– 162
2 Dror, R. et al. (2011) Activation Mechanism of the B2-Adrenergic Receptor. Proc. Natl. Acad. Sci. U. S. A. 108(46):18684– 18689
3 Koensgen, F. et al. (2019) Unsupervised Classification of G-Protein Coupled Receptors and Their Conformational States Using IChem Intramolecular Interaction Patterns J. Chem. Inf. Model. 59(9):3611-3618

PP33
Structural determination of PTPN4 using SAXS restrained Molecular Dynamics

Yasaman Karami[1], Nicolas Wolff[2] and Michael Nilges[1]

1 Department of Structural Biology and Chemistry, Structural Bioinformatics Unit, CNRS UMR 3528, Institut Pasteur, Paris, France.

2 Department of Neuroscience, Channel-Receptors Unit, CNRS UMR 3571, Institut Pasteur, Paris, France.

PTPN4 belongs to the non-receptor protein tyrosine phosphatase (PTP) family and consists of a N-terminal FERM domain, a PDZ (PSD-95/Dlg/ZO-1) domain and a C-terminal catalytic tyrosine phosphatase (PTP) domain. PTPN4 is involved in various biological activities such as T-cell signaling, learning, spatial memory and cerebellar synaptic plasticity. The cleavage of the phosphatase in the cell leads to the enzyme activation and the active form contains the PDZ and PTP domains that are connected by a linker. But, due to the disorder nature of this linker, the joined structure of PTPN4, is not yet determined. It has been shown that the PDZ domain inhibits the catalytic activity of the PTP domain, while the binding of a ligand to PDZ releases the auto-inhibition and activates the phosphatase [1]. We recently proposed a Bayesian model for automatic weighting of SAXS data combined with Molecular Dynamics (MD) simulations, in order to find optimal structural ensemble for PTPN4 [2]. However, the detailed mechanism of PDZ domain for modulating the phosphatase are not yet fully understood. In addition, the experimental data from NMR and SAXS, suggests a compact conformation for the PDZ-PTP domain, as well as a dis- order nature for the linker. To better understand the role of PDZ domain on the catalytic activity of the phosphatase, MD simulations can be carried out. However, such simulations are highly dependent on the accuracy of force fields. One possible approach to tackle such problem, is to integrate experimental data into MD simulations. Metainference is a Bayesian inference approach that integrates experimental data with prior distribution of models, while considering the effects of conformational averaging and errors. On the other hand, it has been shown that metainference can be combined with metadynamics to accelerate the exploration of conformational space [3]. In this study we determined the structural ensembles of the PDZ-PTP tandem of the protein PTPN4, in the presence and absence of the ligand, using extensive metainference metadynamics hybrid-resolution SAXS-driven simulations, following a similar recent study [3]. For each system, we performed multi-replica simulations for 500 ns, leading to 100 µs of simulation time. This study allowed us to elucidate the functional dynamics of the PTPN4 and better understand the molecular mechanisms that control the catalytic activity of phosphatase.

[1] Shin-ichiro Kina, Tohru Tezuka, Shinji Kusakawa, Yasushi Kishimoto, Sho Kakizawa, Koichi Hashimoto, Miho Ohsugi, Yuji Kiyama, Reiko Horai, Katsuko Sudo, et al. Involvement of protein-tyrosine phos- phatase ptpmeg in motor learning and cerebellar long-term depression. European Journal of Neuro- science, 26(8):2269–2278, 2007.

[2] Yannick Spill, Yasaman Karami, Pierre Maisonneuve, Nicolas Wolff, and Michael Nilges. Automatic bayesian weighting for saxs data. Frontiers in Molecular Biosciences, 8:495, 2021.

[3] Cristina Paissoni, Alexander Jussupow, and Carlo Camilloni. Determination of protein structural en- sembles by hybrid-resolution saxs restrained molecular dynamics. Journal of Chemical Theory and Computation, 16(4):2825–2834, 2020.

PP34
Navigating the chemical space using structure-based guided generative modelling

Stéphanie Labouille*[a], Nicolas Devaux[a] Brice Hoffmann[a]

a Iktos, 65 rue de Prony 75017, Paris, France
* E-mail: stephanie.labouille@iktos.com

The past decade has witnessed a growing interest in the application of generative modelling to drug discovery problems.[1] In this approach, a deep generative model is trained on a dataset of molecules (e.g. CheMBL) to generate novel molecules, enabling one to span an infinitely larger chemical space than accessible through virtual screening of known databases. Taking advantage of the recent advances in the field of language processing, algorithms such as long short-term memory (LSTM) neural networks have been used, with great success, to generate large libraries of de novo drug-like molecules.[2] Such algorithms typically use the SMILES representation (Simplified Molecular Input Line Entry System), which encodes a molecule as a string of ASCII characters. They generate new SMILES character by character, having learnt the probability of a new character given all the previous characters in the string.

When combined with predictive models via reinforcement learning,[3] generative models can be biased toward the generation of molecules with the desired molecular profile (e.g. biological activity, physico-chemical and pharmacokinetic properties). In principle, this strategy can help navigate the chemical space more efficiently and more effectively than brute-force approaches, by focusing on promising regions rather than sampling at random. In practice, the exploration is often limited by the applicability domain of the predictive models used to guide the generation, therefore restricting access to truly novel chemical space. Physics-based methods, such as molecular docking, have by construction a much wider applicability domain and thus offer a way to overcome this limitation.[4]

Here, we investigate the behaviour of three popular optimisation algorithms (Hill-Climbing, Augmented Likelihood and Proximal Policy Optimization) when coupled with structure-based scoring functions. Focusing on a well-known case study (COX-2), we compare their ability to balance exploration and exploitation during the generation. We show that without any knowledge of what constitutes a good ligand for the target, these algorithms can guide the generation toward known scaffolds and even known binders, while also suggesting new scaffolds with improved docking scores. Finally, we show that incorporating structural knowledge about the target during the generation leads to de novo molecules that satisfy key protein-ligand interactions, making it a powerful tool for drug design.

1 J. Meyers, B. Fabian, N. Brown, De novo molecular design and generative models, Drug Discovery Today 2021, https://doi.org/10.1016/j.drudis.2021.05.019.
2 M. H. S Segler, T. Kogej, C. Tyrchan, M. P. Waller, Generating focused molecule libraries for drug discovery with recurrent neural networks ACS Cent. Sci. 2018, 4 (1), 120–131.
3 M. Olivecrona, T. Blaschke, O. Engkvist, H. Chen, Molecular de-novo design through deep reinforcement learning J. Cheminform. 2017 9, 48.
4 M. Thomas , R. T. Smith , N. M. O'Boyle , C. de Graaf, A. Bender, Comparison of structure- and ligand-based scoring functions for deep generative models: a GPCR case study, J. Cheminform. 2021, 13, 39.

PP35
Towards lattice-based interactive pharmacophore exploration

Etienne Lehembre[a], Jean-Luc Lamotte[b], Abdelkader Ouali[a], Christophe Couronne[a], Bruno Crémillleux[a], Alban Lepailleur[b], Jean-Luc Manguin[a], Ronan Bureau[b], Albrecht Zimmermann[a], Bertrand Cuissart[a]

a GREYC, Normandie Univ., UNICAEN, CNRS – UMR 6072, 14000 Caen, France
b Centre d'Etudes et de Recherche sur le Médicament de Normandie, Normandie Univ, UNICAEN, CERMN, 14000 Caen, France

In medicinal chemistry, a pharmacophore denominates a spatial arrangement of chemical features shared by a group of molecules which is responsible for a favourable interaction with binding site of the studied 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]. The connections between the computed pharmacophores provide a hierarchical organization: the pharmacophore network. The latter is structured by pharmacophores' size and contains a large number of them. The current work aims at developing a structure which a drug designer can use to support his analysis without having to repeatedly mine pharmacophores. For this purpose, we enhance the knowledge on the pharmacophore network organization by taking into account parent-children relations and/or grouping the pharmacophores into equivalence classes, i.e. sets of pharmacophores occurring in exactly the same molecules. Additionally, each pharmacophore is annotated with one or several quality measures (e.g. confidence or growth rate measurements), information that will be exploited later-on.

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

With a view towards Interactive Pattern Mining [2], we also use the lattice to simulate the effects of users or functions interacting with the search space during the search itself. Each non-explored branch in the lattice represents a pharmacophore and its descendants which would not have been explored during the interactive mining, allowing users to understand the repercussions of their interaction. This can be considered a training step for users, who do, after all, make up an indispensable ingredient of interactive pattern mining.

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

PP36
Comparison between homology-based and experimentally determined structures on prostanoid GPCR receptor

Camille Lescanne[a], *, Maud Jusot[b], *, Anita Rayar[a], Brice Hoffmann[b], Jean-Philippe Annereau[a], Luis Briseño-Roa[a]

a. Medetia, Institute Imagine, 24 Boulevard du Montparnasse, 75015 Paris, France
b. IKTOS, 65 rue de Prony, 75017 Paris, France email: contact@medetia.com
* equal contribution

Background
G-protein-coupled receptors (GPCR) are arguably one of the most intensively studied drug target families – one third of all small molecule approved are GPCR ligands. From the approximately 800 GPCRS encoded in the human genome, only about 70 structures have been determined in complex with either an agonist or antagonist. This lack of structural knowledge – generally due to the challenges of experimental determination structure of membrane protein – is further accrued by the fact that GPCR are often dynamic. With the advent of novel computational methods on the drug discovery process, the absence of experimentally determined structures or reliable computationally generated models is not anymore a rate limiting point in the structure-based drug discovery process.

Objectives
The specific objective of the present study is to benchmark computationally generated homology models of EP2 receptor created from the GPCR prostanoid receptor family members, against structures recently published and solved by EM-cryo. Our final intent is to use our models to identify lead-like molecules amenable to be developed in a particular set of renal rare diseases with currently no available treatment.

Method
A large number of homology models of EP2 prostaglandin receptor were created using previously reported structures of prostaglandin type-E receptors EP3 and EP4 as templates. As model evaluation is a challenging task, different steps for model selection were defined from the use of simple modelling scores to a more complex comparison between docking and experimental values on a proprietary database containing more than 600 prostanoid receptor ligands.

Results
Approximately 1000 models were created, from which around 30 with the best energy and structural descriptors were retained. A final selection <10 models was done by re-docking the co-crystalized ligand and a proprietary database. From the docked compounds, we focused the analysis on Taprenepag and Evatanepag, thanks to the recent publication of EP2 structure resolved by EM-cryo modelling at 2,8Å with these two compounds. The comparison of the EP2 models and the experimental structures as well as the docking poses found for both Evatanepag and Taprenepag in the models compare to the experimentally determined structures, showed the good quality of our models.

Conclusions
Our work is allowing us to parse through several proprietary focused and public libraries of prostanoid ligands. We are confident that our work will enable us to bring forward molecules that are relevant to our pathologies and specific to our receptors of interest. Our final aim is to discover a therapeutic molecule to treat a family of currently untreated diseases.

PP37
Protein-Ligand binding affinity prediction using combined molecular dynamics and deep learning approaches

Libouban, PY. [1], Aci-Sèche, S. [1], Tresadern, G. [2], Bonnet, P. [1]

1 Institute of Organic and Analytical Chemistry (ICOA), UMR7311, Université d'Orléans, CNRS, Pôle de chimie rue de chartres - 45067 Orléans Cedex 2, France
2 Computational Chemistry, Janssen Research & Development, Janssen Pharmaceutica N. V., B-2340 Beerse, Belgium

Interactions of small molecules with proteins are essential to pharmaceutical research. Indeed, drugs bind to the active site of proteins, in order to prevent or modulate their interaction with their natural ligands. The stronger a molecule binds to its target, the lower dose is required for its effect, increasing the chance to become a drug even if other important parameters need to be evaluated.

Although in vitro experiments were developed to measure the affinity of protein-ligand complexes, they remain long and expensive. Nowadays artificial intelligence methods, and especially deep learning algorithms like convolutional neural networks, are used to develop statistical models that predict the affinity of these complexes (1, 2). For example, the neural networks of Pafnucy or Kdeep use the 3D structures of protein-ligand complexes to predict the binding affinities of compounds. They are trained on the PDBbind dataset, which gather the structural information of more than 17000 protein-ligand complexes, as well as their affinity. Regarding their performance, these 2 models achieve a correlation coefficient of 0.78 and 0.82 respectively on the 2016 PDBbind core set (comparative assessment of scoring functions (CASF)), but drawbacks of possible bias related to only learning from the ligand structures have been suggested (3). One of the main limitations to improve the statistical models is the lack of structural data, since it requires extensive experimental determination of complexes.

This project aims to improve our ability to predict the binding affinity of protein-ligand complexes using recent deep learning methods. To achieve this goal, we develop a protocol combining deep learning and molecular dynamics (MD) simulations. Additional structures extracted from MD simulations will be used as input for the models. In addition to acting as a data augmentation tool, MD simulations will add temporal information. The information of the movement of the ligands during the simulation was already evaluated to accurately determine correct docking poses (4).

We will carry out 10 MD simulations of 10 ns for each complex using all the complexes included in the PDBbind, leading to more than 170000 simulations. In order to create efficient statistical models, we will use a neural network able to analyse both spatial and temporal information, like the convolution LSTM (5). The first results show a tendency between the movement of the ligands in the binding site and its affinity to the protein. We hope that the developed models will better predict the binding affinity of protein-ligand complexes.

1. Stepniewska-Dziubinska MM, Zielenkiewicz P, Siedlecki P. Development and evaluation of a deep learning model for protein-ligand binding affinity prediction. Bioinformatics (Oxford, England). 2018;34(21):3666-74.
2. Jiménez J, Škalič M, Martínez-Rosell G, De Fabritiis G. KDEEP: Protein–Ligand Absolute Binding Affinity Prediction via 3D-Convolutional Neural Networks. Journal of chemical information and modeling. 2018;58(2):287-96.
3. Sieg J, Flachsenberg F, Rarey M. In Need of Bias Control: Evaluating Chemical Data for Machine Learning in Structure-Based Virtual Screening. Journal of chemical information and modeling. 2019;59(3):947-61.
4. Liu K, Kokubo H. Exploring the Stability of Ligand Binding Modes to Proteins by Molecular Dynamics Simulations: A Cross-docking Study. Journal of chemical information and modeling. 2017;57(10):2514-22.
5. Shi X, Chen Z, Wang H, Yeung D-Y, Wong W-k, Woo W-c. Convolutional LSTM Network: a machine learning approach for precipitation nowcasting. Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 1; Montreal, Canada: MIT Press; 2015. p. 802–10.

PP38
VTX: High-performance molecular structure and dynamics visualization software

Maxime Maria[1,2], Simon Guionniere[1], Nicolas Dacquay[1], Yassine Naimi[3], Jean-Philip Piquemal[4], Guillaume Levieux[5], and Matthieu Montes[1]

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

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

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

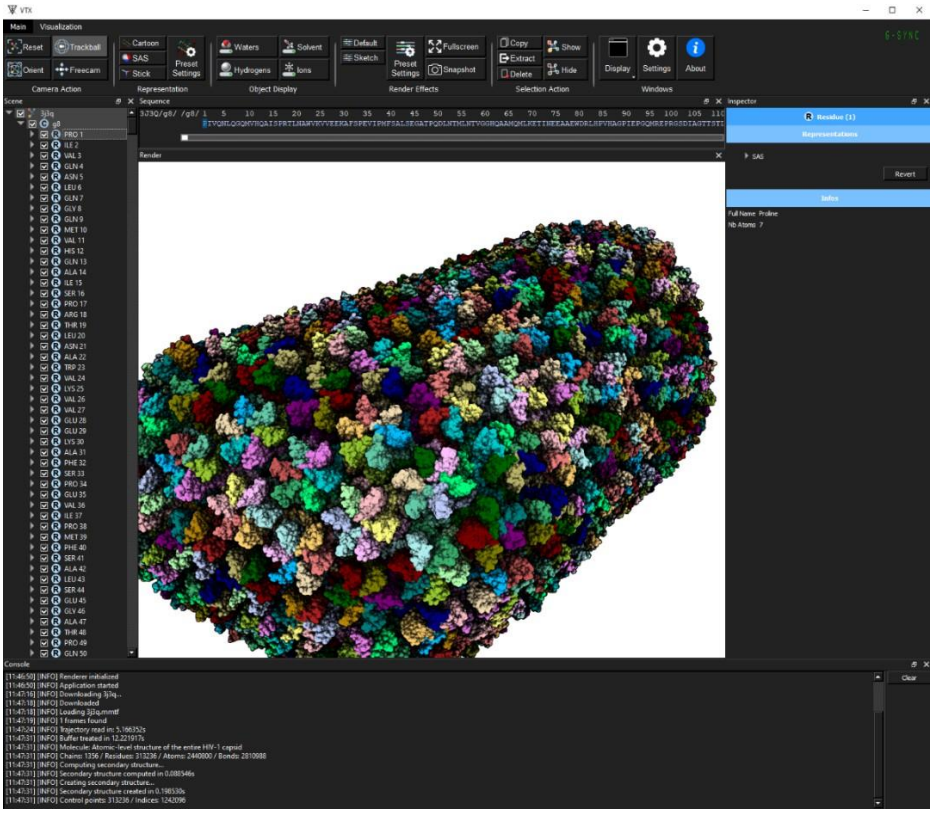


Figure 1. Illustration of VTX Graphical user interface on the structure of the human HIV Capsid (2.5 million atoms, PDB ID: 3J3Q).

PP39
An homodimeric multi-domain protein interacting with ssDNA and dsDNA : The challenges of RelSt3 relaxase modelisation

Dominique Mias-Lucquin[a], Bernard Maigret[a], Isaure Chauvot de beauchene[a]

A Université de Lorraine, CNRS, Inria, LORIA, F-54000 Nancy, France

Spread of antibiotic resistance genes by conjugative mechanisms represent a major public health issue. Among the elements involved in these horizontal gene transfers, integrative and conjugative elements (ICEs) are the most common. In Streptococcus, RelSt3 relaxase is a protein that binds to a sequence-specific dsDNA, and to the ssDNA region of a hairpin to cut it prior to its transfer. RelSt3 is active in the form of an homodimer, each monomer formed by 3 domains linked by highly flexible long loops

Here, I want to expose the challenges encountered during the building of a model of this protein and of its interactions with ss- and dsDNA :

- Modeling of the protein domains and dimeric core domain; Of the three domains of RelSt3, two could be built by homology modeling. However the third, while consensually described as primarily organized in alpha helices, resists our modeling attempts.

- Linking the domains; The three domains of RelSt3 are linked to each other through long (10 and 13 aa) flexible linkers, and a precise location of these domains relative to each other is necessary to bind DNA on known sites.

- Localizing the HtH domains relative to each other on dsDNA; It was experimentally shown that HtH domains interact with two short (13 nt) symmetric repeats, separated by a 9 nt spacer. Due to the symmetric nature of the protein, it is suggested that each HtH domain interacts with a different repeat. This carries out additional constraints to take into account for their positioning on the DNA, relative to each other and relative to the central domain.

- Modeling the ssDNA conformation in contact with the protein; The precise interaction sites are known for both partners. But the fragment-based approach considered here, in which ssDNA sections are fragmented into 3-mers for an efficient docking, was developed for ssRNA/protein docking. It is limited here by the low number of ssDNA/protein complexes in the PDB used to build a fragment library, compared to dsDNA/protein complexes.

This poster aims to present our combination of approaches and to trigger discussion on the hindrances we still face.

PP40
in silico investigation of dynamical and structural properties of the major efflux pump of P. aeruginosa : MexBA-OprM.

MOLZA Anne-Elisabeth [a], Lalaimia Ikram[a], Moroldo Clement[a], Gardebien Fabrice[a], Etchebest Catherine[b],

a INSERM UMR_S 1134,DSIMB, DE-ASSEMBLY team, Faculté des Sciences et Technologies, Université de La Réunion, 15 Avenue René Cassin, BP 7151, 97715 Saint Denis Messag Cedex 09, La Réunion, France
b INSERM UMR_S 1134,DSIMB, Université de Paris 7, 6, rue Alexandre Cabanel 75739 Paris Cedex 15, France

Emerging multi-resistant resistance (MDR) bacterial strains are now considered as a global public health problem. Consequently, new strategies are required to fight against the MDR mechanisms. Our project is part of this major medical challenge and aims to counteract these mechanisms. Here, we focus on Pseudomonas aeruginosa (PA), a nosocomial pathogen — responsible for 10% of infections contracted in hospitals — causes multiple infections: urinary, respiratory and gastrointestinal.
In this project, we are particularly interested in the PA tripartite multidrug efflux system **MexBA-OprM** [3], because it is a **major actor of antibiotic resistance**. Actually, MexAB-OprM efflux pump recognize a large variety of antimicrobial compounds, and lead to the reduction of drug accumulation in bacterial cell by exporting these drugs to cell outside, contributing to MDR in P. aeruginosa accordingly.
A key efficiency mechanism lies in the bacterial capacity of activating these three component complexes. In order to counteract this mechanism, our strategy consists in finding molecules able to limit the pump assembly. Thus, the targeted sites are those located in the protein-protein interfaces. Moreover, as assembly may require conformational changes, residues involved in motions essential for pump activation are also considered.
In this aim, we used a combination of molecular modeling, normal mode analyses, coarse-grained molecular dynamics, structure or ligand-based virtual screening and molecular docking methods.
Our first results of virtual screening based on a dataset of protein-protein inhibitors provided a few sets of molecules interacting with interfacial regions of MexA protein. Moreover, we performed in-depth normal mode analyses that have highlighted important dynamics features. In particular, we have identified regions that are involved in conformational changes which themselves could be necessary for protein function. These residues are now considered as targets in a new stage of virtual screening.

Preliminary results are encouraging to achieve our goals: i) provide a complete description of the molecular mechanisms of the multidrug efflux system; ii) identify and rational design of new inhibitors capable of preventing the function of the efflux pump using MD simulations, drug design, virtual screening methods; iii) to develop alternative and rational therapeutic strategies based on the use of molecules capable of **inhibiting the formation of complexes involved in other types of infectious pathologies**.

Li XZ, Nikaido H, Poole K. Role of mexA-mexB-oprM in antibiotic efflux in Pseudomonas aeruginosa. Antimicrob Agents Chemother. 1995;39(9):1948-1953. doi:10.1128/aac.39.9.1948
Pang Z, Raudonis R, Glick BR, Lin TJ, Cheng Z. Antibiotic resistance in Pseudomonas aeruginosa: mechanisms and alternative therapeutic strategies. Biotechnol Adv. 2019 Jan-Feb;37(1):177-192. doi: 10.1016/j.biotechadv.2018.11.013. Epub 2018 Nov 27. PMID: 30500353.
Glavier M, Puvanendran D, Salvador D, Decossas M, Phan G, Garnier C, Frezza E, Cece Q, Schoehn G, Picard M, Taveau JC, Daury L, Broutin I, Lambert O. Antibiotic export by MexB multidrug efflux transporter is allosterically controlled by a MexA-OprM chaperone-like complex. Nat Commun. 2020 Oct 2;11(1):4948. doi: 10.1038/s41467-020-18770-5. PMID: 33009415; PMCID: PMC7532149.

PP41
Structural analysis of interaction between SARS-CoV-2 spike protein and the human ACE2 receptor

S.Naceri [1]*, M.Ghoula[1]*, S.Sitruk[1], G.Moroy[1], D.Flatters[1], A-C.Camproux[1]

1. Université de Paris, BFA, UMR 8251, CNRS, ERL U1133, Inserm, F-75013 Paris, France

Corresponding Author: sarah.naceri@etu.u-paris.fr

Paper Reference: Molecular Dynamics Simulations of Influenza A Virus NS1 Reveal a Remarkably Stable RNA-Binding Domain Harboring Promising Druggable Pockets. Abi Hussein H, Geneix C, Cauvin C, Marc D, Flatters D, Camproux AC. Viruses. 2020 May 14;12(5):537. doi: 10.3390/v12050537.

In 2019, the emergence of the highly pathogenic SARS-CoV-2 coronavirus, which spread rapidly in 2020, led to an intensive search on this virus for a rapid development of vaccines and became a global public health priority. At the same time, the search for drug candidates to inhibit the virus mechanism and reduce the overall infection has also become a priority. The Spike protein is a target protein of interest because it allows the virus to enter human cells by interacting with human ACE2 (Angiotensin-converting enzyme 2). Several three-dimensional structures of the Spike protein, are currently available. These structures were mainly resolved by cryo-electron microscopy, but also to a lesser extent by X-ray crystallography. Thus, at the end of March 2020, 10 structures of the spike protein were available from the Protein Databank (PDB) [1], reaching now more than 336 available structures.
The aim of this work was to analyze and characterize the interaction between the Spike Receptor Binding Domain (RBD) and the human ACE2. Then, the surface properties of the RBD was explored to identify pockets that could be recognized by therapeutic molecule to block the RBD-ACE2 interaction. The druggability of a target (its ability to bind drug-like molecules), specifically of its binding site, can be predicted from its 3D structure [2] using physicochemical and geometrical parameters to characterize the pockets. In our study, two RBD-ACE2 complex structures (6M0J and 6LZG PDB) were used to understand the interaction mechanism of both proteins. As proteins are known to be highly flexible [3], the RBD-ACE2 complex and the isolated RBD domain were studied through Molecular Dynamics simulations (MD) using GROMACS software [4] in order to identify the proteins's movements, predict pockets emergence of the isolated RBD during MD and characterize key residues involved in the interaction of the complex. We ran 20 simulations of 100ns each to cover a wide range of trajectories and to sample different conformational spaces of our systems. An extensive pocket search was conducted to detect druggable pockets in the RBD protein along the simulations using the PockDrug software [5]. A multivariate statistical method has been applied to analyze the protein pockets extracted throughout the MD. The free binding energy of the complex were computed using the Molecular Mechanics Poisson-Boltzmann Surface Area (MM/PBSA) method [6] to identify key residues in this RBD-ACE2 interaction. Altogether, our study helped us to identify interesting druggable pockets comprising crucial key residues for the RBD-ACE2 interaction and that can be easily targeted by efficient inhibitors in order to prevent the virus infection.

Berman, H. M. et al., The protein data bank. Nucleic Acids Res, 2000
Abi Hussein, H. et al., Global vision of druggability issues, applications and perspectives, DRUG Discov Today, 2017
Regad, L. et al., Exploring the potential of structural alphabet-based tool for mining multiple target conformations and target flexibility insight. PLoS One, 2017
Abraham, M. J. et al., GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX, 2015
Borrel, A. et al., PockDrug-Server: a new web server for predicting pocket druggability on holo and apo proteins. Nucleic Acids Research, 2015
Kumari, R. et al., g_mmpbsa – A GROMACS tool for hight-throughput MM-PBSA calculations. J. Chem. Inf. Model, 2014

PP42
Insect odorant co-receptor ORco agonists binding mechanism

Jody Pacalon*(a), Guillaume Audic*(b), Christophe Moreau(b), Jérémie Topin(a) and Jérôme Golebiowski(a)(c)

(a) Université Côte d'Azur (UCA), UMR 7272, Faculté des Sciences Parc Valrose 28 Avenue Valrose, Nice (France).
(b) Institut de Biologie Structurale (IBS), 71 Avenue des Martyrs, 38000 Grenoble (France).
(c) Department of Brain and Cognitive Science, DGIST, Daegu, Republic of Korea.
* Contact: jpacalon@unice.fr

Insects live in an olfactory world, as finding food, mating partners, avoiding predators, or communicating is mostly based on odorant emission and detection. As insects represent a major challenge in our society, both in agriculture (~US$70 billion per year) and healthcare (~US$6.9 billion per year), discovery of chemicals acting on their behaviours is crucial.

Insects' olfactory neurons sense chemicals through activation of multimeric ion channels where highly variable odorant receptors are coupled with a conserved co-receptor (ORco). While ORs are sensitive to a large diversity of volatile compounds, only few synthetic agonists are known to activate ORco.

To identify the binding site and the ligand diffusion pathway into ORco, multiple molecular dynamics simulations (~20 µs in total) were carried. An ion channel made up of four ORco protomers was embedded into a lipid bilayer and surrounded by several agonists. The simulation analysis provided a rational approach to guide in vivo functional assays. We identified and described the binding pathway through which the insect ORco recognize its ligands. This finding opens the way to the rational design of insect repellents.

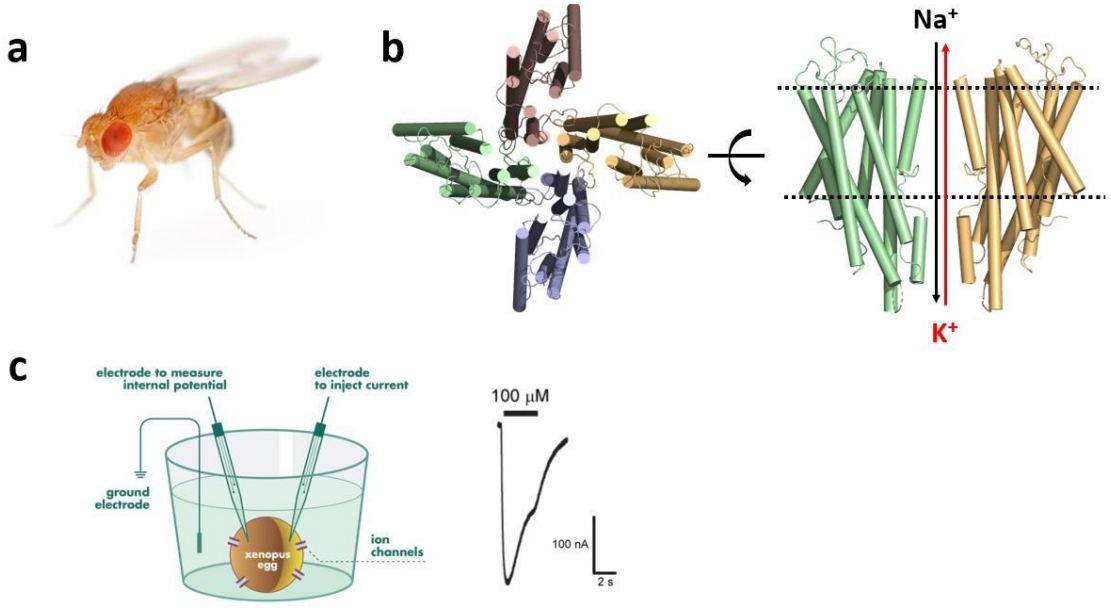


Fig: (a) Drosophila Melanogaster (D. mel) picture. (b) D. Mel ORco tetramer system top and side-view. (c) Voltage clamp recordings from Xenopus oocytes.

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No 682286) and from the Edmond Roudnitska foundation.

PP43
Design of GPCR agonists by free energy calculations

Panel Nicolas[a], Nour Aldin Kahlous[a], Duc Duy Vo[a], Hübner Harald[b], Stefanie Kampen[a], Andreas Luttens[a], Jan Kihlberg[c], Pierre Matricon[a], Pacalon Jody[a], Gmeiner Peter[b], Carlsson Jens[a]

a Science for Life Laboratory, Department of Cell and Molecular Biology, Uppsala University,
BMC Box 596, SE-751 24 Uppsala, Sweden
b Department of Chemistry and Pharmacy, Friedrich Alexander University, Schuhstrasse 19, 91052 Erlangen, Germany.
c Department of Chemistry-BMC, Uppsala University, Box 576, SR-75123 Uppsala, Sweden

G protein-coupled receptors (GPCRs) represent the largest membrane protein family and are targeted by around 40% of the marketed drugs.1 GPCRs exist in an ensemble of conformations that can be stabilized by the binding of a ligand or by the interaction with intracellular partners, e.g. G proteins. Over the past decade the increasing number of available crystal structures of GPCRs has provided the opportunity to get access to those different states and understand the activation process at an atomic level. Upon receptor activation, the binding site undergoes conformational changes leading to an increase of protein-agonist interactions and affinity. This affinity change is not observed for antagonists and suggests the existence of specific and subtle interactions that are required to trigger receptor activation.2,3 Identification of new agonists could help to understand the receptor activation.

We have applied molecular docking and molecular dynamics free energy perturbation calculations (MD/FEP) to design GPCR agonists of the prototypical β2 adrenergic receptor (β2R). The method was first tested on a retrospective set of ligands and was able to discriminate agonists from antagonists. Molecular docking was then used to screen a large virtual library against the active conformation of the receptor. Six non-catechol ligands were then successfully synthesized and experimentally validated, leading to the identification of three new β2R partial agonists and one antagonist with nanomolar to micromolar affinities. Two of the initial hits were extended to a series of seven analogs showing weak to partial agonist activity, in agreement with predictions.

The results demonstrate that MD simulations of GPCRs in active and inactive conformations can enable design of ligands with specific signaling profiles. In contrast to pharmacophore-based models, our approach is not dependent on training sets and can hence be applied directly to design agonists of other therapeutically targets in the GPCR family.

1 Congreve, M.; de Graaf, C.; Swain, N. A.; Tate, C. G. Impact of GPCR Structures on Drug Discovery. Cell 2020, 181 (1), 81–91.
2 Staus, D. P.; Strachan, R. T.; Manglik, A.; Pani, B.; Kahsai, A. W.; Kim, T. H.; Wingler, L. M.; Ahn, S.; Chatterjee, A.; Masoudi, A.; Kruse, A. C.; Pardon, E.; Steyaert, J.; Weis, W. I.; Prosser, R. S.; Kobilka, B. K.; Costa, T.; Lefkowitz, R. J. Allosteric Nanobodies Reveal the Dynamic Range and Diverse Mechanisms of G-Protein-Coupled Receptor Activation. Nature 2016, 535 (7612), 448–452.
3 Warne, T.; Edwards, P. C.; Doré, A. S.; Leslie, A. G. W.; Tate, C. G. Molecular Basis for High-Affinity Agonist Binding in GPCRs. Science 2019, 364 (6442), 775–778.

PP44
Identification and analysis of scaffold hopping situations using protein-ligand interaction fingerprints

Philippe Pinel [a], Brice Hoffmann [b], Stéphanie Labouille [c], Nicolas Drizard [d]

a philippe.pinel@iktos.com
b brice.hoffmann@iktos.com
c stephanie.labouille@iktos.com
d nicolas.drizard@iktos.com

An important issue in drug discovery is the ability to identify new molecules with optimized biological and chemical characteristics for a given protein target, potentially belonging to a different structural family than that of known hits. In this context, scaffold hopping refers to identification of molecules having similar (or better) activities but dissimilar core chemical structures[1].

Pharmacophore approaches - commonly used in scaffold hopping strategies[2] - are limited by the fact that they do not bear information about the molecule binding mode, while protein-ligand interaction fingerprints (IFPs) can solve this issue.

We introduce the HIFP (Hierarchical Interaction Fingerprint), a residue-based IFP encoding for 10 interactions, with additional polar and pi-interactions from known IFPs from the litterature[3, 4, 5]. We propose a hierarchical structure of the fingerprint that takes into account similarities between different interactions that may be of the same global type (i.e. polar, pi-system or hydrophobic): for example, salt bridges and H-bonds will both belong to the polar branch of the hierarchy, and therefore, would not be “orthogonal” as in IFPs. Consequently, molecules that interact with the same residues but with slightly different types of interactions but belonging to the same branch, will have similar HIFP fingerprints, although their IFPs may be dissimilar.

We explore the PDB database to search for scaffold hopping situations using HIFP and other traditional IFPs. We identify these situations when molecules binding to the same protein have different scaffolds, as assessed by a low Tanimoto similarity according to their Morgan fingerprint, but high Tanimoto similarity according to their interaction fingerprint.

We observe that different IFPs bear different and complementary information: some scaffold hopping cases are recovered by all or by specific IFPs. The interest of the proposed HIFP is that it overall recovers a larger panel of scaffold hopping situations than all other tested IFPs, because it better highlights similarities between molecules with slightly different types of interactions with the protein than traditional IFPs.

1 Bajorath J. Computational scaffold hopping: cornerstone for the future of drug design?, Future Med. Chem., 2017, 629-631.
2 Schneider G. “Scaffold-Hopping” by Topological Pharmacophore Search: A Contribution to Virtual Screening, Angew. Chem., 1999, 2894-2896.
3 Marcou G., Rognan D. Optimizing Fragment and Scaffold Docking by Use of Molecular Interaction Fingerprints, J. Chem. Inf. Model., 2007, 195-207.
4 Chupakhin V., Simple Ligand–Receptor Interaction Descriptor (SILIRID) for alignment-free binding site comparison, Computational and Structural Biotechnology Journal, 2014, 33-37.
5 Da C., Kireev D. Structural Protein−Ligand Interaction Fingerprints (SPLIF) for Structure-Based Virtual Screening: Method and Benchmark Study, J. Chem. Inf. Model., 2014, 2555−2561.

PP45
Implementation of a soft grading system for chemistry in a Moodle plugin.

Plyer Louis[a], Marcou Gilles[b]

a Université de Strasbourg, Faculté de Chimie
b Université de Strasbourg, Faculté de Chimie, Laboratoire de Chémoinformatique - UMR7140

We propose a novel approach for grading chemical structures drawings for remote teaching, integrated to the Moodle platform (Figure 1). Online platforms such as Moodle, are unevenly adapted to different disciplines. The platform uses a binary grading system, which often fails to give a nuanced evaluation of the answers. This is particularly true in the case of chemistry drawing, where most questions simply cannot be evaluated on a true/false basis. Specifically, a strict comparison of the candidate and the expected drawings are insufficient when some tolerance is deemed acceptable. To alleviate this constrain, the herein proposed grading workflow is based on computing the similarity between the chemical drawings. It is implemented as a Moodle plugin, using Chemdoodle engine for drawing structures, and communicating with a REST server to compute the similarity using ISIDA[1] descriptors and Tanimoto coefficient. The plugin is highly adaptable to any academic user; additionally, both embedding and similarity measures can be configured. This work has benefited from a state aid managed by the National Research Agency under the program “Investissements d’avenir” with the reference ANR-20-NCUN-0004 DEPHY.

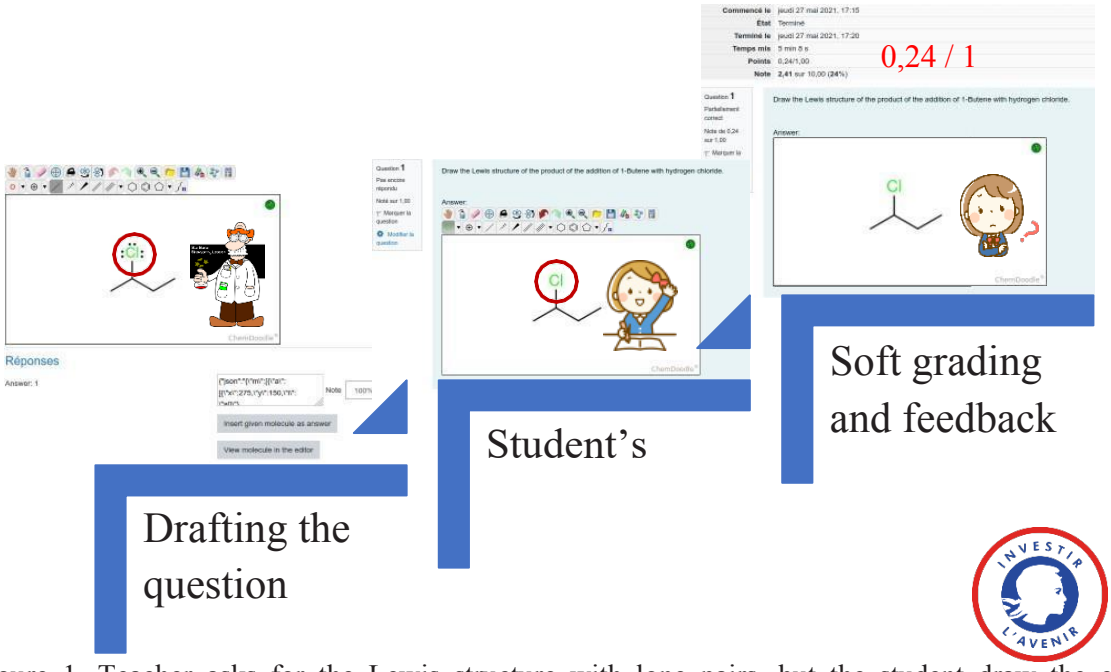


Figure 1. Teacher asks for the Lewis structure with lone pairs, but the student draw the correct structure forgetting the lone pairs. Therefore, the grading system automatically assign a grade of 0,24 out of 1 considering the correct and erroneous constituents of the answer.

1 F. Ruggiu, G. Marcou, A. Varnek and D. Horvath ISIDA Property-labeled fragment descriptors Mol. Informatics, 2010, 29, 855 – 868

PP46
The Design of non-peptide and selective ligands acting as XIAP inhibitors.

Marc RAGUI(1)*, Charline KIEFFER(1), Anne-Sophie VOISIN-CHIRET (1), Jana SOPKOVA-DE OLIVIERA SANTOS (1)

Correspondence: marc.ragui@unicaen.fr

(1) CERMN, EA 4258, Boulevard Becquerel 14032 Caen Cedex, 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 a 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 (also cIAPs) BIR domains releasing caspases and reactivating intrinsic signaling pathways leading to apoptosis. Overexpression of XIAP is involved in cancer and autoimmune diseases like Multiple sclerosis. That is why XIAP is considered a potential target especially, for cancer therapeutics[2].

In the market, there is no approved drug acting as XIAP inhibitor. Most of the molecules in clinical trials are either peptide or peptidomimetics[3] prone to peptidases actions. Also, the lack of XIAP-selectivity, even for the only non-peptidomimetic ligand was shown[4], and this could be the cause of adverse effects during patient's treatments[5]. So, the design of small and selective chemical compounds will be of great interest.

This project aims to design selective non-peptide ligands for BIR2 and BIR3 domains of XIAP, but also selective versus BIRs of other IAPs (cIAP) by optimizing previously synthesized ligands in the laboratory. The first part of this work relies on using molecular dynamics simulations associated with Poisson-Boltzmann surface methods (MM-PBSA)[6] to predict ligand-binding affinity of the hits in silico and compare theoretical results with experimental ones.

The validation of MM-PBSA for our system was done by running dynamic simulations of 50 ns on 4 different ligand-BIR3 XIAP complexes. The additive forcefield CHARMM36m was used, associated with hydrogen mass repartitioning as parametrization type, and the interaction entropy (IE)[10] as a method for calculating the entropic part. This protocol leads us to coefficient correlation (r) equals to 0.99 between the experimental results and the predicted ones.

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. Chessari, G. et al. Fragment-Based Drug Discovery Targeting Inhibitor of Apoptosis Proteins: Discovery of a Non-Alanine Lead Series with Dual Activity Against cIAP1 and XIAP. J. Med. Chem. 58, 6574–6588 (2015).
4. Ward, G. A. et al. ASTX660, a Novel Non-peptidomimetic Antagonist of cIAP1/2 and XIAP, Potently Induces TNFα-Dependent Apoptosis in Cancer Cell Lines and Inhibits Tumor Growth. Mol. Cancer Ther. 17, 1381–1391 (2018).
5. Fulda, S. Targeting Inhibitor of Apoptosis Proteins for Cancer Therapy: A Double-Edge Sword? J. Clin. Oncol. 32, 3190–3191 (2014).
6. Genheden, S. & Ryde, U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. Expert Opinion on Drug Discovery (2015) doi:10.1517/17460441.2015.1032936.
7. Zwanzig, R. W. High-Temperature Equation of State by a Perturbation Method. I. Nonpolar Gases. J. Chem. Phys. 22, 1420–1426 (1954).
8. Barbault, F. & Maurel, F. Simulation with quantum mechanics/molecular mechanics for drug discovery. Expert Opin. Drug Discov.10, 1047–1057 (2015).
9. Laio, A. & Parrinello, M. Escaping free-energy minima. Proc. Natl. Acad. Sci. 99, 12562 LP – 12566 (2002).
10. Duan, L., Liu, X. & Zhang, J. Z. H. Interaction entropy: A new paradigm for highly efficient and reliable computation of protein– ligand binding free energy. J. Am. Chem. Soc. 138, 5722–5728 (2016).

PP47
Roles of glycans in dynamics of insulin receptor

Rajas Rao[a], Hua Wong[a], Manuel Dauchez[a,b], Stéphanie Baud[a,b]

a Laboratoire MEDyC, CNRS UMR-7369, Université de Reims Champagne-Ardenne, Reims.
b Plateforme Modélisation Moléculaire Multiéchelle, Université de Reims Champagne-Ardenne , Reims.

Background: Glycosylations are one of the most abundant post-translational modifications, with about 10% of the structures on PDB having glycosylations as post-translational modifications. Despite its abundance, their role in glycoprotein dynamics and function is not well understood. In insulin receptor (IR), earlier experimental studies have shown that removal of sialic acids from its glycans is closely associated with the onset of insulin resistance[1] and degradation in IR signalling. However, how desialylation affects IR dynamics from structural perspective is not well understood.

Objective: We attempted to understand how removal of sialic acids from glycans of insulin receptor affects its dynamics and functioning.

Methods: We performed a total 17.2μs of all-atom molecular dynamics simulations of glycosylated IR in various glycosylation states. We studied the impact of desialylation on the IR dynamics on local as well as global level.

Results: Though desialylation did not have impact on global IR dynamics, it had a significant impact on IR dynamics at a local level. In particular, insulin-binding residues were more flexible in desialylated trajectories, and some insulin-binding residues were destabilized. We also observed that glycan antennae interacted with insulin-binding residues at lower frequency in desialylated condition, compared to sialylated condition. Using an in-house developed technique that maps regions of protein surface that are under the shadow of glycan, we observed that the IR region binding with the C-terminal helix domain of IR is under the glycan shadow in sialylated condition, suggesting a possible role of glycans in mediating protein-protein interactions. Finally, we found that residues destabilized in desialylated condition were found to be part of an allosteric pathway.

Conclusion: While several experimental studies highlighted the importance of glycosylations in context of various physiological processes, we here demonstrate how changes in saccharide composition in glycans affects glycoprotein dynamics and glycoprotein functioning at an atomic level. Our work sheds some interesting insights on the role of saccharide residues in the onset of insulin resistance.

1 Blaise S., Romier B., Kawecki C., Ghirardi M., Rabenoelina F., Baud S., Duca L., Maurice P., Heinz A., Schmelzer C.E., Tarpin M. “Elastin-derived peptides are new regulators of insulin resistance development in mice”. Diabetes. Nov 1;62(11):3807-16, (2013).

PP48
Deeprank-GNN: A Graph Neural Network Framework to Learn Interaction Patterns from Protein-Protein Interfaces

Manon Réau [a], Nicolas Renaud [b], Alexandre M.J.J. Bonvin [a]

a Computational Structural Biology Group, Bijvoet Centre for Biomolecular Research, Faculty of Science - Chemistry, Utrecht University, Padualaan 8, 3584 CH Utrecht, the Netherlands

b Netherlands eScience Center, Science Park 140, 1098 XG, Amsterdam, The Netherlands

Protein-protein interactions (PPIs) are essential in all cellular processes of living organisms including cell growth, structure, communication, protection and death. Acquiring knowledge on PPI is fundamental to understand normal and altered physiological processes and propose solutions to restore them. In the past decades, a large number of PPI structures have been solved by experimental approaches (e.g., X-ray crystallography, nuclear magnetic resonance, cryogenic electron microscopy). Given the remarkable success of Convolutional Neural Network (CNN) in retrieving patterns in images[1], CNN architectures have been developed to learn interaction patterns in PPI interfaces[2,3].

We have developed Deeprank[2] (https://github.com/DeepRank/deeprank), an open-source configurable deep learning framework for data mining PPIs using 3D-CNNs. Deeprank maps atomic and residue-level features from PPIs to 3D grids and applies 3D CNNs to learn problem-specific interaction patterns. Deeprank was applied to two problems: 1) the classification of biological vs. crystallographic PPIs, and 2) the scoring of models of protein-protein complexes generated by docking. Deeprank was shown to compete with- or outperform state-of-the-art methods in both scenarios.

CNNs however come with major limitations: First, they are sensitive to the input PPI orientation, and it may require data augmentation (i.e. multiple rotations of the input data) for the network to forget about the orientation in the learning process; second, the size of the 3D grid is unique for all input data, which does not reflect the variety in interface sizes observed in experimental structures and may be problematic for large interfaces that do not fit inside the predefined grid size. A solution to this problem is to use instead Graph Neural networks (GNN). By definition, graphs are non-structured geometric structures and do not hold orientation information. They are rotational invariant and can easily represent interfaces of varying sizes. We have therefore developed Deeprank-GNN that converts PPI interfaces into graphs and uses those to learn interaction patterns. We benchmarked the performance of Deeprank-GNN in scoring docking models from the CAPRI score set. Results show that it performs equally or outperforms state-of-the-art scoring functions (HADDOCK, Deeprank, DOVE, iScore) on 10/13 complexes. Of note, the current version of Deeprank-GNN maps basic conservation, physico-chemical and geometric features to the graph nodes and does not use any energetic terms. This reveals that energetic terms are not essential for the network to correctly learn favorable interactions.

Deeprank-GNN is freely available from https://github.com/DeepRank/Deeprank-GNN/

1 Krizhevsky A, Sutskever I, Hinton GE, ImageNet classification with deep convolutional neural networks. Adv Neural Inf Process Syst 25, 2012
2 Nicolas Renaud, Cunliang Geng, Sonja Georgievska, Francesco Ambrosetti, Lars Ridder, Dario Marzella, Alexandre M.J.J. Bonvin, Li C Xue, DeepRank: A deep learning framework for data mining 3D protein-protein interfaces, bioRxiv, 2021.01.29.425727
3 Wang X, Terashi G, Christoffer CW, Zhu M, Kihara D. Protein docking model evaluation by 3D deep convolutional neural networks. Bioinformatics. 2020 ;36(7):2113-2118.

PP49
A survey of the fragment space. Commercial libraries.

J. Revillo Imbernón[a], G. Marcou[b], E. Kellenberger[a]

a UMR 7200 CNRS-Université de Strasbourg, Laboratoire d'Innovation Thérapeutique, 67400 Illkirch, France

b UMR 7140 CNRS-Université de Strasbourg, Laboratory of Chemoinformatics, 4 rue B. Pascal, 67000 Strasbourg, France

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

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

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

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

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

1 Romasanta A. K. S., van der Sijde P., Hellsten I., Hubbard R. E., Keseru G. M., van Muijlwijk-Koezen J., de Esch I. J. P., When fragments link: a bibliometric perspective on the development of fragment-based drug discovery, Drug Discovery Today, 2018, 23, 1596-1609
2 Lipinski C., Hopkins A., Navigating chemical space for biology and medicine, Nature, 2004, 432, 855-861
3 Hann M. M., Keserü G. Finding the sweet spot: the role of nature and nurture in medicinal chemistry. Nat Rev Drug Discov 2012, 11, 355–365
4 Lovering F., Bikker J., Humblet C., Escape from Flatland: Increasing Saturation as an Approach to Improving Clinical Success, J. Med. Chem. 2009, 52, 6752-6756
5 Congreve M., Carr, R., Murray, C. & Jhoti, H. A 'Rule of Three' for fragment-based lead discovery. Drug Discov. Today, 2003, 8, 876–877
6 Firth N. C., Brown N., Blagg J., Plane of Best Fit: A Novel Method to Characterize the Three-Dimensionality of Molecules, J. Chem. Inf. Model, 2012, 52, 2516-2525
7 Kireeva N., Baskin I., Gaspar H. A., Horvath D., Marcou G., Varnek A., Generative Topographic Mapping (GTM): Universal Tool for Data Visualization, Structure-Activity Modeling and Dataset Comparison, Mol. Inf., 2012, 31, 301-312

PP50
A system to compare shapes based on a Wave Kernel Signature Map applied to protein surfaces

Léa Sirugue[a], Florent Langenfeld[a], Matthieu Montes[a]

a Laboratoire GBCM, EA 7528, Conservatoire National des Arts et Métiers, Hesam université, 2 rue Conté, 75003 Paris, France

Many methods exist to compare macroscopic objects in the field of computer vision. Few of them have been applied to protein surfaces, to the best of our knowledge. Proteins feature unique shapes that are challenging to compare and discriminate. In this work, we present a new method to compare proteins using their surfaces only.
Our descriptor is based on the Wave Kernel Signature (WKS)[1] descriptor and is called Projected WKS Map (PWKSM). The quick parsing of this descriptor is advantageous for a rapid protein comparison in the context of big data. The WKS descriptor is projected onto a unit sphere using a conformal transformation and then mapped onto a 2D plane. Dense maps are created by interpolating the WKS values on the 2D plane. These maps are the final descriptors called Projected WKS Maps. The use of a conformal mapping on a unit sphere reduces the loss of information due to the projection from a Euclidian space to a Riemann space.
We extracted surficial homologs (i.e., proteins with high surface similarity but low sequence similarity) with sequence homology below 30% from the PPI4DOCK2 dataset to compare our method to state-of-the-art methods (WKS[1], HKS[3] and FPFH[4]). The experimental results on this dataset show performances similar to the state-of-the-art methods in computer vision. Moreover, a comparison is in average faster or equivalent with our descriptor than with other methods.

1 Aubry, M., Schlickewei, U., & Cremers, D. (2011, November). The wave kernel signature: A quantum mechanical approach to shape analysis. In 2011 IEEE international conference on computer vision workshops (ICCV workshops) (pp. 1626-1633). IEEE.
2 Yu, J., & Guerois, R. (2016). PPI4DOCK: large scale assessment of the use of homology models in free docking over more than 1000 realistic targets. Bioinformatics, 32(24), 3760-3767.
3 Sun, J., Ovsjanikov, M., & Guibas, L. (2009, July). A concise and provably informative multi-scale signature based on heat diffusion. In Computer graphics forum (Vol. 28, No. 5, pp. 1383-1392). Oxford, UK: Blackwell Publishing Ltd.
4 Rusu, R. B., Blodow, N., & Beetz, M. (2009, May). Fast point feature histograms (FPFH) for 3D registration. In 2009 IEEE international conference on robotics and automation (pp. 3212-3217). IEEE.

PP51
With or without sugar? A search for lectin inhibitors.

Jérémie Topin[a], Sakonwan Kuhaudomlarp[b,c,d], Annabelle Varrot[b], Anne Imberty[b]

a Université Côte d'Azur, ICN, UMR 7272 CNRS, 06108 Nice, France

b Univ. Grenoble Alpes, CNRS, CERMAV, 38000 Grenoble, France

c Department of Biochemistry, Faculty of Science, Mahidol University, Bangkok 10400, Thailand

d Center for Excellence in Protein and Enzyme Technology, Faculty of Science, Mahidol University, Bangkok 10400, Thailand

Because of the antimicrobial resistance crisis, lectins are considered novel drug targets. Interactions between oligosaccharides and lectins constitute a key step in the first stage of recognition and tissue adhesion in several bacterial infections. Recent advances in glycobiology revealed the essential role of lectins for deciphering the glycocode by specific recognition of carbohydrates. Ligands competing with human glycoconjugates for lectins binding are thus promising candidates to counteract lung injury, mortality, and cellular invasion.

Histo-blood group epitopes are fucosylated branched oligosaccharides with well-defined conformations in solution that are recognized by receptors, such as lectins from pathogens. We have extensively studied the flexibility of histo-blood group antigens combining experimental and multiple molecular dynamics simulations. Our findings show that conformational adaptation of oligosaccharides is of paramount importance in cell recognition and should be considered when designing anti-infective glyco-compounds.[1] We have identified a novel divalent ligand from a focused galactoside(Gal)-conjugate array which binds to the lectin LecA with a nanomolar affinity. The flexibility of the spacer studied by molecular dynamics simulations favored optimal contact with the protein surface and resulted in a gain in enthalpy.[2]

The development of non-carbohydrate mimics would be an alternative strategy to find high affinity ligands and to explore new avenues for lectin inhibition. It will then enlarge the chemical space for novel inhibitors. Indeed, since carbohydrate-protein interactions are mostly governed by a complex arrangement of hydrogen bonds and hydrophobic contacts, the success of non-carbohydrate analogues is challenging. The virtual screening of the National Cancer Institute (NCI) diversity set IV lead to the identification of a catechol as the first non-carbohydrate lectin ligand that binds bacterial and mammalian calcium(II)-binding lectins. This result gives rise to a fundamentally new class of glycomimetics.[3]

Altogether, our results demonstrate that the use of numerical approaches is of major importance to reveal biomolecular mechanisms at spatial and temporal scales that remain difficult to observe experimentally.

1 J. Topin, M. Lelimousin , J. Arnaud, A. Audfray, S. Pérez, A. Varrot, A. Imberty . The Hidden Conformation of Lewis x, a Human Histo-Blood Group Antigen, Is a Determinant for Recognition by Pathogen Lectins, ACS Chem Biol. 2016 Jul 15;11(7):2011-20. doi: 10.1021/acschembio.6b00333. Epub 2016 May 20. PMID: 27198630.
2 A. Novoa, T. Eierhoff, J. Topin, A. Varrot, S. Barluenga, A. Imberty, W. Rçmer, and N. Winssinger ; A LecA Ligand Identified from a Galactoside- Conjugate Array Inhibits Host Cell Invasion by Pseudomonas aeruginosa, Angewandte Chemie International Edition, 2014, DOI: 0.1002/ange.201402831.
3 S. Kuhaudomlarp, E. Siebs, E. Shanina, J. Topin, I. Joachim, P. da Silva Figueiredo Celestino Gomes, A. Varrot, D. Rognan, C. Rademacher, A. Imberty, A. Titz, Non-Carbohydrate Glycomimetics as Inhibitors of Calcium(II)-binding Lectins, Angewandte Chemie International Edition, 2020, doi.org/10.1002/ange.202013217

PP52
LIT-PCBA: An Unbiased Data Set for Machine Learning and Virtual Screening

Viet-Khoa Tran-Nguyen[a], Guillaume Bret[a], Didier Rognan[a],*

a Laboratoire d'Innovation Thérapeutique, Université de Strasbourg, 74 Route du Rhin, 67400 Illkirch- Graffenstaden, France

* Correspondence: rognan@unistra.fr

Serious biases related to commonly used datasets for retrospective structure-based virtual screening studies (e.g. DUD, DUD-E, ChEMBL) have been reported in recent years[1,2]. The composition of each dataset has been heavily biased, as the quantity of active compounds is usually too high; the potency of presumably inactive compounds always remains unknown; and the actives are too similar (in 2D) to each other and to the crystallographic reference structures deposited on Protein Data Bank, while remarkably different from the inactive counterparts. Such datasets do not mimic chemolibraries used for high throughput screening in reality, tend to overestimate the performance of virtual screening methods, and are not recommended for benchmarking purposes. The need to design a novel and unbiased database dedicated to structure-based in silico screening approaches therefore arises. We herewith present the newly designed LIT-PCBA database, consisting of 21 datasets representing 11 protein families of pharmaceutical interest (including kinases, GPCRs, nuclear receptors and other targets), which was constructed based on the experimental results of biological tests deposited on PubChem's BioAssays, thus confirming the potency of active and inactive compounds[3]. All substances were prepared and filtered in such a way that assay artifacts (false positives) as well as artificial enrichment were prevented according to our selection rules and those previously explained by Rohrer SG and Baumann K[4]. The ratio between the number of active compounds and that of inactives has been greatly reduced, and the potency of remaining actives is remarkably lower than that found in the DUD-E database. Retrospective virtual screening results using two ligand-based methods (2D geometry similarity search by ECFP4 and 3D geometry similarity search with ROCS) and a structure-based approach (molecular docking with Surflex-Dock) show that screening performances (ROC AUC, BEDROC AUC, EF1%) varied depending on the PDB template structure that was used for each set and the method that was employed, and there is little structural bias that remains among the compounds that constitute most datasets. The LIT-PCBA database can therefore be used to compare the real accuracy of scoring functions in future benchmarking research.

1 Chaput L et al. Benchmark of four popular virtual screening programs: construction of the active/decoy dataset remains a major determinant of measured performance, J. Cheminform., 2016, 8, 56.
2 Sieg J et al. In Need of Bias Control: Evaluating Chemical Data for Machine Learning in Structure-Based Virtual Screening, J. Chem. Inf. Model., 2019, 59, 947-961.
3 https://pubchem.ncbi.nlm.nih.gov/ (accessed Jul 31, 2019).
4 Rohrer SG, Baumann K. Maximum Unbiased Validation (MUV) Data Sets for Virtual Screening Based on PubChem Bioactivity Data, J. Chem. Inf. Model., 2009, 49, 169-184.

PP53
Modélisation par homologies d'Adénylyl Cyclases et de leurs interactions

Mahesh Velusamy[a], Yohan Legueux_Cajgfinger[b], Pierre Vincent[b], Régis Blaise[b], Laurent Duca[a],
Stéphanie Baud[a], Isabelle Limon[b], Manuel Dauchez[a]

a UMR CNRS 7369 MEDyC, Chaire MAgICS, Université de Reims Champagne-Ardenne, Reims
b UMR CNRS 8256 B2A, Institut de Biologie Paris-Seine, Sorbonne Université, Paris.

L'athérosclérose et les pathologies cardiovasculaires associées est la 1ère cause de mortalité mondiale (OMS, 2015). La pause de stent entraîne des problématiques par exemple de restenose. Dans ce contexte, le contrôle des Cellules Musculaires Lisses Vasculaires (CMLV) est un objectif pour contrer cette resténose et la famille des Adenylyl Cyclases (AC), protéines membranaires, une famille de protéines qui produit, à partir de l'ATP, de l'AMPc, est une cible prioritaire. Elles sont au nombre de 9 isoformes ayant une distribution tissulaire spécifique et des propriétés régulatrices associées.

Dans ce travail, nous avons entrepris la modélisation moléculaire par homologie de l'AC8, isoforme exprimée spécifiquement dans les CMLV et une de ses formes tronquées l'AC8E. Les ACs sont des protéines transmembranaires présentant deux cassettes de 6 hélices transmembranaires (cassettes dites M1 et M2). La forme tronquée AC8E présente la délétion des hélices 1 à 5 de la cassette M1 et perturbent donc les mécanismes fonctionnelles de ces protéines et de leurs associations par dimérization.

A partir de la structure de l'AC9 obtenue récemment par Cryo-EM[1] nous avons proposé des modèles des AC8, AC8E et AC3. L'AC9 étant incomplète les domaines manquant ont dû être obtenu par modélisation par homologie. Les différents modèles ont été optimisés et insérés dans les environnements membranaires. La conséquence des troncations sur les mécanismes d'action sera présentée et les résultats confrontant les interactions possibles de l'AC8 et AC8E avec l'AC3 sur des données de PEPscan sont analysés en terme de séquences potentielles d'intérêt. Ces peptides transmembranaires sont susceptibles de déstabiliser les dimères et les tests in vitro préliminaires semblent confirmer les travaux de modélisation.

1 . C. Qi, S. Sorrentino, O. Medalia, V. Khorkov, "The structure of a membrane adenylyl cyclase bound to an activated stimulatory G protein", Science, 364, 389-394, (2019).

PP54
Structural basis of the transmembrane domain of the yeast mitofusin

Raphaelle Versini[1,2], Antoine Taly[2] and Patrick Fuchs[1]

1 Laboratoire des BioMolecules, Sorbonne University, Ecole Normale Supérieure, PSL Research University, CNRS, Paris, France

2 Laboratoire de Biochimie Théorique, CNRS, Institut de Biologie Physico-Chimique-Fondation Edmond de Rothschild, PSL Research University, Paris, France

Outer mitochondrial membrane (OMM) fusion is an important process for the cell and organism survival, as its dysfonction is often linked to neurodegenerative diseases. The OMM fusion is mediated by members of the dynamin-related protein (DRP) family, named mitofusins. Fzo1, the only mitofusin homologue of the yeast Saccharomyces cerevisiae and embedded in the OMM, was modeled in literature by homology with the mitofusin related bacterial dynamin-like protein (BDLP) as template[1]. However BDLP does not possess any transmembrane part. Thus, the structure of the Fzo1 transmembrane domain, made of two putative helices TM1 and TM2, had to be determined using ab initio methods. One study in literature predicted the structure of Fzo1 transmembrane domain using the webserver PREDIMMER.

Coarse-grained simulations using a force field such as MARTINI are usually used to enhance sampling. However, MARTINI2 has been shown to overaggregate proteins, issue that has been resolved in the latest version of the force field MARTINI3. In this work, we compare results from MARTINI2 and MARTINI3, and produce an improved model of TM1/TM2 from Fzo1. The clustering method GROMOS was used to extract the models from the simulations. MARTINI2 presented a higher variability in the conformations available, and a greater number of clusters than MARTINI3. The retrieved structures with MARTINI3 were found to be robust and different from previous predictions using the webserver PREDDIMER.

1 De Vecchis D, Cavellini L, Baaden M et al (2017) A membrane-inserted structural model of the yeast mitofusin Fzo1. Sci Rep 7:10217. https://doi.org/10.1038/s41598-017-10687-2

PP55
Applicability of graph neural networks to predict binding affinities from protein-ligand structures.

Mikhail Volkov * [1,2], Joseph-André Turk [2], Nicolas Drizard [2], Nicolas Martin [2], Brice Hoffmann [2], Didier Rognan [1]
* E-mail: mvolkov@unistra.fr

1 Laboratoire d'innovation thérapeutique, UMR7200 CNRS-Université de Strasbourg, 74 Route du Rhin, F-67400 Illkirch
2 Iktos, 65 rue de Prony 75017, Paris, France

The recent progress in retrieval of structural data on proteins available in public databases such as PDBBind [1] made it possible to apply deep learning models to the prediction of binding affinities from protein-ligand 3D structures. Multiple NN architectures targeted at this problem have been proposed [2, 3], including the message passing neural networks (MPNN) [4], which can be applied to raw graph representation of chemical information without a prior feature engineering. Despite that, recent publications show that the structural data available for training of these models are strongly biased to ligand and protein information [5]. This issue limits the applicability domain of models trained on currently available datasets.
In the current work, we present a deep learning model based on the MPNN architecture that can be trained on various combinations of graph representations of a protein-binding site, a ligand and an interaction pattern. The rotationally invariant graph representation of protein binding site based on the characteristic properties of amino-acid residues [6] and protein ligand interaction patterns [7] were developed. Our MPNN model demonstrated performance close to the state-of-the-art models in the field (R2 = 0.66 on the PDBBind 2013 core set). We showed that the maximal performance of the model could be reached using a combination of protein and ligand inputs without explicit information about protein-ligand interactions. Then, in order to remove the bias coming from ligand and protein structures, we introduce an iterative unbiasing procedure by removing samples for which the binding affinity is more easily predicted by a random forest model trained on generic descriptors of ligands or proteins alone. The impact of this method on model training and evaluation is studied.

1. Liu, Z., Li, Y., Han, L., Li, J., Liu, J., Zhao, Z., ... & Wang, R. (2015). PDB-wide collection of binding data: current status of the PDBbind database. Bioinformatics, 31(3), 405-412.
2. Stepniewska-Dziubinska, M. M., Zielenkiewicz, P., & Siedlecki, P. (2018). Development and evaluation of a deep learning model for protein–ligand binding affinity prediction. Bioinformatics, 34(21), 3666-3674.
3. Jiménez, J., Skalic, M., Martinez-Rosell, G., & De Fabritiis, G. (2018). K deep: protein–ligand absolute binding affinity prediction via 3d-convolutional neural networks. Journal of chemical information and modeling, 58(2), 287-296.
4. Karlov, D. S., Sosnin, S., Fedorov, M. V., & Popov, P. (2020). graphDelta: MPNN scoring function for the affinity prediction of protein–ligand complexes. ACS omega, 5(10), 5150-5159.
5. Yang, J., Shen, C., & Huang, N. (2020). Predicting or pretending: artificial intelligence for protein-ligand interactions lack of sufficiently large and unbiased datasets. Frontiers in pharmacology, 11, 69.
6. Schmitt, S., Kuhn, D., & Klebe, G. (2002). A new method to detect related function among proteins independent of sequence and fold homology. Journal of molecular biology, 323(2), 387-406.
7. Da Silva, F., Desaphy, J., & Rognan, D. (2018). IChem: A Versatile Toolkit for Detecting, Comparing, and Predicting Protein–Ligand Interactions. ChemMedChem, 13(6), 507-510.

PP56
Multiscale modelling of the Extracellular Matrix (ECM)

Hua Wong[a], Jean-Marc Crowet[a], Manuel Dauchez[a], Sylvie Ricard-Blum[b], Stéphanie Baud[a], Nicolas Belloy[a].

a UMR CNRS 7369 MEDyC, Chaire MAgICS, Université de Reims Champagne-Ardenne, Reims
b UMR CNRS 5246 IMSCB, Equipe PESA, Université de Lyon Claude Bernard, Lyon.

The extracellular matrix is a complex three-dimensional network of molecules that provides cells with a complex microenvironment. The major constituents of the extracellular matrix such as collagen, elastin and associated proteins form supramolecular assemblies contributing to its physicochemical properties and organization. The structure of proteins and their supramolecular assemblies such as fibrils can be studied at the atomic level (e.g., by X-ray crystallography, Nuclear Magnetic Resonance and cryo-Electron Microscopy) or at the microscopic scale. However, many protein complexes are too large to be studied at the atomic level and too small to be studied by microscopy methods. Most extracellular matrix components fall into this intermediate scale, so-called the mesoscopic scale preventing their detailed characterization. Simulation and modelling are some of the few powerful and promising approaches that can deepen our understanding of mesoscale systems. We have developed a set of modelling tools to study the self-organization of the extracellular matrix and large motion of macromolecules at the mesoscale level, by taking advantage of the dynamics of articulated rigid bodies as a mean to study a larger range of motion at the cost of atomic detail.

PP57
Design of RNA oligonucleotides against the Beta-Secretase 1 (BACE1) enzyme involved in Alzheimer's disease

Taher YACOUB [a], Roy GONZALEZ-ALEMAN [a,b], Fabrice LECLERC [a]

a Université Paris-Saclay, CEA, CNRS, Institute for Integrative Biology of the Cell (I2BC), 91198, Gif-sur- Yvette, France.
b Laboratorio de Química Computacional y Teórica , Facultad de Química, Universidad de La Habana, Zapata y G, Vedado 10400, La Habana, Cuba.

Background: Amyloid-β 42 oligomers (Aβ-42) aggregating in plaques in the brain, plays a critical role in the occurrence of Alzheimer's disease. The β-secretase 1 (BACE1) enzyme cleaves the initiates the Amyloid Protein Precursor (APP), the limiting step of the reaction that generates the Aβ-42 fragments[1]. Different BACE1 inhibitors, targeting the catalytic site, failed in clinical trials because of side effects[2]. This may be due in particular to some toxicity (dose effects) or lack of specificity and/or selectivity (off- target effects) due to interactions with homologous enzymes such as BACE2.

Objectives: To improve the specificity, we consider two strategies. The first is to establish specific interactions with the catalytic site of BACE1. The second is to target a distant site (exosite) with an effect on the catalytic site, having a lower level of homology with BACE2. Our objective is to design de novo the most specific active RNA oligonucleotides, standard or chemically modified, targeting the catalytic site or the exosite.

Methods: The design of these oligonucleotides relies on fragment-based and structure-based in-silico approaches. We identified, using a clustering method, some representative 3D structures (RCSB PDB, https://www.rcsb.org/) of BACE1 having different conformational states at the catalytic site and exosite. The fragment-based strategy builds on the MCSS method (Multiple Copy Simultaneous Search)3. MCSS allows to map chemical functional groups at the surface of a target, making possible to perform virtual screening using pre-defined or customized fragment libraries3,4. It has been recently applied to the prediction of the binding mode and selectivity of nucleotide ligands5. Nucleotide residues mapped at the surface of BACE1 by MCSS are selected based on their score and used as fragments to generate oligonucleotides. Further optimization of the connected nucleotides generates oligonucleotides ranked by score to select potential ligand candidates.

Results: We applied the MCSS method on beforehand selected structures of BACE1 from RSCB PDB which have different conformations. We generated different short selective oligonucleotides. Each oligonucleotide was designed to be specific either for the active site or the exosite of BACE1.
Conclusions: In conclusion, we designed potential RNA ligands against BACE1, an enzyme involved in the onset of Alzheimer's disease. Further studies are still required to rank the ligand candidates using more accurate descriptions of the molecular environment and taking into account the dynamics of the complexes. The more promising oligonucleotide ligands will be proposed for experimental validations.

1. Hampel H, Vassar R, De Strooper B, Hardy J, Willem M, Singh N, Zhou J, Yan R, Vanmechelen E, De Vos A, Nisticò R, Corbo M, Imbimbo BP, Streffer J, Voytyuk I, Timmers M, Tahami Monfared AA, Irizarry M, Albala B, Koyama A, Watanabe N, Kimura T, Yarenis L, Lista S, Kramer L, Vergallo A. The β-Secretase BACE1 in Alzheimer's Disease. Biological psychiatry, 2021, 89, 745-56.
2. Das B, Yan R. A Close Look at BACE1 Inhibitors for Alzheimer's Disease Treatment. CNS drugs, 2019, 33, 251-63
3. Miranker A, Karplus M. Functionality maps of binding sites: A multiple copy simultaneous search method. Proteins: Struct., Funct., Genet., 1991, 11, 29–34
4. Leclerc F, Karplus M. MCSS-based predictions of RNA binding sites. Theor Chem Acc, 1999, 101, 131–137
5. González-Alemán R, Chevrollier N, Simoes M, Montero-Cabrera L, and Leclerc F. MCSS-Based Predictions of Binding Mode and Selectivity of Nucleotide Ligands. J. Chem. Theory Comput., 2021, 17, 2599–2618

VIRTUAL POSTERS

VP1
Evaluation of AutoDock and AutoDock Vina on the CASF-2013 benchmark

Thomas Gaillard

thomas.gaillard@polytechnique.edu

Laboratoire de Biochimie (CNRS UMR7654), Department of Biology, Ecole Polytechnique, 91128 Palaiseau, France.

Computer-aided protein-ligand binding predictions are a valuable help in drug discovery. Protein-ligand docking programs generally consist of two main components: a scoring function and a search algorithm. It is of interest to evaluate the intrinsic performance of scoring functions, independently of conformational exploration, to understand their strengths and weaknesses, and suggest improvements. The comparative assessment of scoring functions (CASF) provides such an evaluation. Here we add the AutoDock and Vina scoring functions to the CASF-2013 benchmark. We find that these popular, free software docking programs are generally in the first half (AutoDock) and first quarter (Vina) among all methods tested in CASF-2013. Vina is the best of all methods in terms of docking power. We also find that ligand minimization has an important impact, reducing the performance difference between AutoDock and Vina.

VP2
Conception of 3D models of reduced models of Tau protein aggregates for conception and synthesis of potential Tau aggregation disruptors

Johanna GIOVANNINI*[a,b], Marie JOUANNE[a], Jana SOPKOVA-DE OLIVEIRA SANTOS[a], Marco CATTO[b] and Anne Sophie VOISIN-CHIRET[a]

[a] Centre d'Etudes et de Recherche sur le Médicament de Normandie, Normandie Université, Unicaen, CERMN, 14000 Caen, FRANCE. [b]Dipartimento di Farmacia – ScienzedelFarmaco, UniversitàdegliStudi di Bari 'Aldo Moro', 70125 Bari, ITALIA.

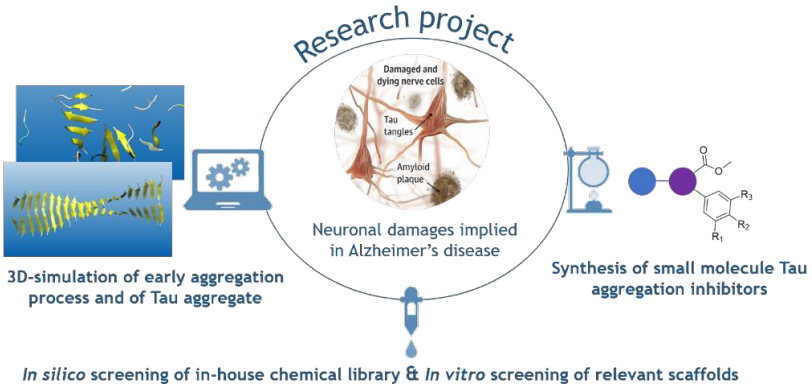
* Correspondence: johanna.giovannini@unicaen.fr

Alzheimer's disease is a neurodegenerative illness characterized by short term memory confusion, executive performance disturbance and space and time orientation function disruption. Alzheimer's disease patient brain studies constantly reveal two types of damages: amyloid plaques and neurofibrillary tangles. Both are related to abnormal protein aggregation: beta-amyloid peptide (Aβ) for amyloid plaques and tau protein for neurofibrillary tangles. For both proteins, key-peptide sequences were identified as responsible for early oligomerization, initiating the whole amyloidogenic process.[1,2] In fact, those peptides adopt a beta-sheet structuration and pile themselves up, guiding the protein aggregation initiation.

We are aiming to synthesize small molecules as protein-protein interaction disruptors in order to prevent early stage aggregation.

The present work was initiated by a conformational analysis of the Tau key hexapeptide implied in Tau aggregation, called PHF6 (Paired Helical Filament hexapeptide). We built preformed PHF6 aggregates and assessed their stabilities through Molecular Dynamic (MD) simulations and analyses of intra and intermolecular interactions. Then, MD simulations of PHF6 aggregates with palmatine, a Tau aggregation disruptor[3] were launched and mechanisms of aggregation disruption will be proposed. Then, similarity screening of our in-house chemical library[4](d) based on palmatine and in vitro tests provided 40 scaffolds as starting points for the rational design and synthesis of small molecules that could disturb the amyloid fibril interactions.

Finally, we set up a new kind of 3D-simulation to comprehend early Tau protein aggregation process that bring up together disordered PHF6 gradually aggregating. This simulation was also carried out with palmatine in order to assess its aggregation inhibition capacity in very early stages.



1. Ahmed, M., Davis, J., *et al.*, Nat. Struc. Mol. Biol., 2010, 17, 561-567.
2. Von Bergen, M., Barghorn, S., *et al.*, J. Biol. Chem., 2001, 276, 48165-48174.
3. Haj, E., Losev, Y., *et al.*, BBA – General Subjects, 2018, 1862, 1565-1575.
4. cermn.unicaen.fr/plateformes/chimiotheque/

VP3
Study by the DFT of the structural and electronic properties of Isoquercetin

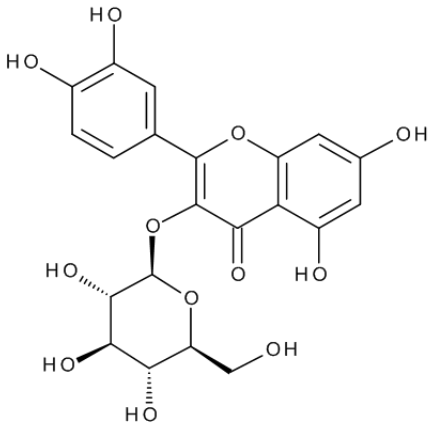
Samiha Karoui,[1*] Abdelatif Messaoudi,[1] Serge Antonczak[2]
samiha.karoui@univ-batna.dz

a Laboratoire Chimie des Matériaux et des Vivants: Activité-Réactivité (LCMVAR), Département chimie, Faculté des Sciences de la Matière, Université Batna 1, Algérie.
b Institut de Chimie de Nice, Universitè Côte d'Azur, 06108 Nice, France.

Keywords: Molecular modeling, Flavonoid, Anti-oxidant, Antiradical, Activated forms.

Abstract

Flavonoids are well known to be widely distributed in many fruits and vegetables. They showed various biological activities including anti-inflammatory, anti-oxidant and antiradical properties.
The calculations were made using quantum chemistry methods, choosing the DFT calculation method at the B3LYP / calculation level (6-31 g (d) for C and H atoms, 6-31 + g(d) for O atoms), which allowed us to study the electronic, structural and energetic properties of various activated forms of Isoquercetin.



A comparison of experimental spectral data (1H NMR, 13C NMR, IR and UV-Vis) with those obtained theoretically was performed.

[1] S. fiorucci, J Golebiowski, D. Cabrol-bass, S. Antonczak, DFT study of quercetin activated forms involved in antiradical, antioxidant, and prooxidant biological processes, J. Agric. Food Chem, 2007,55(3),903-911.
[2] Sandhar HK, Kumar B, Prasher S, Tiwari P, Salhan M, Sharma P. A review of phytochemistry and pharmacology of flavonoids. Internationale Pharmaceutical Sciencia2011,1, 25-41.

VP4
Role of the geometric characteristics on the carbon nanotube conductance: a molecular dynamic simulations study.

Alia Mejri[a], Guillaume Herlem[a] and Fabien Picaud[a]

[a]*Laboratoire de Nanomédecine, Imagerie et Thérapeutique, Université Bourgogne-Franche-Comte (UFR Sciences et Techniques), EA 4662, Centre Hospitalier Universitaire de Besançon, 16 route de Gray, 25030 Besançon, France.*

During the past decade, understanding mass transport through nanoscale channels have received great attention, particularly water flow and ions. Computer simulations such as molecular dynamics become widely used to sense water and ions motion inside graphene nanopores providing supplementary information to experiments[1].
In this work, we performed all atom molecular dynamic simulations of a flow of solvated ions moving inside carbon nanotubes under the application of an external electric field allowing relevant measurements of the ionic current established in the internal area of the tube. The geometric characteristics of carbon nanotube (tube length, chirality, diameter and chemical functions at the end) were investigated in order to report their effect on the conductance measurements.
Moreover, several water models were tested in our simulations. From these, a set of three sites rigid and non-polarizable water models were selected: TIP3P, SPC/E and TIP4/2005[2]. We will show that these models are responsible for water flow and ions motion differences through the carbon cage and consequently have a potential impact on the ionic conductance determination of the system.

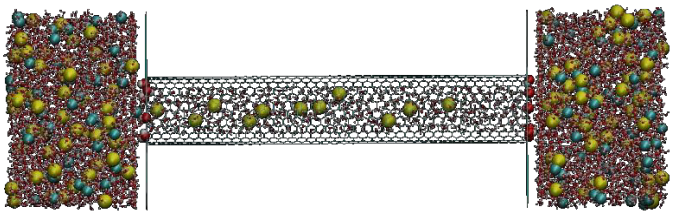


Figure: Ion transport inside carbon nanotube connecting two electrolyte reservoirs.

1. Wang, X.; Shi, G.; Liang, S.; Liu, J.; Li, D.; Fang, G.; Liu, R.; Yan, L.; Fang, H. Unexpectedly high salt accumulation inside carbon nanotubes soaked in dilute salt solutions. Phys. Rev. Lett. 2018, 121, 226102.
2. Liu, L.; Patey, G. N. Simulated conduction rates of water through a (6,6) carbon nanotube strongly depend on bulk properties of the model employed. Adv. Mat. 2018, 30, 1702419.

This work was supported by the ANR-18-CE09-0011-02 IONESCO project of the French “Agence Nationale de la Recherche”.

VP5
Contributions of molecular dynamics simulations to the understanding of rhamnolipid interactions with biomimetic fungal membranes

Nely Rodríguez-Moraga[1], Catherine Sarazin[1], Sébastien Buchoux[1,*]

[1] Laboratoire de Génie Enzymatique et Cellulaire (GEC) UMR7025 CNRS, Université de Picardie Jules Verne

In agriculture, the intensive use of chemical pesticides is environmentally unfriendly and potentially harmful to human health. Therefore, the development of new approaches with less environmental and health impacts is quite a crucial challenge. Within this context, this project is focused on a family of natural amphiphilic glycolipids called rhamnolipids (RLs). They are produced by some microorganisms, present antibacterial and / or antifungal properties, and remain harmless to humans and stimulating plant defenses. They are thus a reasonable alternative to agrochemicals. However, even though they most likely interact directly with the lipids from the target cell membrane, the mode of action of these molecules is barely known in detail. Previouslyto this project, solid-state NMR experiments carried out in our laboratory showed a fluidification of liposomes when the RLs were added to a model membrane containing ergosterol, which is a fungal-specific membrane sterol. Interestingly, this effect was less remarkable for a stigmasterol-containing plant model and there were no differences in other models containing sitosterol or cholesterol. In this project, we used molecular dynamics (MD) simulations as a complementary approach to those experiments. We focus on all atom (AA) MD simulations with the purpose of giving a better insight of the specific interactions taking place between the RLs and lipids of different plasma membrane models. Multiple parameters have been analyzed, *(i)* effect of different sterols on the lipid dynamics in the presence and the absence of RLs by order parameter measurements, *(ii)* RLs localization within the membranes by means of density profile calculations, *(iii)* RLs interactions with the different membrane components through inter-molecular polar contact distributions.While it is hard to find significant differences in order parameters between different membrane models, same RL-induced fluidization has been obtained. Besides, the formation of specific H-bonds between RLs and some types of sterols could be a starting point to understand how RLs target different types of membranes due to different interaction mechanisms with several membrane components. Coarse grained (CG) MD simulations have also carried out with the goal of simulating more complex and larger fungal membranes in a longer time scale. RL-induced fluidization was also observed by the means of order parameter measurements. Moreover, RLs have shown to affect membrane properties as thickness, membrane area and area per lipid. However, no destabilization of the membrane was observed in the presence of RLs even in longer time scales with bigger models.

VP6
Increasing knowledge of odors and molecular structures linkages of smell compounds by comparing UMAP method to other classification approaches

Marylène Rugard[a], Thomas Jaylet[a], Olivier Taboureau[b], Anne Tromelin[c], Karine Audouze[a]

[a] Université de Paris, T3S, Inserm UMR-S 1124, 75006 Paris, France
[b] Université de Paris, Inserm U1133, CNRS UMR 8251, 75006, Paris, France
[c] Centre des Sciences du Goût et de l'Alimentation, AgroSup Dijon, CNRS, INRAE, Université Bourgogne Franche-Comté, F-21000 Dijon, France

The olfactory perception begins at the olfactory epithelium level with the activation of olfactory receptors (ORs) by the binding of odorants[1]. The olfactory system can discriminate a huge number of odors that would reach 1 trillion[2]. Odor structure relationships in olfaction is a challenging area and a key element in understanding the olfactory system[3-5].
This study aims to highlight the relationships between the structure of smell compounds and their odors. For this purpose, 6038 odorant compounds and their known associated odors (162 odor notes) were compiled. We assessed four dimensional reduction techniques (PCA, MDS, t-SNE and UMAP[6]) and two clustering methods (k-means and agglomerative hierarchical clustering AHC) applied to the molecular structures of these 6038 smell compounds encoded by 1024-bit fingerprints. An analysis of the distribution of odor notes and molecular substructures represented in the different clusters was performed.
The less significant results were obtained using the t-SNE, as well concerning the blurred spatial arrangement of the elements in the 2D-space than the overlapping of clustering partitions obtained by k-means and AHC. The MDS and PCA calculations provided better but average results, except for PCA-AHC for which results were a slightly better. All the results and analyses put forward the precision of UMAP in aggregations of the elements according to the cluster areas that were reflected by the high degree of specificity of odor notes regarding the clusters. Indeed, as UMAP is based on the fact that manifold structure exists in the data, UMAP calculation is able to find these structures in the noise of a dataset which is suitable for data visualization.
The assignment of smell compounds in the 2-two-dimensional space defined by the calculation shows a distribution of odorants into four main areas, each cluster being dominated by few specific odors and chemical structures. The four clusters gather respectively (i) ketones and bicyclic compounds having “balsamic”/“nutty” odor notes; (ii) unsaturated and aromatic compounds carrying “woody” odor; (iii) aldehydes, sulfur compounds and amines with “sulfurous” or “citrus” odors; (iv) esters and long linear carbon chains sharing “fruity”/”fatty” odor notes.
Such association of k-means and AHC clustering with UMAP is the first performed on molecular fingerprints for a dataset related to odors. Therefore, the use of UMAP provides a promising way to improve the understanding of the structure-odor relationships by visualizing high quality embedding of large datasets that were previously unattainable.

1. Buck L, Axel R. A novel multigene family may encode odorant receptors: A molecular basis for odor recognition, Cell, 1991, 175-187.
2. Bushdid C, Magnasco MO, Vosshall LB, Keller A. Humans can discriminate more than 1 trillion olfactory stimuli, Science, 2014, 13701372.
3. Sell CS. The relationship between molecular structure and odour. In: Chemistry and the sense of smell. Oxford: Blackwell Science Publ; 2014. p. 388-419..
4. Genva M, Kemene TK, Deleu M, Lins L, Fauconnier ML. Is it possible to predict the odor of a molecule on the basis of its structure?, Int J Mol Sci, 2019 3018.
5. Licon CC, Bosc G, Sabri M, Mantel M, Fournel A, Bushdid C, et al. Chemical features mining provides new descriptive structure-odor relationships, PLoS Comput Biol, 2019 e1006945.
6. McInnes L, Healy J, Melville J. Umap: Uniform manifold approximation and projection for dimension reduction, arXiv, 2020, 861.

VP7
Understanding Ptch1 dynamics and interactions: from antibioresistance to chemotherapy resistance

M. SIMSIR[1], C. ETCHEBEST[2], R. GAUTIER[1], M. TIBERTI[1], I. MUS-VETEAU[1]

1 Université Côte d'Azur, CNRS, IPMC, 660 Route des Lucioles, 06560 Valbonne, France
2 Université Sorbonne Paris Cité, Université Paris Diderot, Inserm, INTS, Unité Biologie Intégrée du Globule Rouge UMR S1134, DSIMB, Laboratoire d'Excellence GR-Ex, Paris, 75739, France

Resistance to treatment can be found in many diseases. Among the proteins capable of inducing resistance are the RND superfamily which is found throughout all branches of life. Ptch1, a transmembrane protein, receptor of the morphogen Hedgehog, and member of the RNDs, has a cholesterol efflux activity, but has also shown to efflux chemotherapeutic agents and to induce resistance to treatment in cancer cells [1, 2, 3]. Among the common features of multidrug resistance in RND is the ability of these transmembrane proteins to efflux a broad spectrum of substrates and drugs using the proton motive force.
Unlike it's bacterial counterpart, AcrB, which was widely studied, structural data are available on Ptch1 only for few conformation which made challenging the structural characterization of the efflux mechanism of Ptch1 and its inhibition.
With molecular dynamics we highlighted dynamic behaviors of Ptch1 by comparing our results to AcrB key conformational changes. Having structures with very close conformations, we managed to explore a conformational space further away from that of the available structures. Although the time scale of our simulations was not long enough to witness the transport of the cholesterol present in the protein during our simulations, we identified preliminary conformational changes, in particular in the transmembrane domain.
Molecular docking was used in order to identify possible binding sites of transported drugs by Ptch1 and of its inhibitors [4, 5, 6]. The most populated binding site has been found in the center of the protein, where cholesterol was also identified in some of the resolved structures.
As one of the first in silico studies of Ptch1's dynamics, we paved the way towards the understanding of its transport activity and managed to highlight key features for further studies.

[1] M. Bidet, A. Tomico, P. Martin, H. Guizouarn, P. Mollat, and I. Mus-Veteau. The Hedgehog receptor patched functions in multidrug transport and chemotherapy resistance. Molecular Cancer Research, 2012, 1496-1508.
[2] L. Fiorini, M.A Tribalat, L. Sauvard, J. Cazareth, E. Lalli, I. Broutin, O. P Thomas and I. Mus-Veteau. Natural paniceins from mediterranean sponge inhibit the multidrug resistance activity of Patched and increase chemotherapy efficiency on melanoma cells. Oncotarget, 2015, 22282.
[3] A. Hasanovic, C. Ruggiero, S. Jung, I. Rapa, L. Signetti, M. Ben Hadj, M. Terzolo, F. Beuschlein, M. Volante, C. Hantel, E. Lalli and I. Mus-Veteau. Targeting the multidrug transporter Patched potentiates chemotherapy efficiency on adrenocortical carcinoma in vitro and in vivo. International journal of cancer, 2018, 199-211..
[4] L. Signetti, N. Elizarov, M. Simsir, A. Paquet, D. Douguet, F. Labbal, ... & I. Mus-Veteau. "Inhibition of Patched Drug Efflux Increases Vemurafenib Effectiveness against Resistant BrafV600E Melanoma." *Cancers* 12.6 (2020): 1500.
[5] A. Hasanovic, M. Simsir, F.S. Choveau, E. Lalli, & I. Mus-Veteau. "Astemizole Sensitizes Adrenocortical Carcinoma Cells to Doxorubicin by Inhibiting Patched Drug Efflux Activity." *Biomedicines* 8.8 (2020): 251.
[6] N. Durand, M. Simsir, L. Signetti, F. Labbal, R. Ballotti, & I. Mus-Veteau. "Methiothepin Increases Chemotherapy Efficacy against Resistant Melanoma Cells." *Molecules* 26.7 (2021): 1867.

PARTICIPANTS

ACI-SECHE	Samia	ICOA - UMR7311	Orléans	France
ALFERKH	Lina	Laboratoire CiTCoM, CNRS UMR 8038	Paris	France
AMAYA	Diego	INRIA-LORIA	Laxou	France
ANDRE	François	I2BC-CNRS/CEA/Université Paris Saclay	Gif-sur-Yvette	France
ANDREANI	Jessica	CEA I2BC	Gif-sur-Yvette	France
ANDRE-LEROUX	Gwenaëlle	MaIAGE, INRAE	Jouy-en-Josas	France
ANNEREAU	Jean-Philippe	Medetia	Paris	France
AOUIDATE	Adnane	ICOA UMR 7311	Orléans	France
ARAUJO-ROCHA	Mario	ITODYS CNRS UMR 7086	Paris	France
AUFFINGER	Pascal	UPR9002/CNRS	Strasbourg	France
AVSTRIKOVA	Mariia	Université de Strasbourg, CNRS	Strasbourg	France
BARBAULT	Florent	ITODYS CNRS UMR 7086	Paris	France
BARBE	Sophie	INSA, UMR CNRS 5504 /INRA 792	Toulouse	France
BARTOCCI	Alessio	Laboratoire des Fonctions Moleculaires	Strasbourg	France
BAUD	Stéphanie	CNRS UMR 7369 MEDyC	Reims	France
BEDART	Corentin	U1286 - INFINITE	Lille	France
BELLAICHE	Adam	Université de Paris	Paris	France
BELLOY	Nicolas	CNRS UMR 7369 MEDyC	Reims	France
BENKAIDALI	Lydia	ICOA	Orleans	France
BLANC	Florian	Theoretical Biophysics Department, Max Planck Institute of Biophysics	Frankfurt am Main	Germany
BONACHERA	Fanny	CNRS	Strasbourg	France
BONNET	Pascal	ICOA - UMR7311	Orléans	France
BORT	William	University of Strasbourg	Strasbourg	France
BOUCHIBA	Younes	INSA TBI	Toulouse	France
BOULARD	Yves	CEA/DRF/I2BC	Gif-sur-Yvette	France
BOURG	Stéphane	ICOA - UMR7311	Orléans	France
BRET	Guillaume	LIT - UMR7200 (Université de Strasbourg - CNRS)	Illkirch-Graffenstaden	France
BRET	Hélène	CEA/DRF/I2BC	Gif-sur-Yvette	France
BRISENO-ROA	Luis	Medetia	Paris	France
BUREAU	Ronan	CERMN, Université de Caen Normandie	Caen	France
CAMPROUX	Anne	BFA - CNRS UMR8251- INSERM U1133	Paris	France
CARVAILLO	Jean-Charles	CNRS UMR 9198 (I2BC)	Gif-Sur-Yvette	France
CECCHINI	Marco	UMR7177 -- CNRS	Strasbourg	France
CHALOIN	Laurent	IRIM - CNRS - University of Montpellier	Montepellier	France
CHAVATTE	Philippe	U1286 - INFINITE	Lille	France
CHAVENT	Matthieu	IPBS / CNRS	Toulouse	France
CHEDIK	Lisa	INRS	Vandoeuvre-lès-Nancy	France
CHIESA	Luca	Laboratoire d'Innovation Thérapeutique	Illkirch-Graffenstaden	France
COLLIANDRE	Lionel	Evotec AG	Toulouse	France

COLLOC'H	Nathalie	CNRS - SCTD	Vandoeuvre-lès-Nancy	France
CONG	Xiaojing	Institut de génomique fonctionnelle	Montpellier	France
CRETIN	Gabriel	INTS / Université de Paris	Paris	France
CROWET	Jean-Marc	CNRS UMR 7369 MEDyC	Reims	France
CUISSARD	Bertrand	CNRS GREYC UMR6072	Caen	France
CZAPLICKI	Georges	IPBS-CNRS	Toulouse	France
DAFNIET	Bryan	BFA - CNRS UMR8251- INSERM U1133	Paris	France
DAUCHEZ	Manuel	CNRS UMR 7369 MEDyC	Reims	France
DE BEAUCHENE	Isaure	CNRS, LORIA	Nancy	France
DE BREVERN	Alexandre G.	INSERM DRPA07	Paris	France
DE MALGLAIVE	France	Medetia	Paris	France
DE RUYCK	Jérôme	UGSF - UMR8576	Villeneuve d'Ascq	France
DEJAEGERE	Annick	IGBMC Gie CERBM	Illkirch-Graffenstaden	France
DELALANDE	Olivier		Rennes	France
DELAUNAY	Maxence	Université Paris-Saclay	Châtenay-Malabry	France
DEPENVEILLER	Camille	CNRS UMR 7369 MEDyC	Reims	France
DESSAUX	Delphine	TBI - INSA Toulouse	Toulouse	France
DETROYER	Ann	L'Oréal, Research & Innovation	Aulnay-sous-Bois	France
DEVIGNES	Marie-Dominique	CNRS, LORIA	Vandoeuvre-lès-Nancy	France
DIHARCE	Julien	UMRS-1134 EQUIPE 2 DSIMB INTS	Paris	France
DION	Charles	Iktos	Paris	France
DOUGUET	Dominique	GDR 2066 - IPMC CNRS	Valbonne	France
DRIZARD	Nicolas	Iktos	Paris	France
DUDAS	Balint	CiTCoM - UMR8038	Paris	France
EFREMOV	Roman	M.M. Shemyakin & Yu.A. Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences	Moscow	Russia
EGUIDA	Merveille		Illkirch-Graffenstaden	France
ETCHEBEST	Catherine	UMRS-1134 EQUIPE 2 DSIMB INTS	Paris	France
FAGNEN	Charline	CERMN - Université de Caen	Caen	France
FARCE	Amaury	U1286 - INFINITE	Lille	France
FEREY	Nicolas	LISN - CNRS - Université Paris Saclay	Orsay	France
FIORUCCI	Sébastien	UMR 7272 CRNS	Nice	France
FLATTERS	Delphine	BFA - CNRS UMR8251- INSERM U1133	Paris	France
GAILLARD	Thomas	Ecole Polytechnique	Palaiseau	France
GALENTINO	Katia	Université de Strasbourg; CNRS	Strasbourg	France
GALOCHKINA	Tatiana	INTS / Université de Paris	Paris	France
GANTZER	Philippe	IFP Energies nouvelles	Rueil-Malmaison	France
GELLY	Jean-Christophe	UMRS-1134 EQUIPE 2 DSIMB INTS	Paris	France

GESLIN	Damien	Cermn, Unicaen	Caen	France
GHEERAERT	Aria	Université Savoie Mont-Blanc	Chambéry	France
GHOULA	Mariem	Paris Diderot University	Paris	France
GIOVANNINI	Johanna	Centre d'Etudes et de Recherche sur le Médicament de Normandie	Caen	France
GONZALEZ ALEMAN	Roy	I2BC UMR9198 CNRS - CEA - Univ. Paris Saclay	Gif-sur-Yvette	France
GUILLAUME	Valentin	Institut Pasteur de Lille	Lille	France
HAMMAMI	Islem	CNRS-SCTD	Vandoeuvre-lès-Nancy	France
HERNANDEZ ALBA	Oscar	CNRS UMR 7178	Strasbourg	France
HLADIS	Matej	UMR 7272 CNRS - Université Côte d'Azur	Nice	France
HOFFMANN	Brice	Iktos	Paris	France
HORVATH	Dragos	Laboratoire de Chemoinformatique, UMR 7140	Strasbourg	France
IORGA	Bogdan	CNRS-ICSN	Gif-sur-Yvette	France
JACQUEMARD	Célien	Laboratoire d'Innovation Thérapeutique	Illkirch-Graffenstaden	France
JUSOT	Maud	Iktos	Paris	France
KARAMI	Yasaman	Institut Pasteur	Paris	France
KELLENBERGER	Esther	Laboratoire d'Innovation Thérapeutique UMR7200	Illkirch-Graffenstaden	France
KHAKZAD	Hamed	INSERM - DRPA11	Paris	France
KLIMCHUK	Olga	University of Strasbourg	Strasbourg	France
KOCH	Mathilde	Arkema	Pierre Bénite	France
KOUKEBENE FOGHA	Jade	INSERM DRPA07	Paris	France
KREZEL	Pascal	ICOA	Orléans	France
LABOUILLE	Stéphanie	Iktos	Paris	France
LAGARDE	Nathalie	Laboratoire GBCM	Paris	France
LAMBERT	Roman	Apoptosis Research Center, NUI Galway	Galway	Ireland
LAMOTTE	Jean-Luc	CERMN - UNICAEN	Caen	France
LANGENFELD	Florent	Laboratoire GBCM - CNAM	Paris	France
LAURENT	Adèle	CEISAM UMR 6230	Nantes	France
LAURENT	Xavier	Discngine	Paris	France
LAVILLE	Pierre	Université de Paris, BFA CNRS UMR 8251, CMPLI Inserm U1133	Paris	France
LE GUILLOUX	Vincent	Discngine	Paris	France
LECLERC	Fabrice	I2BC UMR 9198 CNRS - CEA - Univ. Paris Saclay	Gif-sur-Yvette	France
LEHEMBRE	Etienne	CNRS GREYC UMR6072	Caen	France
LEPAILLEUR	Alban	CERMN - Université Caen Normandie	Caen	France
LESCANNE	Camille	Medetia	Paris	France
LIBOUBAN	Pierre-Yves	Institut de chimie organique et analytique (ICOA)	Orléans	France
MANGUIN	Jean-Luc	CNRS GREYC UMR6072	Caen	France
MAO	Xiaojun	Institut de Chimie Organique et Analytique - ICOA UMR 7311	Orléans	France
MARCOU	Gilles	Université de Strasbourg	Strasbourg	France

MAROUN	R. Charbel	SABNP, INSERM/UEVE	Evry	France
MEJRI	Alia	Laboratoire de Nanomédecine Imagerie et Thérapeutiques	Besançon	France
MIAS-LUCQUIN	Dominique	LORIA	Vandoeuvre-lès-Nancy	France
MILAN RODRIGUEZ	Paula	Laboratoire des Biomolecules (Sorbonne Universite)	Paris	France
MINOLETTI-HOCHEPIED	Claire	Sanofi	Chilly-Mazarin	France
MITEVA	Maria	INSERM ERL U1268 MCTR, CitCom UMR 8038 CNRS	Paris	France
MOLZA	Anne-Elisabeth	INSERM UMR_S 1134, DSIMB	Saint Denis	France
MONIOT	Antoine	Université de Lorraine	Vandoeuvre-lès-Nancy	France
MONTES	Matthieu	Laboratoire GBCM	Paris	France
MOROY	Gautier	Université de Paris, BFA, UMR 8251, CNRS, ERL U1133, Inserm	Paris	France
MULLER	Christophe	Evotec	Toulouse	France
NACERI	Sarah		Paris	France
NEDEV	Hristo	ICSN/CNRS	Gif sur Yvette	France
OUALI	Abdelkader	CNRS GREYC UMR6072	Caen	France
PACALON	Jody	Université Côte D'Azur	Nice	France
PANEL	Nicolas	Uppsala Biomedicinska Centrum (BMC)	Uppsala	Sweden
PATIENT	Guillaume	U1286 - INFINITE	Lille	France
PETITJEAN	Michel	BFA - CNRS UMR8251- INSERM U1133	Paris	France
PEYRAT	Gautier	Institut de Chimie Organique et Analytique	Orléans	France
PEZESHKIAN	Weria	Faculty of Science and Engineering, Molecular Dynamics, Biomolecular Sciences and Biotechnology	Groningen	The Netherlands
PINEL	Philippe	Iktos	Paris	France
PLYER	Louis	Université de Strasbourg	Strasbourg	France
QUIGNOT	Chloé	Institut Curie	Paris	France
RAGUI	Marc	CERMN laboratory (University of Caen)	Caen	France
RAO	Rajas	Université Reims Champagne-Ardenne	Reims	France
RAYAR	Anita	Medetia	Paris	France
REAU	Manon	Utrecht University	Utrecht	The Netherlands
REHIOUI	Hajar	CNRS GREYC UMR6072	Caen	France
RENAULT	Nicolas	U1286 - INFINITE	Lille	France
REVILLO IMBERNON	Julia	University of Strasbourg	Illkirch-Graffenstaden	France
RODRIGUEZ	Julia	CNRS	Illkirch-Graffenstaden	France
RODRIGUEZ-MORAGA	Nely	GEC UMR7025 CNRS, Université de Picardie Jules Verne	Amiens	France
ROGNAN	Didier	CNRS	Illkirch-Graffenstaden	France
RUGARD	Marylène	INSERM UMRS 1124 T3S	Paris	France

SACQUIN MORA	Sophie	Laboratoire de Biochimie Théorique, CNRS UPR9080	Paris	France
SELLAMI	Asma	Conservatoire national des arts et métiers	Paris	France
SINGH	Raghvendra Pratap	UMET/INRAE, University of Lille	Villeneuve d'Ascq	France
SIRUGUE	Lea	Conservatoire National des Arts et Metiers	Paris	France
SOPKOVA-DE OLIVEIRA SANTOS	Jana	CERMN, University of Caen Normandy	Caen	France
STOTE	Roland	IGBMC Gie CERBM	Illkirch-Graffenstaden	France
TABOUREAU	Olivier	Modélisation Computationnelle des interactions protéine-ligand (U1133), Unité de Biologie Fonctionnelle et Adaptative (UMR 8251), Université de Paris	Paris	France
TALY	Antoine	CNRS	Paris	France
TELLES DE SOUZA	Paulo Cesar	Molecular Microbiology and Structural Biochemistry (MMSB, UMR 5086), CNRS, University of Lyon	Lyon	France
TOPIN	Jérémie	Université Côte d'Azur	Nice	France
TRAN NGUYEN	Viet Khoa	Laboratoire d'Innovation Thérapeutique	Illkirch-Graffenstaden	France
TROMELIN	Anne	INRAE-CSGA	Dijon	France
TUBIANA	Thibault	Department of Chemistry, University of Bergen, Norway	Bergen	Norway
TURK	Joseph-André	Iktos	Paris	France
VARNEK	Alexandre	University of Strasbourg	Strasbourg	France
VELUSAMY	Mahesh	CNRS UMR 7369 MEDyC	Reims	France
VERSINI	Raphaëlle	Laboratoire de Biochimie Théorique, CNRS, Université de Paris, UPR 9080	Paris	France
VOLKOV	Mikhail	University of Strasbourg	Illkirch-Graffenstaden	France
VUCINIC	Jelena	IRIT/CRCT	Toulouse	France
WONG	Hua	CNRS UMR 7369 MEDyC	Reims	France
YACOUB	Taher	I2BC UMR9198 CNRS - CEA - Univ. Paris Saclay	Gif-sur-Yvette	France
ZABOLOTNA	Yuliana	University of Strasbourg	Strasbourg	France
ZIMMERMANN	Albrecht	CNRS GREYC UMR6072	Caen	France

POSTERS LIST

PP-1	AMAYA-RAMIREZ Diego	Leveraging Zernike polynomials for peptide conformer identification
PP-2	ANDRE-LEROUX Gwenaelle	The structure of the universal bacterial DNA repair protein Mfd dictates the pathogenicity of Bacillus cereus strains
PP-3	AOUIDATE Adnane	Identification of New Small Molecules Antiviral Agents as Protein-Protein Inhibitors of the ACE2 : Coronavirus Spike Interactions to Prevent the Entry of SARS-CoV-2 to the host Cell
PP-4	AVSTRIKOVA Maria	Allosteric modulation of the nicotinic acetylcholine receptor α7 by ivermectin
PP-5	BARBAULT Florent	DockSurf: a molecular modelling software for predicting protein/surface association
PP-6	BEDART Corentin	SINAPs: A software tool for analysis and visualization of interaction networks of molecular dynamics simulations
PP-7	BORT William	Discovery of novel chemical reactions by Deep Generative Recurrent Neural Network
PP-8	BOUYSSET Cedric	Integrative structural modeling reveals functional molecular switches of human G protein-coupled bitter-taste receptors
PP-9	CARVAILLO Jean-Charles	Molecular modelling of hepatitis B virus Core protein C-terminal domain: Implications for nucleocapsid trafficking
PP-10	CHALOIN Laurent	Small-molecule inhibitors of SARS-CoV-2 interfering with Spike / ACE2 interactions
PP-11	CHAVENT Matthieu	Sugar stay or sugar go: role of the sugar moiety on the biophysical properties of mycobacterial mycocerosate containing lipids
PP-12	COLLIANDRE Lionel	Application of Bayesian Optimization in Drug Discovery
PP-13	CRETIN Gabriel	PYTHIA: Deep Learning Approach For Local Protein Conformation Prediction
PP-14	CROWET Jean-Marc	LIMONADA: a database dedicated to the simulation of complex lipid membranes
PP-15	DAFNIET Bryan	Development of a chemogenomics library for phenotypic screening
PP-16	DE BREVERN Alexandre	Investigate the impact of PTMs on the protein backbone conformation
PP-17	DELAUNAY Maxence	Des3PI: a Computational Fragment-based Approach to Design Peptides Targeting Protein-Protein Interactions
PP-18	SOPKOVA-DE OLIVEIRA SANTOS Jana	Insights into the human 5-HT4 receptor binding site: exploration of activation conformational changes
PP-19	DEPENVEILLER Camille	Simulating large extracellular matrix molecules as dynamic chains of rigid bodies
PP-20	DESSAUX Delphine	Computational design of enzymes for cell-free biocatalysis
PP-21	DIHARCE Julien	Molecular mechanisms governing oligomerization of Translocator Protein (TSPO) and their role in cholesterol translocation highlighted by coarse-grained approaches
PP-22	DOUGUET Dominique	SenSaaS: Shape□based Alignment by Registration of Colored Point□based Surfaces
PP-23	DRIZARD Nicolas	A Molecular Bio-Assays Simulator to Unravel Predictors Hacking during Optimisation

PP-24	DRIZARD Nicolas	Virtual Screening of Large Scale Libraries Guided with Docking and Deep Learning
PP-25	FEREY Nicolas	Interface tangible modulaire, articulée et sans marqueur dédiée à la pédagogie et à la recherche en biologie moléculaire
PP-26	FOGHA Jade	Evolution, structure and dynamics of IL-3 and IL-3R alpha interaction
PP-27	GALENTINO Katia	Rational design of allosteric modulators of biomolecular motors
PP-28	GANTZER Philippe	New Tool for the Comparison of Molecular Structures Generations
PP-29	GESLIN Damien	Visualization of the cluster-based partitioning of active and inactive pharmacophores. Application to the BCR-ABL pharmacophore network
PP-30	GONZALEZ-ALEMAN Roy	MCSS-Based Predictions of Binding Mode and Selectivity of Nucleotide Ligands
PP-31	HAMMAMI Islem	Identification of biological targets for argon
PP-32	JACQUEMARD Celien	ATOLL: A visualization tool to compare transmembrane domains structures
PP-33	KARAMI Yasaman	Structural determination of PTPN4 using SAXS restrained Molecular Dynamics
PP-34	LABOUILLE Stéphanie	Navigating the chemical space using structure-based guided generative modelling
PP-35	LEHEMBRE Etienne	Towards lattice-based interactive pharmacophore exploration
PP-36	LESCANNE Camille	Comparison between homology-based and experimentally determined structures on prostanoid GPCR receptor
PP-37	LIBOUBAN Pierre-Yves	Protein-Ligand binding affinity prediction using combined molecular dynamics and deep learning approaches
PP-38	MARIA Maxime	VTX: High-performance molecular structure and dynamics visualization software
PP-39	MIAS-LUCQUIN Dominique	An homodimeric multi-domain protein interacting with ssDNA and dsDNA : The challenges of RelSt3 relaxase modelisation
PP-40	MOLZA Anne-Elisabeth	in silico investigation of dynamical and structural properties of the major efflux pump of P. aeruginosa: MexBA-OprM
PP-41	NACERI Sarah	Structural analysis of interaction between SARS-CoV-2 spike protein and the human ACE2 receptor
PP-42	PACALON Jody	Insect odorant co-receptor ORco agonists binding mechanism
PP-43	PANEL Nicolas	Design of GPCR agonists by free energy calculations
PP-44	PINEL Philippe	Identification and analysis of scaffold hopping situations using protein-ligand interaction fingerprints
PP-45	PLYER Louis	Implementation of a soft grading system for chemistry in a Moodle plugin
PP-46	RAGUI Marc	The Design of non-peptide and selective ligands acting as XIAP inhibitors

PP-47	RAO Rajas	Roles of glycans in dynamics of insulin receptor
PP-48	REAU Manon	Deeprank-GNN: A Graph Neural Network Framework to Learn Interaction Patterns from Protein-Protein Interfaces
PP-49	REVILLO IMBERNON Julia	A survey of the fragment space. Commercial libraries
PP-50	SIRUGUE Lea	A system to compare shapes based on a Wave Kernel Signature Map applied to protein surfaces
PP-51	TOPIN Jeremy	With or without sugar? A search for lectin inhibitors
PP-52	TRAN-NGUYEN Viet-Khoa	LIT-PCBA: An Unbiased Data Set for Machine Learning and Virtual Screening
PP-53	VELUSAMY Mahesh	Modélisation par homologie d'anelylyl cyclases et de leurs interactions
PP-54	VERSINI Raphaelle	Structural basis of the transmembrane domain of the yeast mitofusin
PP-55	VOLKOV Mikhail	Applicability of graph neural networks to predict binding affinities from protein-ligand structures
PP-56	WONG Hua	Multiscale modelling of the Extracellular Matrix (ECM)
PP-57	YACOUB Taher	Design of RNA oligonucleotides against the Beta-Secretase 1 (BACE1) enzyme involved in Alzheimer's disease

PV-1	GAILLARD Thomas	Evaluation of AutoDock and AutoDock Vina on the CASF-2013 benchmark
PV-2	GIOVANNINI Johanna	Conception of 3D models of reduced models of Tau protein aggregates for conception and synthesis of potential Tau aggregation disruptors
PV-3	KAROUI Samiha	Study by the DFT of the structural and electronic properties of Isoquercetin
PV-4	MEJRI Alia	Role of the geometric characteristics on the carbon nanotube conductance: a molecular dynamic simulation study
PV-5	RODRIGUEZ-MORAGA Nely	Contributions of molecular dynamics simulations to the understanding of rhamnolipid interactions with biomimetic fungal membranes
PV-6	RUGARD Marylene	Increasing knowledge of odors and molecular structures linkages of smell compounds by comparing UMAP method to other classification approaches
PV-7	SIMSIR Meline	Understanding Ptch1 dynamics and interactions: from antibioresistance to chemotherapy resistance

	Wednesday 29/09/2021			Thursday 30/09/2021			Friday 01/10/2021		
8h									
	ID	Name		S3	Session #3	ARAUJO-ROCHA Mario			
9h	O	Organisation & General				AUFFINGER Pascal	S5	Session #5	HOFFMANN Brice
	S1	Visualization and graphism			Keynote Lecture #4	BARBE Sophie			HLADIS Matej
	S2	Simulation of biosystems					S6	Keynote Lecture #7	HORVATH Dragos
	S3	Methodological developments							
10h	S4	Integrative modeling			Flash posters - GGMM #2 - Session B			Coffee break	
	S5	Data-driven drug discovery			Poster session & Coffee break				
	S6	Current topics in chemoinformatics							PEYRAT Gautier
11h		SFCi			Session #3	BOUCHIBA Younes		Session #6	REHIOUI Hajar
						TUBIANA Thibault			SELLAMI Asma
						MILAN-RODRIGUEZ Paula			TELLES de SOUZA Paulo
12h	O	Reception & Opening of "Journées GGMM"		O	Opening of "Journées SFCI" & Lunch break				TURK Joseph-André
							O	GGMM award conference SFCi award conference Poster awards & Closing	
13h		Welcome introduction						General assemblies : GGMM, SFCI, GDR BigDataChim & Lunchbox	
14h	S1	Keynote Lecture #1	PEZESHKIAN Weria	S4	Keynote Lecture #5	MITEVA Maria			
		Session #1	GALOCHKINA Tatiana		Flash posters - SFCI #1 - Session C				
15h			GONZALEZ-ALEMAN Roy		Session #4	KHAKZAD Hamed			
			BEDART Corentin			ZABOLOTNA Yuliana			
			LANGENFELD Florent			ALFERKH Lina			
			GELLY Jean-Christophe			DUDAS Balint			
16h		Flash posters - GGMM #1 - Session A				SACQUIN-MORA Sophie			
		Poster session & Coffee break			Flash posters - SFCI #2 & PV - Session D				
					Poster session & Coffee break				
17h	S2	Keynote Lecture #2	DE RUYCK Jérôme	S5	Keynote Lecture #6	DETROYER Ann			
		Session #2	BLANC Florian						
18h			FAGNEN Charline		Session #5	BENKAIDALI Lydia			
			BARTOCCI Alessio			EGUIDA Merveille			
			GHOULA Mariem			GHEERAERT Aria			
			BELLAICHE Adam						
19h	O	Dinner		O	Break				
					Extra dinner				