YMM 2023
Young Modellers Conference

15-17 mai • Toulouse

Pr. Siewert Jan Marrink
Groningen Biomolecular Sciences
and Biotechnology Institute and Zernike Institute
for Advanced Materials,
University of Groningen, Netherlands

Pr. Lucie Delemotte
Department of Applied Physics,
Science for Life Laboratory,
Stockholm, Sweden

Pr. Jean-Philip Piquemal
Laboratoire de Chimie Théorique (LCT),
Sorbonne Université, Paris, France

Dr. Noelia Ferruz
Department of Structural and Molecular Biology,
Molecular Biology Institute of Barcelona, Spain

Pr. Elisa Fadda
Department of Chemistry and Hamilton Institute,
Maynooth University, Ireland

cecam · LAAS CNRS · Qubit · INSA · GDR · CARBIOS · evotec · ANITI · tbi · OneAngstrom · Schrödinger

Dear colleagues and friends,

On behalf of all the members of the board, scientific and organizing committees, I am delighted to welcome you to Toulouse for the 23rd congress of the Molecular Graphics and Modelling Group, where we will celebrate the 40th anniversary of our learned society.

This congress undoubtedly demonstrates the dynamism and vitality of our association. This year, we wanted to open the conference to an international audience, and it is already a success as we will be welcoming nearly 200 participants from 17 different countries, and the lineup of invited speakers is exceptional. We would like to thank Ms. Lucie Delemotte (Stockholm), Noelia Ferruz (Barcelona), and Elisa Fadda (Maynooth), as well as Mr. Siewert-Jan Marrink (Groningen) and Jean-Philippe Piquemal (Paris) for agreeing to share their expertise with us.

Our habits are not lost, we will be gathered to meet, work, dine, and exchange in a friendly and enjoyable atmosphere over the course of three days. The program is dense with 5 plenary lectures, 30 oral communications, and 2 poster sessions, but the topics covered will provide a comprehensive overview of the different fields of molecular modeling. We will also have a roundtable discussion on the opportunities available to young modelers *beyond academic careers.*

Our congress is by definition multidisciplinary and brings together experts from biology, biochemistry, medicine, chemistry, physics, computer science, mathematics, and statistics, providing them with a unique event to meet every two years. We will have the opportunity to benefit from the latest research and innovations in our field and to discuss the challenges and opportunities we face in our daily practice.

Finally, we will have the pleasure of hearing from the two winners of the GGMM Prize, Merveille Eguida and Jelena Vucinic, who were selected by the GGMM board.

At the end of this congress, the general assembly of the association will be held with the presentation of the moral and financial report, as well as a review of the analysis of responses to the GGMM questionnaire distributed last year to take stock of our community.

I wish you all an excellent congress and hope you will enjoy your stay here with us.

Sébastien Fiorucci,
President of GGMM

Organisation comittee

Isabelle André (TBI, INSA-CNRS-INRAE, Toulouse, France)
Sophie Barbe (TBI, INSA-CNRS-INRAE, Toulouse, France)
Marie Brut (LAAS-CNRS, Toulouse, France)
Matthieu Chavent (IPBS, CNRS, Toulouse, France)
Juan Cortes (LAAS-CNRS, Toulouse, France)
Nicolas Destainville (LPT, Toulouse, France)
Jérémy Esque (TBI, INSA-CNRS-INRAE, Toulouse, France)
Friederike Schmid (Johannes Gutenberg University, Mainz, Germany)
Lukas Stelzl (Johannes Gutenberg University, Mainz, Germany)

Scientific comittee

Gwenaelle André (Maiage, INRAE, Jouy-en-Josas)
Isabelle André (TBI, INSA-CNRS-INRAE, Toulouse, France)
Marc Baaden (LBT, IBPC, CNRS, Paris)
Sophie Barbe (TBI, INSA-CNRS-INRAE, Toulouse, France)
Matthieu Chavent (IPBS, CNRS, Toulouse)
Annick Dejaegere (Université de Strasbourg, Strasbourg)
Jérémy Esque (TBI, INSA-CNRS-INRAE, Toulouse, France)
Sebastien Fiorucci (Université Côte d'Azur, Nice)
Friederike Schmid (Johannes Gutenberg University, Mainz, Germany)
Lukas Stelzl (Johannes Gutenberg University, Mainz, Germany)

We would like to thank our Sponsors

Useful Information

CONFERENCE LOCATION

the GGMM2023 conference will be held at the
Mercure Toulouse Centre Compans hotel
Boulevard Lascrosses, 8 Esplanade Compans Caffarelli
31000 Toulouse

HOTELS

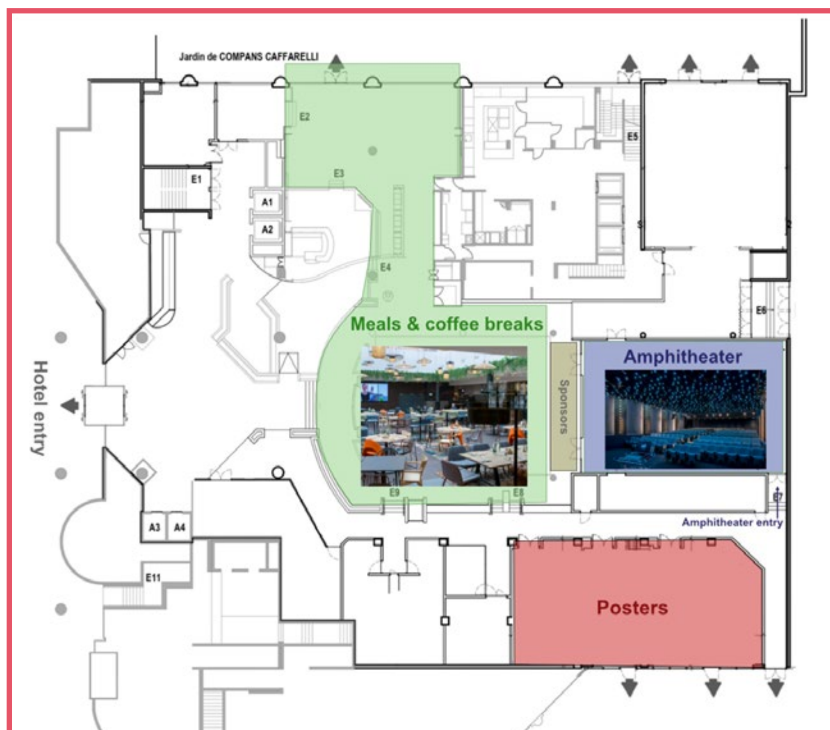
The bedrooms will be split between two hotels:

hotel Mercure Compans Caffarelli

Novotel Compans Caffarelli

The two hotels are located 5 minutes away on foot (see the map opposite).

MERCURE HOTEL MAP



This conference will be articulated
around 5 topics

MULTISCALE MODELLING

Modelling approaches combining different resolutions (QM, AT, CG, and beyond) applied on biological systems (protein, lipids nucleic acids, ..) and their interactions.

INTEGRATIVE APPROACHES

Approaches coupling modelling with data from bioinformatics (co-evolution, seq. alignments, ...) or from experiments (structural biology, biophysics, ...).

METHODOLOGICAL DEVELOPMENTS

Methodological developments to complete or go beyond MD simulations with a special interest to graphics, AI, and tools developments.

MOLECULAR AND DRUG DESIGN

Modelling approaches for the design of proteins (enzymes, ..) or small molecules, drug discovery and screening.

MOLECULAR INTERACTIONS

Modelling biological complexes: protein-protein, protein-ligands, nucleic acids, sugars, ...

Program

- MULTISCALE MODELLING
- INTEGRATIVE APPROACHES
- METHODOLOGICAL DEVELOPMENTS
- MOLECULAR AND DRUG DESIGN
- MOLECULAR INTERACTIONS

Mon. 15/05

Time	Session
12h30-14h00	Participants arrival
14h00-14h15	**M. Chavent (W)**
14h15-15h00	**Pr. SJ Marrink**
15h00-15h15	talk 1
15h15-15h30	talk 2
15h30-15h45	talk 3
15h45-16h00	talk 4
16h00-16h30	Coffe Break
16h30-16h45	talk 5
16h45-17h00	talk 6
17h00-17h45	**Pr. L Delemotte**
17h45-18h00	talk 7
18h00-20h00	**Poster session 1**
From 20h00	Dinner

Tue. 16/05

Time	Session
09h00-09h15	talk 8
09h15-09h30	talk 9
09h30-09h45	talk 10
09h45-10h00	talk 11
10h00-10h15	talk 12
10h15-10h45	Coffe Break
10h45-11h30	**Pr. JP Piquemal**
11h30-11h45	talk 13
11h45-12h00	talk 14
12h00-12h15	talk 15
12h15-12h30	talk 16
12h30-14h00	Lunch
14h00-14h15	talk 17
14h15-14h30	talk 18
14h30-15h15	**Dr. N Ferruz**
15h15-15h30	talk 19
15h30-15h45	talk 20
15h45-16h00	talk 21
16h00-16h30	Coffe Break
16h30-16h45	talk 22
16h45-17h00	talk 23
17h00-17h15	talk 24
17h15-18h00	**Round Table**
18h00-20h00	**Poster session 2**
From 20h00	Gala Dinner

Wed. 17/05

Time	Session
09h00-09h45	**Pr. E Fadda**
09h45-10h00	talk 25
10h00-10h15	talk 26
10h15-10h45	Coffe Break
10h45-11h00	talk 27
11h00-11h15	talk 28
11h15-11h30	talk 29
11h30-11h45	talk 30
11h45-12h40	**GGMM award talks and poster award**
12h40-13h40	GGMM AG
12h40-14h00	Lunch and departure

Pr. Siewert Jan Marrink

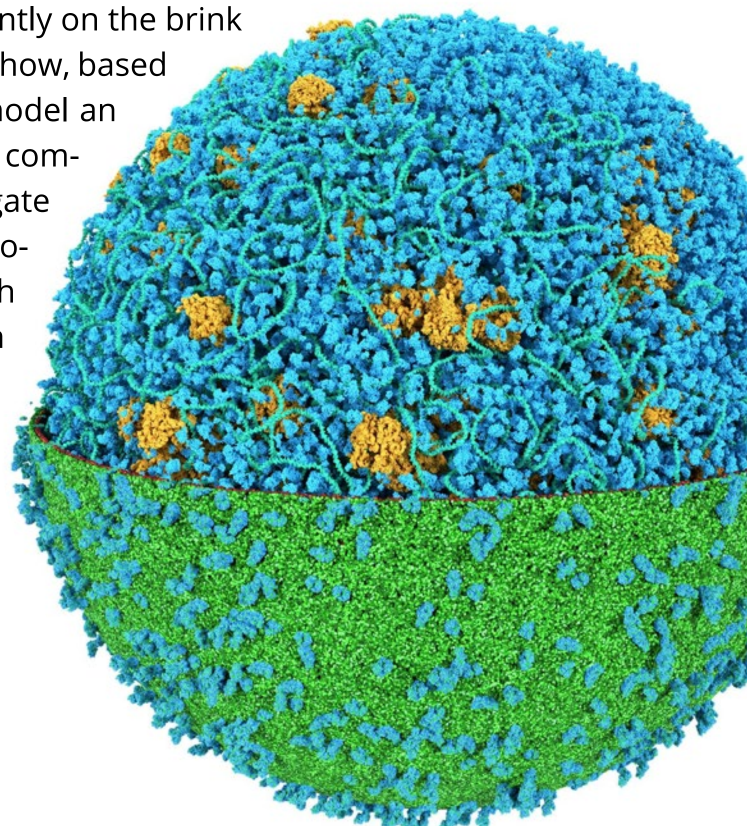
Groningen Biomolecular Sciences
and Biotechnology Institute and Zernike Institute
for Advanced Materials,
University of Groningen, Netherlands

@CG_Martini

cgmartini.nl

SIMULATING ENTIRE CELLS WITH THE MARTINI ECOSYSTEM

The ultimate microscope, directed at a cell, would reveal the dynamics of all the cell's components with atomic resolution. In contrast to their real-world counterparts, computational microscopes are currently on the brink of meeting this challenge. In this talk, I show how, based on the Martini ecosystem, we are able to model an entire cell, the JCVI-syn3A minimal cell, at full complexity. This step opens the way to interrogate the cell's spatio-temporal evolution with molecular dynamics simulations, an approach that can be extended to other cell types in the near future.



NOTES

TALK N°1

COARSE-GRAINED SIMULATIONS: INSIGHT INTO SUPRAMOLECULAR ORGANIZATION OF POLYMER PRODRUGS AND THERMO-RESPONSIVE POLYMERS FOR DRUG ENCAPSULATION

PING GAO[a,b], JULIEN NICOLAS[a], AND TÂP HA-DUONG[b]

[a] BioCIS, UMR CNRS 8076, Faculté de Pharmacie, Université Paris-Saclay, France

[b] Institut Galien Paris-Saclay, CNRS UMR 8612, Faculté de Pharmacie, Université Paris-Saclay, France

Contact: ping.gao@universite-paris-saclay.fr

Traditional drug therapies suffer from poor biocompatibility and pharmacokinetics, resulting in low efficacy and severe side effects. The emergence of polymer nanoparticles offers new ideas for efficient drug encapsulation with enhanced controlled release[1]. However, the success rate of these nanomedicines in clinical trials is extremely low (~5%), with uncontrolled "burst" release and poor drug delivery capacity being a possible trigger. With these current problems in mind, the search for an ideal drug delivery system has become a long-term objective. Coarse-grained molecular dynamics (CGMD) simulations have been introduced as a powerful tool to provide valuable information about the supramolecular arrangement of the polymer chains within nanoparticles and may be of considerable interest to better understand the drug release kinetics.

Two strategies can be followed to encapsulate a drug into polymer nanoparticles, either by chemically linking the drug to a self-assembling polymer chain (so called polymer prodrug) or by physically co-aggregate drugs with self-assembling polymers (see Figure 1). We first introduce CG model of polyisoprene-based prodrugs in which either a gemcitabine (Gem) or a paclitaxel (Ptx) drug is attached to the polymer by using either a propanoate or a diglycolate linker. Our CGMD simulations provided crucial information concerning the spatial organization of the different components (drug, linker, polymer) of the nanoparticles and revealed that the linker cleavage sites were poorly accessible to the solvent[2]. Our theoretical study was further applied to optimized polymer prodrugs with a more suitable linker (diglycolate-TEG). Importantly, our predictions were supported by drug release and in vitro cytotoxicity experiments, which demonstrates that our CG models can guide the design of new polymer prodrug nanoparticles with improved drug release efficiency.

Meanwhile, CGMD simulations were applied to vinyl copolymers with upper critical solution temperature (UCST) property in order to assess the CG model capability to reproduce their liquid-liquid phase separation. The theoretical transition temperature was found with a discrepancy smaller than 10 K to experimental measurements[3]. CGMD simulations of either Gem or Ptx were also performed in presence of the vinyl copolymers to further study their encapsulation and drug-loading efficiency. Gem showed extremely high-water exposure and exhibited «burst release» with increasing temperature. Ptx were rarely accessible to solvent due to drug aggregation and complete burial in the nanoparticle core, suggesting poor Ptx release efficiency. These results highly support that coarse-grained modeling can be used as a general and predictive tool for drug delivery applications.

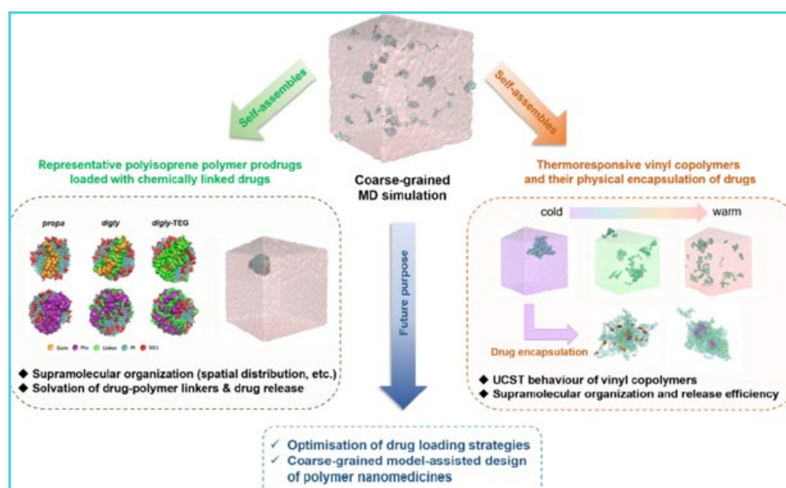


Figure 1: Coarse-grained simulations of polymer prodrug (left) and physical-encapsulated drug (right) delivery nanoparticles.

References

[1] Nicolas, J.; Mura, S.; Brambilla, D.; Mackiewicz, N.; Couvreur, P. Design, functionalization strategies and biomedical applications of targeted biodegradable/biocompatible polymer-based nanocarriers for drug delivery. Chem. Soc. Rev. 2013, 42, 1147.

[2] Gao, P.; Nicolas, J.; Ha-Duong, T., Supramolecular Organization of Polymer Prodrug Nanoparticles Revealed by Coarse-Grained Simulations, J. Am. Chem. Soc., 2021, 143, 17412.

[3] Bossion, A.; Zhu, C.; Guerassimoff, L.; Mougin, J.; Nicolas, J. Vinyl copolymers with faster hydrolytic degradation than aliphatic polyesters and tunable upper critical solution temperatures. Nat. Commun., 2022, 13, 2873.

MICROTUBULE ASSEMBLY AS A MOLECULAR BROWNIAN RATCHET: FROM ATOMISTIC TO SUPER COARSE-GRAINED MODELING

MAXIM IGAEV[a], MAKSIM KALUTSKII[a], HELMUT GRUBMÜLLER[a]
[a] Max Planck Institute for Multidisciplinary Sciences

Microtubules (MTs), mesoscopic cellular filaments, grow by the addition of GTP-bound tubulin dimers at their dynamic flaring ends. They operate as chemomechanical energy transducers with stochastic transitions between phases of growth and shortening upon hydrolyzing GTP to GDP. During this astounding process, known as “dynamic instability”, MTs can push and pull on other cellular compartments. However, we do not yet fully understand exactly how MTs self-assemble and transmit mechanical force. Moreover, physical constraints present within the cell cause MTs to bend and fracture, resulting in intricate collective phenomena such as softening and self-repair, which are still elusive.

Despite the importance of MTs, their highly dynamic nature makes it hard to study their structure and dynamics experimentally. On the other hand, there are no multiscale computational approaches to predict the impact of subtle tubulin changes driven by GTP hydrolysis on the large-scale behavior of MTs. To address this gap, we use atomistic, explicit-solvent simulations to scrutinize the microsecond dynamics of complete GTP- and GDP-MT end models[1]. Our findings show that the post-hydrolysis MT end is exposed to higher activation energy barriers for straight lattice formation, which strongly reduces its probability to elongate. To extrapolate this knowledge to real-life timescales, we are now developing a novel coarse-grained model of MT assembly based on the discrete elastic rod representation and parametrized fully ab initio. Overall, our study provides an information-driven Brownian ratchet mechanism for MT self-assembly and offers new insights into the mechanoenzymatics of and force generation by MTs.

References

[1] Igaev, M. and Grubmüller, H. Bending-torsional elasticity and energetics of the plus-end microtubule tip, PNAS, 2022, 119(12): e2115516119.

EXPLORING THE IMPACT OF BACTERIAL LIPOPOLYSACCHARIDES IN ANTIMICROBIAL RESISTANCE AND INNATE IMMUNE MODULATION

ALEJANDRA MATAMOROS-RECIO[a]
JUAN FELIPE FRANCO-GONZALEZ[a]
SONSOLES MARTÍN-SANTAMARÍA[a]
[a] Center for Biological Research "Margarita Salas" (CIB-CISC). C. Ramiro de Maeztu 9, 28040 Madrid, Spain

Gram-negative bacteria lipopolysaccharides (LPS) are major players in bacterial infection. LPS is released via formation of outer-membrane vesicles (OMVs) and recognized by Toll-like receptor 4 (TLR4) on host innate immune cells. TLR4 activation is associated with certain autoimmune diseases, noninfectious inflammatory disorders, and neuropathic pain. Therefore, TLR4 has risen as a promising therapeutic target.[1] Nonetheless, structural studies deepening into the full-length TLR4 activation are scarce due to the high complexity of the receptor.

Besides its role in bacterial infection, LPS is also a major player in Antimicrobial Resistance (AMR). AMR is a worldwide health emergency. ESKAPE pathogens include the most relevant AMR bacterial families. In particular, Gram-negative bacteria stand out due to their cell envelope complexity, which exhibits strong resistance to antimicrobials, being a key element the chemical structure of LPS, inflecting the membrane permeability to antibiotics.[2]

In this context, we have addressed the following objectives by means of computational chemistry techniques i) to investigate the implication of the altered permeability in resistant bacteria due to the membrane composition and the chemical structure of the LPS; and ii) to deepen into the molecular recognition events of TLR4 at the atomic and molecular levels to gain insight into its activation and signaling.

To those aims, we have modeled heterogeneous lipid liposomes as OMVs models using coarse-grained MD simulations to capture the role of LPS in the membrane properties and morphology of ESKAPE Gram-negative bacteria. The reported antimicrobial peptides Cecropin B1, JB95, and PTCDA1-kf were employed to unveil their implications for membrane disruption.[3]

Furthermore, we employed molecular modeling techniques to characterize the recognition processes of TLR4 agonist and antagonist modulators, including small molecule modulators,[4] monosaccharide-based synthetic glycolipids with potential activity as vaccine adjuvants,[5] and singular LPSs from opportunistic bacteria, accounting for molecular strategies of host evasion. We have also modeled, by all-atom MD simulations, the most realistic and complete 3D models of the active full TLR4 complex to date, embedded into a model membrane, accounting for the active (agonist) state of the receptor, providing an analysis at both atomic/molecular and thermodynamic levels of the TLR4 assembly.[6]

Our computational studies open promising starting points to observe and predict the activity of new antimicrobial agents able to modulate the bacterial envelope properties, overcoming the AMR problem. Additionally, our work unveils relevant molecular aspects for TLR4 activation and adaptor recruitment in the innate immune pathways, contributing to the discovery of new immunomodulators.

References

[1] A. Matamoros-Recio, J.F. Franco-Gonzalez, R.E. Forgione, A. Torres-Mozas, A. Silipo, S. Martín-Santamaría. Understanding the antibacterial resistance : computational explorations in bacterial membranes. ACS Omega, 2021, 6(9), 6041.

[2] J. M. Billod, A. Lacetera, J. Guzmán-Caldentey, S. Martín-Santamaría. Computational approaches to Toll-Like receptor 4 modulation. Molecules, 2016, 21, 8, 994.

[3] J.F. Franco-Gonzalez, A. Matamoros-Recio, A. Torres-Mozas, B. Rodrigo-Lacave, S. Martín-Santamaría. Lipid-A-dependent and cholesterol-dependent dynamics properties of liposomes from gram-negative bacteria in ESKAPE. Sci. Rep., 2022, 12, 19474.

[4] P. Gratal, A. Mediero, A. Lamuedra, A. Matamoros-Recio, C. Herencia, G. Herrero-Beaumont, S. Martín-Santamaría, R. Largo. 6-Shogaol (enexasogoal) treatment improves experimental knee osteoarthritis exerting a pleiotropic effect over immune innate signalling responses in chondrocytes. Br. J. Pharmacol. 2022, 179(22), 5089-5108.

[5] A. Romerio, N. Gotri, A. R. Franco, V. Artusa, M. Monsoor-Shaik, S. T. Pasco, U. Atxabal, A. Matamoros-Recio, M. Mínguez-Toral, J. D. Zalamea, A. Franconetti, N. G. A. Abrescia, J. Jimenez-Barbero, J. Anguita, S. Martín-Santamaría, F. Peri. J. New Glucosamine-Based TLR4 Agonists: Design, Synthesis, Mechanism of Action, and In Vivo Activity as Vaccine Adjuvants. Med. Chem. 2023, 66(4), 3010–3029.

[6] A. Matamoros-Recio, J. F. Franco-Gonzalez, L. Perez-Regidor, J.M. Billod, J. Guzman-Caldentey, S. Martin-Santamaria. Full-atom model of the agonist LPS-bound Toll-like receptor 4 dimer in a membrane environment. Chem. Eur. J., 2021, 27(62), 15406.

CO-EXISTING PHASES AND CONDENSATES OF PROMOTER AND GENE CONDENSATES IN COARSE-GRAINED MOLECULAR DYNAMICS SIMULATIONS

ARYA CHANGIARATH SIVADASAN [abc], LUKAS S. STELZL[abc]

[a] Institute of Physics, Johannes Gutenberg University Mainz, Staudingerweg 7, 55128 Mainz, Germany
[b] Faculty of Biology, Johannes Gutenberg University Mainz, 55128 Mainz, Germany
[c] Institute of Molecular Biology (IMB), 55128 Mainz, Germany

The formation of distinct phase-separated condensates of biological macromolecules has been proposed to provide specific regulation in cells. How different condensates with specific biomolecules coexist spatiotemporally and perform distinct functions in cells is not well understood. We study the conditions where phase-separated condensates are actually able to provide for distinct environments either as multiphase condensates or by forming entirely different condensates with coarse-grained molecular dynamics simulations[1]. Here, we are looking at an example of gene transcription process in which initiation and elongation processes may take place in different condensates called the promoter and gene body condensates respectively[2]. We consider the ternary system comprising CTD of RNA Pol II (CTD), Phosphorylated CTD (pCTD), and HRD of (HRD). It is evident from the simulations that CTD undergoes homotypic phase separation, whereas, pCTD or HRD does not show any homotypic phase-separation behavior. However, pCTD and HRD co-phase separate to form a heterotypic condensate that shares an interface with the CTD condensate. We probe into intermolecular homotypic and heterotypic interactions in the mixture which give rise to the specificity of biomolecules in being different condensates. Furthermore, the composition and the charge distribution of biomolecules in different condensates could play an important role in determining the interfacial tension between the condensates which we compute and the formation of different morphology. We investigate how the morphology of multiphase condensates is crucial in providing different domains for specific regulations in cells.

References

[1] G. L. Dignon, W. Zheng, Y. C. Kim, R. B. Best, and J. Mittal, PLOS Computational Biology 14, 1 (2018) .
[2] P. Cramer, Nature 573, 45 (2019).

COARSE GRAINED DYNAMICS IN MOLECULAR SIMULATION: GENERALIZED LANGEVIN EQUATION

HADRIEN VROYLANDT[a], LUDOVIC GOUDENÈGE[b]
PIERRE MONMARCHÉ[c,d], FABIO PIETRUCCI[e]
BENJAMIN ROTENBERG[f]

[a] Institut des Sciences du Calcul et des Données, Sorbonne Université, F-75005 Paris, France
[b] CNRS, FR 3487, Fédération de Mathématiques de CentraleSupélec, CentraleSupélec, Université Paris-Saclay, 91190 Gif-sur-Yvette, France
[c] Laboratoire Jacques-Louis Lions, Sorbonne Université, F-75005 Paris, France
[d] Laboratoire de Chimie Théorique, Sorbonne Université, F-75005 Paris, France
[e] Muséum National d'Histoire Naturelle, UMR CNRS 7590, Institut de Minéralogie, de Physique des Matériaux et de Cosmochimie, Sorbonne Université, F-75005 Paris, France
[f] Physicochimie des Électrolytes et Nanosystèmes Interfaciaux, Sorbonne Université, CNRS, F-75005 Paris, France

Generalized Langevin equations[1] with non-linear forces and memory kernels[2,3] are commonly used to describe the effective dynamics of coarse-grained variables in molecular dynamics. Such reduced dynamics play an essential role in the study of a broad class of processes, ranging from chemical reactions in solution to conformational changes in biomolecules or phase transitions in condensed matter systems.

Using a parametrization based on hidden auxiliary variables[4], we obtain a generalized Langevin equation by maximizing the statistical likelihood of the observed trajectories. Both the memory kernel and random noise are correctly recovered by this procedure. This data-driven approach provides a reduced dynamical model for multidimensional collective variables, enabling the accurate sampling of their long-time dynamical properties at a computational cost drastically reduced with respect to all-atom numerical simulations. The present strategy, based on the reproduction of the dynamics of trajectories rather than the memory kernel or the velocity-autocorrelation function, conveniently provides other observables beyond these two, including, e.g., stationary currents in nonequilibrium situations or the distribution of first passage times between metastable states.

References

[1] Zwanzig, R., Nonequilibrium statistical mechanics (Oxford University Press, 2001).

[2] Hadrien Vroylandt, and Pierre Monmarché. Position-dependent memory kernel in generalized Langevin equations: theory and numerical estimation. J. Chem. Phys. 156, 244105 (2022).

[3] Hadrien Vroylandt. On the derivation of the generalized Langevin equation and the fluctuation-dissipation theorem EPL (Europhysics Letters), 2022, 140, 62003

[4] Hadrien Vroylandt, Ludovic Goudenège, Pierre Monmarché, Fabio Pietrucci, Benjamin Rotenberg. Likelihood-based parametric estimator for memory kernel in molecular dynamics PNAS 119 (2022)

INTEGRATED MOLECULAR DESIGN: RECENT DEVELOPMENTS IN THE SAMSON PLATFORM

STEPHANE REDON[a]
[a] OneAngstrom
https://www.oneangstrom.com

In recent years, SAMSON[1] has emerged as a popular platform for molecular design, offering an integrated environment for scientists to model, simulate, and analyze complex molecular systems. This talk will provide a look at the latest developments in the platform, focusing on three advancements: the extension of the data model for coarse-grained systems, new tools for Python development, and the introduction of SAMSON AI, a large language model (LLM)-based assistant.

First, we will explore the extension of the data model, which has been specifically designed to facilitate the modeling of coarse-grained systems[2]. Coarse-grained models are crucial for studying large-scale biomolecular systems, as they enable the simplification of atomic details while retaining essential features of the system. The extended data model allows for efficient representation and manipulation of coarse-grained entities, in particular when preparing molecular dynamics simulations.

Next, we will delve into the new tools available for Python development within the SAMSON platform. These tools are designed to empower researchers by providing a user-friendly interface for scripting and automating molecular design tasks. We will showcase the enhancements in the Python API, including improved documentation, and the integrated Python editor. We will show how, by leveraging these tools, users can create custom scripts, integrate machine learning models, perform complex data analysis, and automate workflows.

Lastly, we will introduce SAMSON AI, an innovative, integrated LLM-based assistant designed to change the way researchers interact with the SAMSON platform. SAMSON AI harnesses the power of natural language processing and machine learning to provide context-aware assistance, making it easier for users to navigate the platform and access relevant information and documentation, set up calculations, analyze results, and create new applications for molecular design.

References

[1] https://www.samson-connect.net

[2] R. Alessandri, S. Thallmair, C.G. Herrero, R. Mera- Adasme, S.J. Marrink, P.C.T. Souza, A Practical Introduction to Martini 3 and its Application to Protein-Ligand Binding Simulations, in A Practical Guide to Recent Advances in Multiscale Modeling and Simulation of Biomolecules [AIP Publishing (online), Melville, New York, 2023].

Pr. Lucie Delemotte

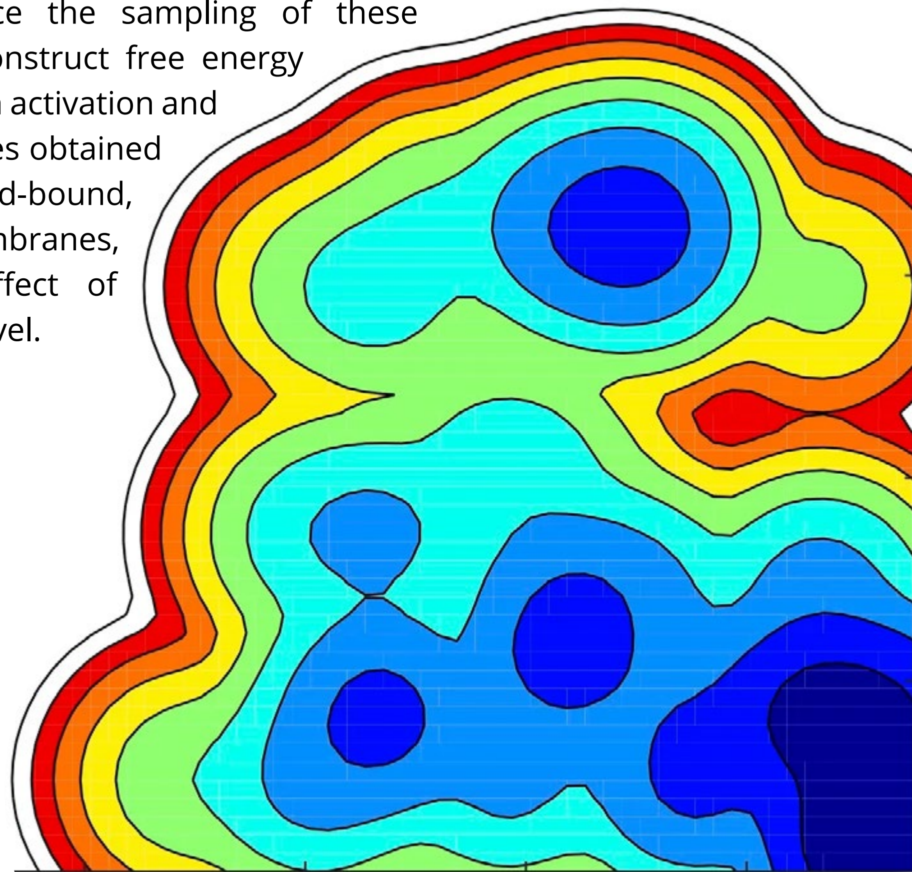
Department of Applied Physics,
Science for Life Laboratory,
Stockholm, Sweden

@delemottelab

biophysics.se

DATA-DRIVEN ENHANCED SAMPLING OF CONFORMATIONAL CHANGES IN MEMBRANE PROTEINS

Membrane proteins engage in cellular communication and do so by cycling between various conformational states. Molecular dynamics simulations provide exquisite resolution insights into these processes but often are too short to be able to resolve full conformational cycles. In this talk, we will discuss how experimental data can be used to enhance the sampling of these conformational changes and construct free energy landscapes of membrane protein activation and gating. Comparing the landscapes obtained under different conditions (ligand-bound, pH, embedded in different membranes, etc...) help rationalize the effect of perturbations at the atomistic level.



NOTES

TALK N°7

EMMIVOX: BAYESIAN INFERENCE APPROACH TO DETERMINING STRUCTURAL ENSEMBLES USING CRYO-EM AND SIMULATIONS

SAMUEL HOFF[a], MASSIMILIANO BONOMI[a]

[a] Institut Pasteur, Université Paris Cité, CNRS UMR 3528, 75015 Paris, France

Relatively recent advances in cryo-EM have led to a large increase in the number of high-resolution structures solved and deposited within the Protein Data Bank. However, the way protein models are represented has not kept pace as static single structure models continue to be widely accepted as accurate depictions of proteins despite their underlying conformational heterogeneity and dynamics. While advances in image processing allows for the construction of multiple distinct conformational states directly from raw 2D images, reconstruction of dynamic regions is still challenging[1]. Particularly, functionally important highly entropic regions of proteins including highly-dynamic side chains, are often averaged out during 3D map reconstruction, resulting in low map resolution and an inability to accurately determine local structures. As biomolecule function is tied to dynamics and structure, understanding these regions is crucial to understanding protein function. Here we present 'EMMIVox', a Bayesian inference approach to determining structural ensembles of biological entities by combining cryo-EM data with Molecular Dynamics (MD) simulations. This new tool builds upon previous works and utilizes the metainference approach[2,3]. EMMIVox, which is able to automatically detect and downweigh noisy experimental data, calculates accurate structural ensembles of proteins and protein complexes including any lipids, small-molecules and ordered water present in experimental maps. Using EMMIVox, we generated single structure models for a variety of high-resolution cryo-EM maps (1.9 Å – 3.5 Å), including membrane proteins and ligand-protein complexes. EMMIVox models outperformed deposited structures in terms of stereochemical descriptors (MolProbity and Clashcore) as well as metrics that quantify how well the model fit the density map (map-model cross correlation and EMRinger). Finally, we determined the structural ensembles of the type 1a tau filament (1.9 Å) and the SPP1 bacteriophage (4 Å) in detail, demonstrating the accuracy of the structural ensemble compared to standard MD simulations. Inference of structural ensembles and their representative populations will have wide ranging applications in structural biology and drug discovery. EMMIVox is available in the Integrative Structural and Dynamical Biology module of the open-source, freely-available PLUMED library (www.plumed.org).

References

[1] Zivanov, J. et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. Elife 7 (2018).

[2] Bonomi, M., Pellarin, R and Vendruscolo, M. Simultaneous Determination of Protein Structure and Dynamics Using Cryo-Electron Microscopy. Biophys. J., 114 (2018), pp. 1604-1613

[3] Bonomi, M., Camilloni, C., Cavalli, A. & Vendruscolo, M. Metainference: A Bayesian inference method for heterogeneous systems. Sci Adv 2, e1501177 (2016).

ON THE AUTOMATIC OPTIMIZATION OF LIPID MODELS IN THE MARTINI FORCE FIELD USING SWARMCG

CHARLY EMPEREUR-MOT[a], KASPER B. PEDERSEN[b], CRISTINA CARUSO[c], MATTIA PERRONE[c], MARTINA CRIPPA[c] PAULO C.T. SOUZA[d], SIEWERT J. MARRINK[e] & GIOVANNI M. PAVAN[a,c]

[a] Department of Innovative Technologies, University of Applied Sciences and Arts of Southern Switzerland, Lugano, Switzerland
[b] Department of Chemistry, Aarhus University, Aarhus, Denmark
[c] Politecnico di Torino, Department of Applied Science and Technology, Torino, Italy
[d] Molecular Microbiology and Structural Biochemistry, UMR 5086 CNRS and University of Lyon, Lyon, France
[e] Groningen Biomolecular Sciences and Biotechnology Institute and Zernike Institute for Advanced Material, University of Groningen, Groningen, the Netherlands

After two decades of continued development of the Martini coarse-grained force field[1] (CG FF), further refining the already rather accurate Martini lipid models has become a demanding task that could benefit from integrative data-driven methods. Automatic approaches are increasingly used in the development of accurate molecular models, but they typically make use of specifically-designed interaction potentials that transfer poorly to molecular systems or conditions different than those used for model calibration. As a proof of concept here we employ *SwarmCG*[2,3], an automatic multi-objective optimization approach facilitating the development of lipid force fields, to refine specifically the bonded interaction parameters in building blocks of lipid models within the framework of the general Martini CG FF. As targets of the optimization procedure, we employ both experimental observables (area per lipid & bilayer thickness) and all-atom molecular dynamics simulations, respectively informing on the supra-molecular structure of the lipid bilayer systems and on their sub-molecular dynamics. In our training sets we simulate at different temperatures in the liquid and gel phases up to 11 homogeneous lamellar bilayers, composed of phosphatidylcholine lipids spanning various tail lengths and degrees of (un)saturation. We explore different CG representations of the molecules and evaluate improvements *a posteriori* using additional simulation temperatures and a portion of the phase diagram of a DOPC/DPPC mixture. Successfully optimizing up to ~80 model parameters within still limited computational budgets, we demonstrate that this protocol allows to obtain improved transferable models, as well as to gain insights on the ability of putative refined CG representations to further enhance the thermodynamic properties of lipid models in Martini simulations.

References

[1] Souza PCT et al. Martini 3: a general purpose force field for coarse-grained molecular dynamics. Nature Methods, 2021, 18, 382-388.

[2] Empereur-Mot C, Pesce L, Doni G, Bochicchio D, Capelli R, Perego C, Pavan GM. SwarmCG: Automatic Parametrization of Bonded Terms in MARTINI-Based Coarse-Grained Models of Simple to Complex Molecules via Fuzzy Self-Tuning Particle Swarm Optimization. ACS Omega, 2020, 5, 32823-43.

[3] Empereur-Mot C, Capelli R, Perrone M, Caruso C, Doni G, Pavan GM. Automatic Multi-Objective Optimization of Coarse-Grained Lipid Force Fields using SwarmCG. J Chem Phys, 2022, 156, 024801.

ENHANCED-SAMPLING COUPLED WITH BIOPHYSICS TO BOOST THE PERFORMANCE OF DRUG SCREENING

XIAOJING CONG, DAMIEN MAUREL, OMOLADE OTUN, CHERINE BECHARA, CARINE BECAMEL ; RÉMY SOUNIER, GUILLAUME LEBON, SÉBASTIEN GRANIER

Institut de Génomique Fonctionnelle, Université de Montpellier, CNRS, INSERM, 34094 Montpellier cedex 5, France.

Proteins and ligands adapt their conformations to each other upon binding. In the case of G protein-coupled receptors (GPCRs)—the targets of nearly 40% of marketed drugs—the receptors may undergo remarkable conformational changes upon ligand binding[1]. Yet, structure-based drug screening is commonly done with one or few receptor conformations, e.g. experimental structures, AlphaFold2 models or homology models. A large ensemble of the receptor conformations relevant for drug screening is often metastable and inaccessible. This is particularly true for agonists and positive allosteric modulators (PAMs) which exert their activity in a highly dynamic process.

We explore the conformational ensembles of GPCRs using enhanced-sampling MD simulations coupled with biophysical measurements, e.g. NMR and HDX-mass spectroscopy. The approach allows to identify distinct receptor conformations for virtual drug screening. In two studies as showcases here, we investigated biased agonists of the μ-opioid receptor[2] and PAMs of the metabotropic glutamate receptor[3]. Specific receptor conformations were identified and used for virtual screening of new biased agonists or PAMs. In each case, we tested less than 50 compounds to obtain new ligands that are active *in vivo* (in preclinical mouse models, unpublished data). Given the recent breakthroughs in cryo-EM and AlphaFold2, high-quality structures and models are available for more and more drug target proteins. Therefore, we expect our approach to find wide applications in drug screening.

References

[1] Weiss WI, Kobilka BK. The Molecular Basis of G Protein-Coupled Receptor Activation, Annu. Rev. Biochem., 2018. 87, 897–919.

[2] Cong X, Maurel D, et al. Molecular insights into the biased signaling mechanism of the μ-opioid receptor, Mol. Cell., 2021, 81, 1-11.

[3] Cong X, Chéron JB, et al. Allosteric Modulation Mechanism of the mGluR5 Transmembrane Domain, J. Chem. Info. Mod.,2019, 59, 2871-2878.

TALK N°10

WASCO: A WASSERSTEIN-BASED STATISTICAL TOOL TO COMPARE CONFORMATIONAL ENSEMBLES OF INTRINSICALLY DISORDERED PROTEINS

JAVIER GONZÁLEZ-DELGADO[a,b], AMIN SAGAR[c], CHRISTOPHE ZANON[a], KRESTEN LINDORFF-LARSEN[d], PAU BERNADÓ[c], PIERRE NEUVIAL[b] AND JUAN CORTÉS[a]

[a] LAAS-CNRS, Université de Toulouse, CNRS, Toulouse, France.
[b] Institut de Mathématiques de Toulouse, Université de Toulouse, CNRS, Toulouse, France.
[c] Centre de Biologie Structurale, Université de Montpellier, INSERM, CNRS, Montpellier, France.
[d] The Linderstrøm-Lang Centre for Protein Science, Department of Biology, University of Copenhagen, Denmark.

The structural investigation of intrinsically disordered proteins (IDPs) requires ensemble models describing the diversity of the conformational states of the molecule. Due to their probabilistic nature, there is a need for new paradigms that understand and treat IDPs from a purely statistical point of view, considering their conformational ensembles as well-defined probability distributions. In this work, we define a conformational ensemble as an ordered set of probability distributions and provide a suitable metric to detect differences between two given ensembles at the residue level, both locally and globally. The underlying geometry of the conformational space is properly integrated, one ensemble being characterized by a set of probability distributions supported on the three-dimensional Euclidean space (for global-scale comparisons) and on the two-dimensional flat torus (for local-scale comparisons). The inherent uncertainty of the data is also taken into account to provide finer estimations of the differences between ensembles. Additionally, an overall distance between ensembles is defined from the differences at the residue level. We illustrate the interest of the approach with several examples of applications for the comparison of conformational ensembles: (i) produced from molecular dynamics (MD) simulations using different force fields, and (ii) before and after refinement with experimental data. We also show the usefulness of the method to assess the convergence of MD simulations, and discuss other potential applications such as in machine-learning-based approaches. The numerical tool has been implemented in Python through easy-to-use Jupyter Notebooks available at https://gitlab.laas.fr/moma/WASCO

References

[1] González-Delgado J, Sagar A, Zanon C, Lindorff-Larsen K, Bernadó P, Neuvial P, Cortés J: WASCO: A Wasserstein-based statistical tool to compare conformational ensembles of intrinsically disordered proteins, 2022 bioRxiv 2022.12.01.518687.

TOWARDS AN UNDERSTANDING OF THE ALLOSTERIC POTENTIATION MECHANISM OF NICOTINIC ACETYLCHOLINE RECEPTOR α7

MARIIA AVSTRIKOVA [a,b], JEAN-PIERRE CHANGEUX [c] AND MARCO CECCHINI [b]

[a] Institut Pasteur, UniversitéParis Cité , CNRS UMR 3571, Channel-Receptors Unit, Paris, France
[b] Institut de Chimie de Strasbourg, CNRS UMR 7177, Université de Strasbourg, Strasbourg, France
[c] Neuroscience Department, Institut Pasteur, Collège de France, Paris, France

Nicotinic acetylcholine receptors (nAChRs) are members of the pentameric Ligand-Gated Ion Channel superfamily – a type of synaptic neurotransmitter receptors that is responsible for various cognitive functions such as attention, memory, learning, etc. [1]. nAChRs are interesting pharmacological targets as they are related to a variety of neurological disorders [2] and hypothetically COVID-19 [3].

nAChRs are allosteric proteins: their response to neurotransmitters can be potentiated by the binding of positive allosteric modulators (PAMs). α7 nAChR has two types of PAMs: type I PAMs only enhance the response to agonists, while type II PAMs also delay the subsequent desensitization of the receptor [4].

The main goal of this project is to explore the mechanism of action of distinct types of PAMs from a structural point of view. This knowledge would open the door to the allosteric drug discovery for α7 nAChR.

First, to study the functioning of the receptor in the absence of PAMs, we performed molecular dynamics (MD) simulations of recently published structures of α7 nAChR [5]. These structures represent the receptor in its three distinct conformational states: active, resting and desensitized. Active state structure possesses an open ion channel, while structures of resting and desensitized states have a closed ion channel. We modelled the physiological environment and performed 300 ns all-atom MD simulations in three independent replicas for each structure.

Analysis of these simulations revealed an unexpected behaviour of the desensitized state structure. This analysis suggests that this structure is capable of local spontaneous transitions between open-like and closed-like conformations at the level of the ion channel. At the same time, the rest of the protein stays in the desensitized conformation.

We hypothesized that the delay of the desensitization, which is distinctive of type II PAMs, could be related to the stabilization of the desensitized conformation with an open channel. To validate this hypothesis, we performed the same MD simulations for α7 nAChR structures bound to different PAMs. These structures are obtained in collaboration with the group of Dr Ryan Hibbs (University of Texas Southwestern Medical Center).

The overall conformation of these structures is consistent with the desensitized conformation. Furthermore, during the simulation, the structures bound to type II PAM keep the ion channel open, while structures bound to type I PAM remain closed, which is coherent with our hypothesis.

References

[1] Cecchini M, Changeux JPP. The nicotinic acetylcholine receptor and its prokaryotic homologues: Structure, conformational transitions & allosteric modulation. Neuropharmacology. 2015, 96(PB):137-149.

[2] Taly A, Corringer PJ, Guedin D, Lestage P, Changeux JP. Nicotinic receptors: Allosteric transitions and therapeutic targets in the nervous system. Nature Reviews Drug Discovery. 2009, 8(9):733-750.

[3] Changeux JP, Amoura Z, Rey FA, Miyara M. A nicotinic hypothesis for Covid-19 with preventive and therapeutic implications. Comptes Rendus - Biologies. 2020, 343(1):33-39.

[4] Papke RL, Lindstrom JM. Nicotinic acetylcholine receptors: Conventional and unconventional ligands and signaling. Neuropharmacology. 2020, 168:108021.

[5] Noviello CM, Gharpure A, Mukhtasimova N, et al. Structure and gating mechanism of the α7 nicotinic acetylcholine receptor. Cell. Published online March 17, 2021, 1-14.

MEDICINAL AND COMPUTATIONAL CHEMISTRY SYNERGY AT EVOTEC

LIONEL COLLIANDRE, CHRISTOPHE BOLDRON
Evotec SE, Molecular Architects, Integrated Drug Discovery, Campus Curie, 195 Route d'Espagne, 31036 Toulouse

Drug design is evolving thanks to the significant increase in the amount of data and computing power as well as the development of cutting-edge in silico technologies including AI/ML. To transform these major developments into drug research accelerator, Evotec has created two main entities, namely in-silico R&D and Molecular Architects which optimally integrate AI methods' developers and, computational & medicinal chemists by background.

In this presentation, we will highlight the synergy of these trained and skilled profiles to apply advanced and integrated design as part of rapid data-rich Design Make Test Analyse (DMTA) cycles.

Pr. Jean-Philip Piquemal

Laboratoire de Chimie Théorique (LCT),
Sorbonne Université, Paris, France

@jppiquem

piquemalresearch.com

HIGH-RESOLUTION MOLECULAR DYNAMICS SIMULATIONS FOR BIOPHYSICS WITH TINKER-HP

I will discuss our strategy for high-resolution molecular dynamics towards biophysical applications. As I detail the various protein targets that are currently under study, I will show how the newly developed multi-GPUs version of the Tinker-HP software [1,2] can accelerate high-resolution molecular dynamics simulations. Indeed, thanks to adaptive sampling and new generation many-body polarizable force fields such as AMOEBA, long (µs) molecular dynamics simulations at enhanced accuracy become possible.[3] As I detail the currently available other enhanced sampling capabilities of the software, I will give some perspectives about the use of new hybrid physically-driven machine learning approaches [4, 5] for condensed phase molecular dynamics.

1) Tinker-HP: a Massively Parallel Molecular Dynamics Package for Multiscale Simulations of Large Complex Systems with Advanced Polarizable Force Fields. L. Lagardère, L.-H. Jolly, F. Lipparini, F. Aviat, B. Stamm, Z. F. Jing, M. Harger, H. Torabifard, G. A. Cisneros, M. J. Schnieders, N. Gresh, Y. Maday, P. Ren, J. W. Ponder, J.-P. Piquemal, Chem. Sci., **2018**, 9, 956-97 (Open Access), DOI: 10.1039/C7SC04531J

2) Tinker-HP: Accelerating Molecular Dynamics Simulations of Large Complex Systems with Advanced Point Dipole Polarizable Force Fields using GPUs and Multi-GPUs systems. O. Adjoua, L. Lagardère, L.-H. Jolly, Arnaud Durocher, Z. Wang, T. Very, I. Dupays, T. Jaffrelot Inizan, F. Célerse, P. Ren, J. Ponder, J-P. Piquemal, J. Chem. Theory. Comput., **2021**, 17, 4, 2034–2053 (Open Access), DOI: 10.1021/acs.jctc.0c01164

3) High-Resolution Mining of SARS-CoV-2 Main Protease Conformational Space: Supercomputer-Driven Unsupervised Adaptive Sampling. T. Jaffrelot Inizan, F. Célerse, O. Adjoua, D. El Ahdab, L.-H. Jolly, C. Liu, P. Ren, M. Montes, N. Lagarde, L. Lagardère, P. Monmarché, J.-P.Piquemal, Chem. Sci., **2021**, 12, 4889 – 4907 (Open Access), DOI: 10.1039/D1SC00145K

4) Scalable Hybrid Deep Neural Networks/Polarizable Potentials Biomolecular Simulations including long-range effects. T. Jaffrelot Inizan, T. Plé, O. Adjoua, P. Ren, H. Gökcan, O. Isayev, L. Lagardère, J.P. Piquemal, **2023**, DOI: 10.48550/arXiv.2207.14276

5) Force-Field-Enhanced Neural Network Interactions: from Local Equivariant Embedding to Atom-in-Molecule properties and long-range effects. T. Plé, L. Lagardère, J.-P. Piquemal, **2023**, DOI: 10.48550/arXiv.2301.08734

NOTES

A NEW METHOD TO PREDICT CONFORMATIONAL MOTIONS IN PROTEINS

AXEL DESCAMPS[a], RODOLPHE VUILLEUMIER[a], NICOLAS CHERON[a]

[a] PASTEUR, Département de chimie, Ecole Normale Supérieure, PSL University, SorbonneUniversité, CNRS, 75005 Paris, France

Even if one could consider the problem of structure prediction as solved for proteins since 2022[1], AlphaFold and RosettaFold still have caveats. Indeed, the prediction of the conformation of mobile motifs remains far from reach from these two programs. One only has to look at the definition of a mobile motif (a sub-structure in the protein that can adopt several stable conformations) to realize that predictions of these structures will remain a big challenge for AI[2]. This is unfortunate since these sub-structures play a major role in protein-protein interactions or in enzymatic reactivity[3], and experimental studies sometimes fail to describe the different conformations of a protein. In this talk, we will present a new methodology which aims at predicting the active structure of several enzymes, as well as describing the mechanism of the motion.

We have developed an approach where steered-MD simulations and simulations under constraints allow us to recover the closed conformations of several protein loops by starting only from the open conformation. During the first steered-MD simulations, we follow the evolution of the potential energies of the dihedral angles from the backbone. This provides a general response of the system to an external stress. This response, which is only mildly dependent on the stress (i.e. on the coordinate chosen for the steered-MD), provides information on the strongest point of the system. A second set of steered-MD simulations are then performed where the potential energies of some dihedral angles are set to 0, allowing the system to overcome the energy barrier. Finally, the system is relaxed first with constraints and then with enhanced sampling.

The method we have developed has several advantages. The first one is that we can start from an experimental crystallographic structure with the open loop which is often easier to obtain than the closed conformation. Moreover, the transition from the inactive to the active structure reveals the mechanism of motion, which is not only an important information for understanding the reactivity, but also for bio-engineering.

References

[1] Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tun-yasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A. et al. Highly accurate proteinstructure prediction with AlphaFold.Nature2021,596, 583–589.

[2] Saldaño, T.; Escobedo, N.; Marchetti, J.; Zea, D. J.; Mac Donagh, J.;Velez Rueda, A. J.; Gonik, E.; Melani, A. G.; Novomisky Nechcoff, J.; Salas, M. N.et al.Impact of protein conformational diversity on AlphaFold predictions; preprint,2021.

[3] Richard, J. P. Protein Flexibility and Stiffness Enable Efficient Enzymatic Catalysis.Journal of the American Chemical Society2019,141, 3320–3331

A NOVEL PROTOCOL FOR PROTEIN-LIGAND BINDING AFFINITY PREDICTION VIA SPATIO-TEMPORAL DEEP LEARNING AND MOLECULAR DYNAMICS SIMULATIONS

LIBOUBAN PY.[a], ACI-SÈCHE S.[a], TRESADERN G.[b], BONNET P.[a]

[a] Institute of Organic and Analytical Chemistry (ICOA); UMR7311, Université d'Orléans, CNRS; Pôle de chimie rue de Chartres - 45067 Orléans Cedex 2, France

[b] 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. Although in vitro experiments are used to measure the binding 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]. These neural networks use the 3D structures of protein-ligand complexes to predict the binding affinity of complexes. They are trained on the PDBbind dataset, which gather the structural information of more than 17 000 protein-ligand complexes, as well as their affinity. Regarding their performance, these models achieve a correlation coefficient around 0.8 on a test set called "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 affinity of protein-ligand complexes using recent deep learning (DL) methods. To achieve this goal, we have developed a protocol combining DL and molecular dynamics (MD) simulations. Additional structures extracted from MD simulations are used as input for the models. In addition to acting as a data augmentation tool, MD simulations also add temporal information of the protein-ligand interactions that can be used to improve models.

We carried out 10 replicate simulations of 10 ns for each complex in a dataset of 6,300 structures. Therefore, we were able to produce 63,000 simulations. In order to create efficient statistical models by learning from these MD simulations, we developed neural networks able to analyse both spatial and temporal information, like the convolutional LSTM[4]. These neural networks combine a CNN able to extract the spatial information from the 3D structures at each time step, while the LSTM keeps track of the evolution of this information over the whole simulation.

We are currently evaluating their performance and expect these neural networks to outperform current state of the art statistical models in the prediction of binding affinity.

References

[1] Stepniewska-Dziubinska MM, Zielenkiewicz P, Siedlecki P. Development and evaluation of a deep learning model for protein-ligand binding affinity prediction, Bioinformatics, 2018, 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, 287-96.

[3] Volkov M, Turk J-A, Drizard N, Martin N, Hoffmann B, Gaston-Mathé Y, et al. On the Frustration to Predict Binding Affinities from Protein–Ligand Structures with Deep Neural Networks, Journal of medicinal chemistry, 2022, 7946-58.

[4] 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, 2015, 802-10.

TALK N°15

COMPUTATIONAL PROTEIN DESIGN WITH ARTIFICIAL INTELLIGENCE

MARIANNE DEFRESNE[a] , THOMAS SCHIEX[b], SOPHIE BARBE[a]

[a] TBI, Université de Toulouse, CNRS, INRAE, INSA, ANITI, Toulouse, France

[b] Université Fédérale de Toulouse, ANITI, INRAE, UR 875, Toulouse, France

Computational Protein Design (CPD) aims to design proteins with enhanced or new functions and/or properties. Structure-based methods target a function via a 3D input backbone, and aim to find sequence(s) folding onto this backbone. Traditional methods minimize a score function combining the energy of the sequence on the input backbone[1] with specific design constraints related to the precise design objective. However, various generally desirable properties such as expressivity, solubility or hydrophilicity are extremely difficult to capture mathematically in such score functions. Recent deep learning-based methods can implicitly capture these general properties by training on natural proteins. But then, specific design constraints become difficult to enforce with deep learning.

In this talk, we aim to take the best of both worlds by combining two artificial intelligence technologies: automated reasoning, augmented with deep learning. We keep the formulation CPD as the optimization of a pairwise decomposable energy function, but we learn the energy on known protein structures. This energy is then optimized exactly for each protein to design using automated reasoning[2]. We in silico tested the quality of the learned energy function, and found it to outperform state-ofthe art hybrid and statistical functions such as those available in Rosetta[1] or KORP[3] in terms of Native Sequence Recovery, a commonly used metric for CPD. We are currently testing it on applied projects we will briefly present.

References

[1] Park, H. et al. Simultaneous Optimization of Biomolecular Energy Functions on Features from Small Molecules and Macromolecules. Journal of

Chemical Theory and Computation, 2016, 12.12, pp. 6201–6212.

[2] Traoré, S. et al. A new framework for computational protein design through cost function network optimization. Bioinformatics, 2013, 29.17, pp.

2129–2136. issn:1367-4803.

[3] Lopez-Blanco, J. and Chacon, P. KORP: knowledge-based 6D potential for fast protein and loop modeling. Bioinformatics, 2019, 35.17, pp. 3013–3019. issn: 1367-4803.

TOWARDS ALL-ATOM SAMPLING OF MEMBRANE CONFIGURATIONS USING NON-EQUILIBRIUM MOLECULAR DYNAMICS

FLO SZCZEPANIAK[a,b], FRANÇOIS DEHEZ[a], BENOIT ROUX[b]

[a] Laboratoire de Physique et Chimie Théoriques 7019 – CNRS, Université de Lorraine, Vandoeuvre-Les-Nanct, France
[b] Department of Biochemistry and Molecular Biology, The University of Chicago, Chicago, IL, USA

In all-atoms simulations, phenomena that can be observed are limited in terms of time-scale, usually up to a few µs. Slow processes taking place on a longer time-scale are not observable. For example, reorganization of lipids around a protein can not be simulated, and therefore the protein-lipid interactions can not be properly modelized. [1] Here, we develop a method to generate all-atoms configurations of membrane. The diffusion of the lipids within the membrane is not simulated per say, but non-equilibrium molecular dynamics is used to exchange the positions of the lipids in the membrane. [2] With Monte-Carlo exchanges performed with alchemical transformation, different possible configurations of the membrane at the thermodynamic equilibrium are generated. It becomes possible to perform an equilibration of the membrane in all-atoms representation at a reduced sampling cost, and thus get a more precise model of the membrane organization. With the example of a membrane and a peptide, we show that this algorithm successfully equilibrates membranes, and considerably improves our ability to model protein-lipid interactions. In a few hundred nanoseconds only, we reorganize the membrane and reach a thermodynamic equilibrium that would need a few microseconds to be observed in classical Molecular Dynamics.

References

[1] Corradi, Valentina, Besian I. Sejdiu, Haydee Mesa-Galloso, Haleh Abdizadeh, Sergei Yu. Noskov, Siewert J. Marrink, et D. Peter Tieleman. « Emerging Diversity in Lipid–Protein Interactions ». Chemical Reviews 119, no 9 (8 mai 2019): 57755848

[2] Fathizadeh, Arman, et Ron Elber. « A Mixed Alchemical and Equilibrium Dynamics to Simulate Heterogeneous Dense Fluids: Illustrations for Lennard-Jones Mixtures and Phospholipid Membranes ». The Journal of Chemical Physics 149, no 7 (21 août 2018): 072325.

TALK N°17

EXTREMELY SCALABLE QM/MM MD SIMULATIONS OF THE CANCER-ASSOCIATED IDH1 ENZYME

BHARATH RAGHAVAN[a,b], MARCO DE VIVO[c], PAOLO CARLONI[a,b]

[a] Computational Biomedicine, Institute of Advanced Simulations IAS-5/Institute for Neuroscience and Medicine INM-9, Forschungszentrum Jülich GmbH, Jülich 52428, Germany

[b] Department of Physics, RWTH Aachen University, Aachen 52074, Germany

[c] Molecular Modeling and Drug Discovery, Italian Institute of Technology, 16163 Genova, Italy

We present the first application of the highly scalable Density Functional Theory (DFT)-based mixed Quantum Mechanical/Molecular Mechanical (QM/MM) Molecular Dynamics (MD) 'MiMiC' interface[2] to a problem of pharmaceutical relevance. We demonstrate the large scaling (almost up 30,000 cores on the JUWELS supercomputer) of this code for the investigation of the chemical reaction performed by the human IDH1 enzyme, whose variants have emerged as excellent therapeutic targets for glioma, glioblastoma and acute myeloid leukemia.[3,4] The system is split into 142 QM atoms and ~130,000 MM atoms, where we show that MiMiC allows us to reach the timescales (5.4 ps/day for BLYP and 0.24 ps/day for B3LYP) required to accurately map the free energy of catalysis of this enzyme, providing an accurate description of the transition state. This may be used for the rational design of transitions state analogues inhibitors, which bind many times stronger than current substrate analogs.[5] This procedure could have a huge impact on computational drug design workflows in the coming years.

References

[1] Ariana Remmel. (2022). Welcome to exascale computing. C&EN Global Enterprise, 100(31), 29–33. https://doi.org/10.1021/cen-10031-cover

[2] Bolnykh, V., Olsen, J. M. H., Meloni, S., Bircher, M. P., Ippoliti, E., Carloni, P., & Rothlisberger, U. (2019). Extreme Scalability of DFT-Based QM/MM MD Simulations Using MiMiC. Journal of Chemical Theory and Computation, 15(10), 5601–5613. https://doi.org/10.1021/acs.jctc.9b00424

[3] Horbinski, C. (2013). What do we know about IDH1/2 mutations so far, and how do we use it? Acta Neuropathologica, 125(5), 621–636. https://doi.org/10.1007/S00401- 013-1106-9

[4] Cerchione, C., Romano, A., Daver, N., DiNardo, C., Jabbour, E. J., Konopleva, M., Ravandi-Kashani, F., Kadia, T., Martelli, M. P., Isidori, A., Martinelli, G., & Kantarjian, H. (2021). IDH1/IDH2 Inhibition in Acute Myeloid Leukemia. Frontiers in Oncology, 11, 345. https://doi.org/10.3389/FONC.2021.639387

[5] Schramm, V. L. (2013). Transition States, Analogues and Drug Development. ACS Chemical Biology, 8(1), 71. https://doi.org/10.1021/CB300631K

THE POWER OF ACTIVE LEARNING IN DRUG DISCOVERY: HOW DOCKAI CAN REVOLUTIONIZE HIT DISCOVERY

MAUD JUSOT[a], PAUL JOIN-LAMBERT[a], GUILLAUME PLUM[a], NICOLAS DRIZARD[a], NICOLAS BROSSE[a], ENNYS GHEYOUCHE[a], CÉDRIC THAO[a], CHRISTOPHER HOUSSEMAN[a], BRICE HOFFMANN[a]

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

Identifying potential drug candidates, also called hit discovery, is the first and most critical step from a drug discovery campaign. Often achieved by virtual screening, large chemical libraries (~106 ligands) are docked into the target protein in order to narrow down a vast pool of compounds to a more manageable set of potential hits that can be further evaluated. This strategy has led to some success over the last two decades[1]. However, with the advent of make-on-demand ultra large chemical libraries[2], the size of the possible chemical search space has increased by orders of magnitude (~109) leading the deployment of traditional virtual screening methods prohibitively long and costly[3]. In order to address this limitation, we have developed a proprietary solution called DockAI, for a highly effective AI-based method for virtual screening of ultra large scale databases of virtual compounds. DockAI is a fast and low cost method. It is based on an active learning approach, which allows us to identify the most promising compounds for further evaluation by docking a small fraction (typically <1%) of the database. DockAI iteratively samples the database and uses the results of each docking round to guide the selection of the next set of compounds to be docked. This approach allows ultimately to converge on the most promising candidates and efficiently identify novel hits. Combining this with another of our approaches to generate molecules by designing de novo synthesizable (virtual) molecules based on available building blocks and chemistry that are not constrained by any database, allows exploration of a previously unexplored chemical space and discovering novel and highly active compounds that were previously unforeseen. DockAI outperformed Schrodinger's comparable approach on a dataset of Dopamine receptor (D4) compounds (n~140 millions): DockAI retrieved ~75% of the hit molecules after docking only 1 million compounds, demonstrating the efficiency of our method. DockAI has been successfully used in real-world projects and is deployed on AWS enabling us to speed up the process and to provide results of our screening campaigns within 24 hours.

References

[1] Lionta, Evanthia, et al. «Structure-based virtual screening for drug discovery: principles, applications and recent advances.» Current topics in medicinal chemistry 14.16 (2014): 1923-1938.

[2] Hoffmann, Torsten, and Marcus Gastreich. «The next level in chemical space navigation: going far beyond enumerable compound libraries.» Drug discovery today 24.5 (2019): 1148-1156.

[3] Acharya, Atanu, et al. «Supercomputer-based ensemble docking drug discovery pipeline with application to COVID-19.» Journal of chemical information and modeling 60.12 (2020): 5832-5852.

Dr. Noelia Ferruz

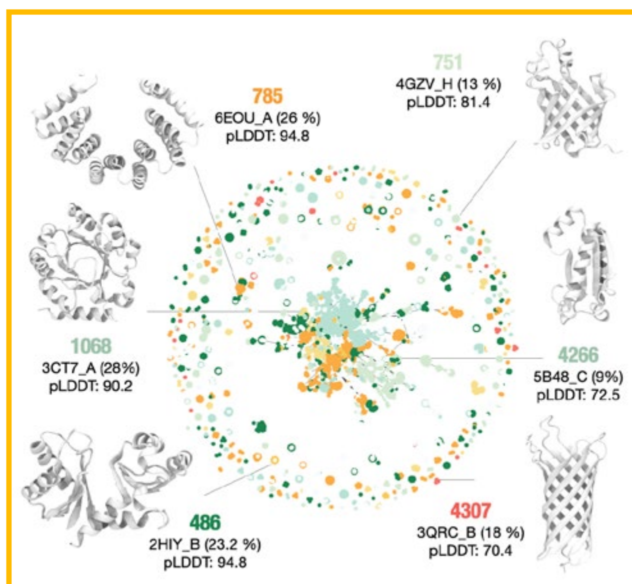
Department of Structural and Molecular Biology,
Molecular Biology Institute of Barcelona, Spain

@ferruz_noelia

aiproteindesign.com

DE NOVO PROTEIN AND ENZYME DESIGN WITH UNSUPERVISED LANGUAGE MODELS

Artificial Intelligence (AI) methods are emerging as incredibly compelling tools in fields such as Natural Language Processing (NLP) and Computer Vision (CV), impacting the tools and applications we use in our daily lives. Language models have recently shown incredible performance at understanding and generating human text, producing text often indistinguishable from that written by humans. Inspired by these recent advances, we trained a language model, ProtGTP2, which effectively learned the protein language and generated sequences in unexplored regions of the protein space. A desirable critical feature in protein design is having control over the design process, i.e., designing proteins with specific properties. For this reason, we trained ZymCTRL, a model trained on enzyme sequences and their associated Enzymatic Commission (EC) numbers. ZymCTRL generates enzymes upon user-defined specific catalytic reactions, thus enabling conditional de novo design of biocatalysts. Our preliminary experimental data shows remarkable success, with high expression rates.



NOTES

HOW LONG-RANGE MUTATION OF THE FGFR3 KINASE HAS AN IMPACT FOR INHIBITORS? A COMPUTATIONAL STUDY

LAURINE VASSEUR[a], LAURENCE LEGEAI-MALLET[b], FLORENT BARBAULT[a]

[a] ITODYS, UMR CNRS 7086, Univ Paris Cité

[b] Laboratoire bases moléculaires et physiopathologiques des ostéochondrodysplasies INSERM UMR 1163, Institut Imagine, Univ Paris Cité

Fibroblast Growth Factor Receptor 3 (FGFR3), belonging to the tyrosine kinase receptors family, plays a pivotal role in regulating cell growth and differentiation[1]. Various mutations of this receptor have demonstrated its implication in several bone pathologies such as achondroplasia – the most frequent form of dwarfism. Therefore, some years ago, the hypothesis that the inhibition of the kinase domain of FGFR3 could counteract dwarfism was posed[2]. To this extent, the use of Infigratinib, that target the ATP binding site of FGFR3, is recently entered the third phase of clinical development[3].

Recently, biologist collaborators identified a specific mutation in the kinase domain of FGFR3 (N540K) that worsened the development of dwarfism. Given the fact that the N540K mutation is neither located in the catalytic site nor the activation loop, a computational study was realized to lead deeper understanding of its specific role. In this sense, molecular docking, molecular dynamics (MD), Gaussian-accelerated MD, and free energy determinations through umbrella sampling were engaged.

Visualization and structural analysis highlight an original role of a loop in FGFR3 that stabilizes the interaction of Infigratinib. Interestingly, when the asparagine is replaced by lysine for the mutant a specific electrostatic network with the charged residues appears and favors the approach of the protein loop toward Infigratinib. Structural properties of the N540K mutated version of FGFR3 exhibit a distinct configuration favoring interaction with specific ligands, thus opening new developments of inhibitors with better selectivity for this receptor.

References

[1] Richette, P.; Bardin, T.; Stheneur, C. Achondroplasia: From Genotype to Phenotype. Joint Bone Spine 2008, 75 (2), 125–130. https://doi.org/10.1016/j.jbspin.2007.06.007.

[2] https://patents.google.com/patent/WO2013088191A1/en

[3] https://bridgebio-announces-positive-phase-2-cohort-5-results-of-infigratinib-in-achondroplasia/

CONFORMATIONAL DYNAMICS OF CHOLINE IMPORT MEDIATED BY A BACTERIAL MEMBRANE TRANSPORT PROTEIN.

ALEN T. MATHEW[a], AHMAD REZA MEHDIPOUR[a]
[a] Center for Molecular Modelling, Ghent University, Ghent, Belgium

Background: LicB is a bacterial membrane transporter that plays a crucial role in choline-mediated decoration of bacterial cell wall teichoic acids & lipopolysaccharides, which contribute to virulence features, antibiotic resistance, & evasion of host immune system in S. pneumonia. Invitro studies involving inactivation of the licb gene, lead to nonviable S. pneumonia suggesting the therapeutic potential of targeting LicB. Although it is suggested that LicB uses the alternate access model for choline import, the exact mechanism by which LicB imports choline into the bacterial cytoplasmic space is unknown. Currently, we have Cryo-EM & X-ray crystallography structures of LicB in the outward open state & occluded state with choline bound[1].

Objectives: Our objective is to explore the conformational dynamics of LicB during the choline import by reconstructing the free energy landscape of LicB to explain the thermodynamics & kinetics of the transport cycle.

Methods: To achieve our objective, we will define collective variables that can accurately differentiate between the various conformations of the protein during the transport cycle. We will use biomolecular simulations coupled with enhanced sampling techniques to explore the phase space of the transport cycle. Using non equilibrium simulation techniques like steered molecular dynamics (MD) & targeted MD, we will generate intermediate states, such as the inward open conformation. We will use Markov state modelling & transition path sampling to explore the kinetics of the conformational changes[2].

Results: Our 1 μs long unbiased simulations of LicB in both the outward open & occluded states within a lipid bilayer mimicking the composition of S. pneumoniae membrane helped us identify the appropriate collective variable. Using non-equilibrium simulations like steered MD & targeted MD, we generated the outward open conformation bound to choline & inward open conformation which were not yet identified experimentally, from an initial occluded state.

Conclusion: We will continue extending our simulations using experimental structures & those generated using non-equilibrium simulations to sample the phase space thoroughly & reconstruct the free energy landscape of the transport cycle to explain the mechanism of LicB-mediated choline import. Further, we will use Markov state modelling to explain the kinetics of the conformational changes during the choline import process. Our work will shed light on the mechanism of choline import by LicB & pave the way for the development of novel antibiotics targeting LicB.

References

[1] Bärland, N. et al. Mechanistic basis of choline import involved in teichoic acids and lipopolysaccharide modification. Science Advances 8, eabm1122 (2022).

[2] Hénin, J., Lelièvre, T., Shirts, M., Valsson, O. & Delemotte, L. Enhanced Sampling Methods for Molecular Dynamics Simulations [Article v1. 0]. Living Journal of Computational Molecular Science 4, 1583-1583 (2022).

ONE CLASS CLASSIFICATION AND MOLECULAR DYNAMICS IDENTIFY ADRB2 LIGANDS WITH SPECIFIC PHARMACOLOGICAL PROFILES USING A SINGLE REFERENCE LIGAND

LUCA CHIESA, ESTHER KELLENBERGER[a]

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

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

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

Molecular dynamics (MD) simulations were used to study the interactions formed by the reference ligand with ADRB2. Machine learning, specifically one class classification, was used to extract relevant binding information from the simulations.

The models were trained to recognize binding patterns comparable to the ones observed in the simulations, while discarding anomalous ones. The method was tested on small datasets containing well characterized agonists, antagonists, and inactive molecules. Docking was performed on an ensemble of structures extracted from the simulations. Each docking pose was classified by the models as reference-like or not. Properly trained models were able to clearly distinguish between agonist, and antagonists and inactives.

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

References

[1] de Graaf C, Rognan D. Selective Structure-Based Virtual Screening for Full and Partial Agonists of the β2 Adrenergic Receptor, J Med Chem., 2008, 51-4978.

[2] Kooistra AJ, Leurs R, de Esch IJP, de Graaf C. Structure-Based Prediction of G-Protein-Coupled Receptor Ligand Function: A β-Adrenoceptor Case Study, J Chem Inf Model, 2015, 55-1045.

[3] Chiesa L, Kellenberger E, One class classification for the detection of β2 adrenergic receptor agonists using single-ligand dynamic interaction data, Journal of Cheminformatics, 2022, 14-74

P53 REACTIVATION VIA COMPUTATIONALLY DESIGNED BINDERS

RIKA TANDIANA[a], GERMAN PATRICIO BARLETTA[a], SARA FORTUNA[a], WALTER ROCCHIA[a]

[a] Italian Institute of Technology (IIT), Via Melen 83, B Block, Genova, Italy

Mutations occurring in p53 proteins have been demonstrated to disrupt their tumor suppression functionality. Research efforts have been directed towards the restoration of p53 functionality as a potential cancer treatment.[1,2] One of the most frequently occurring mutations is Y220C, which impairs the physiological functioning by destabilizing the protein. In this study, our objective is to design a high affinity nanobody towards p53, binding in such a way to stabilize the structure of p53 and recover its function.

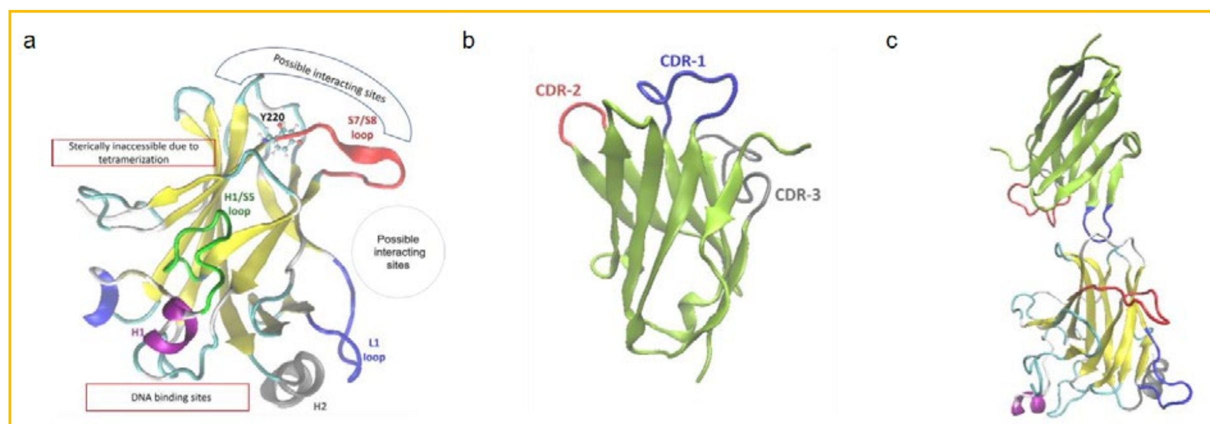


Figure 1: (a) The structure of p53 protein, with Y220 mutation site, L1 loop, H1/S5 loop, and H2 highlighted; (b) The representative structure of VHH with the CDR loops (responsible for antigen recognition and binding) highlighted; (c) The VHH-p53 complex with VHH interacting through one of the CDR loop

First, we studied the effects of Y220C mutation. We performed 2μs of MD simulations of the monomeric p53 WT and of its mutant Y220C (highlighted in Figure 1a) to investigate the structural changes induced by the mutation. Our results suggested increased flexibility in L1, H1/S5 loop, and in S7/S8 loop. Additionally, helix H2 has been observed to lose its helicity as a result of the mutation.

Considering that p53 functions as a tetramer, we identified binding regions far from the p53 tetramerization interface (highlighted in Figure 1a). On those regions, we docked with HADDOCK a series of nanobodies (VHH, Figure 1b) with varying length of CDR3 (between 7 and 18 residues). The resulting poses were initially visually screened based on the binding interfaces, in order to identify the optimum CDR3 length and they underwent 250ns of MD simulation to assess the VHH/p53 complex stability.

Then, selected starting p53/VHH complexes (Figure 1c) underwent PARCE-based optimization to increase the affinity of the VHH.[3,4] PARCE iteratively performs single point mutations on the three CDR residues, followed by minimization and by 30ns of MD simulation. The trajectory is then scored with various scoring functions (namely BLUUES, EvoEF2, PISA, and MMPBSA), to assess whether the mutation results in higher or lower affinity. If the mutation results in better affinity, the mutation is accepted, and the process is repeated until the estimated affinity no longer increases significantly and high affinity VHHs are obtained. These VHHs will be subsequently validated experimentally.

For further optimizing PARCE, benchmark studies are being performed on a set of chain reconstruction methods and available scoring functions with a database of antibody-lysozyme complexes. The information collected from these benchmark studies will be used to improve the efficiency and accuracy of the optimization protocols.

References

[1] Demir, Ö., Barros, E.P., Offutt, T.L., Rosenfeld, M., Amaro, R.E., An integrated view of p53 dynamics, function, and reactivation, Current Opinion in Structural Biology, 2021, 187-194.

[2] Zhang, C., Liu, J., Xu, D., Zhang, T., Hu, W., Feng, Z., Gain-of-function mutant p53 in cancer progression and therapy, J. Mol. Cell. Biol., 2020, 674-687.

[3] Cantarutti, C., Vargas, M.C., Dongmo Foumthuim, C.J., Dumoulin, M., La Manna, S., Fortuna, S., et. al., Insights on peptide topology in the computational design of protein ligands: the example of lysozyme binding peptides, Phys. Chem. Chem. Phys., 2021, 23158-23172.

[4] Soler, M.A., Medagli, B., Fortuna, S., et. al., A consensus protocol for the in silico optimization of antibody fragments, Chem. Commun., 2019, 14043-14046.

EPITOPE IDENTIFICATION USING IN-SILICO APPROACHES, A CASE STUDY: NANOBODIES BINDING TO MGLU5 RECEPTOR

FLORIANE ESHAK[a], ANNE GOUPIL-LAMY[b], FRANCINE ACHER[a]

[a] Université Paris Cité, SPPIN, Paris, France

[b] Biovia Science Council, Dassault Systèmes, Vélizy-Villacoublay, France

Schizophrenia is a mental disorder that affects approximately 1% of the worldwide population. Scientists finally developed a hypothesis that may explain the pathophysiology behind this disorder, which is related to the deficiency of glutamate in the brain. Among glutamate targets, in this study we have focused on metabotropic glutamate receptor 5 (mGlu5). However, targeting mGlu receptors selectively is challenging due to their highly conserved orthosteric binding site. To overcome this, an alternative approach was developed which is the use of nanobodies as allosteric modulators[1] . A nanobody is defined as the variable fragment of the heavy chain-only antibody. Its small size and stability under extreme physical conditions make it an interesting therapeutic agent.

This study is based on previous biological data indicating that nanobody Nb5A is a selective positive allosteric modulator of rat mGlu5. The aim of the study is to identify the epitope of NB5A at rat mGlu5 followed by re-epitoping or re-designing of the nanobody to target human mGlu5. Molecular modeling techniques and artificial intelligence algorithms were used such as homology modeling[2], AlphaFold, IgFold, and ImmuneBuilder to predict the structures of NB5A and rat mGlu5. This was followed by blind rigid-rigid[3] and flexible docking to generate docking poses where two possible epitopes were selected. Among these, only one epitope was identified using molecular dynamics, which was validated through biological experiments.

Finally, the re-eptoping of this nanobody was performed aiming to target both rat mGlu5 and human mGlu5 for therapeutic purposes.

Our study provides insights into the capacity of in-silico approaches and novel artificial intelligence algorithms for epitope identification and re-designing nanobodies for therapeutic purposes.

References

[1] Goudet, C.; Rovira, X.; Rondard, P.; Pin, J.-P.; Llebaria, A.; Acher, F. Modulation of Metabotropic Glutamate Receptors by Orthosteric, Allosteric, and Light-Operated Ligands. In Structure and Function of GPCRs; Lebon, G., Ed.; Springer International Publishing: Cham, 2019; pp 253–284.

[2] Sali, A.; Blundell, T. L. Comparative Protein Modelling by Satisfaction of Spatial Restraints. J Mol Biol 1993, 234 (3), 779–815.

[3] Chen, R.; Li, L.; Weng, Z. ZDOCK: An Initial-Stage Protein-Docking Algorithm. Proteins 2003, 52 (1), 80–87.

FINDING HITS IN LARGE CHEMICAL SPACES BY COMBINING DOCKING WITH DEEP LEARNING

DAVID RINALDO[a], KUN YAO[b], MATTHEW P. REPASKY[c], KARL LESWING[b], ROBERT ABEL [b], STEVEN V. JEROME[d]

[a] Schrödinger GmbH, Glücksteinallee 25, 68163 Mannheim, Germany
[b] Schrödinger, Inc., New York, New York 10036, United States
[c] Schrödinger, Inc., Portland, Oregon 97239, United States
[d] Schrödinger, Inc., San Diego, California 92121, United States

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

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

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

References

[1] Y. Yang et al., Efficient Exploration of Chemical Space with Docking and Deep Learning, J. Chem. Theory Comput., 2021 17, 7106-7119

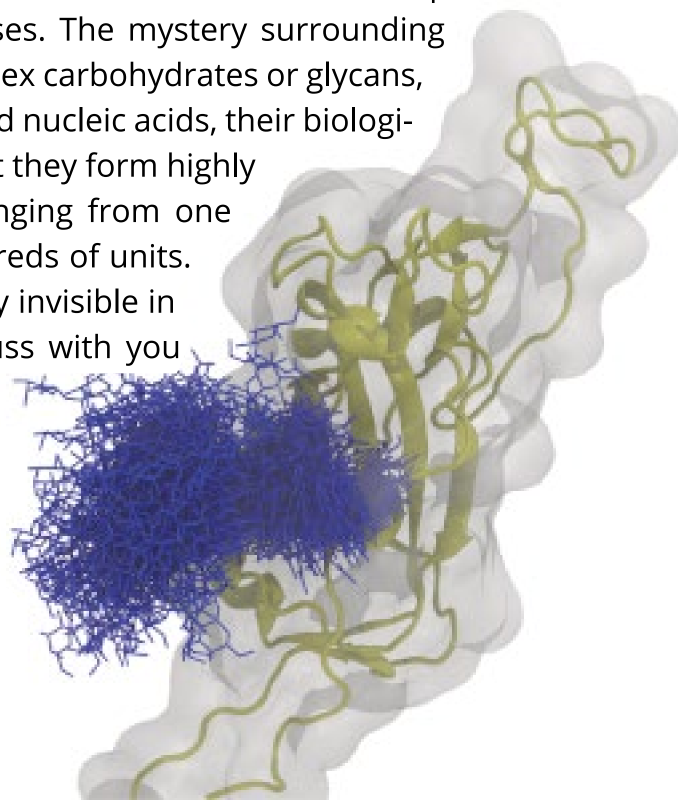
[2] K. Konze et al., Reaction-Based Enumeration, Active Learning, and Free Energy Calculations To Rapidly Explore Synthetically Tractable Chemical Space and Optimize Potency of Cyclin-Dependent Kinase 2 Inhibitors, J. Chem. Inf. Model. 2019, 59, 3782-3793.

Pr. Elisa Fadda

Department of Chemistry and Hamilton Institute, Maynooth University, Ireland

SIMULATIONS OF COMPLEX CARBOHYDRATES (GLYCANS) AND GLYCOPROTEINS, GLYCANS NATURE, STRUCTURE, DYNAMICS AND FUNCTIONS, GLYCAN-PROTEIN RECOGNITION AND INTERACTIONS

Glucose is probably the most widely known sugar, generally recognized for its role in diet as the fundamental component of starch. Yet, the complexity of sugars in biology vastly exceeds cell metabolism, affecting virtually all aspects of the cell life cycle and of its interactions with the matrix, other cells and with pathogens, such as bacteria, toxins and viruses. The mystery surrounding sugars, less colloquially referred to as complex carbohydrates or glycans, derives from the fact that unlike proteins and nucleic acids, their biological synthesis is not template-driven and that they form highly flexible branched polymeric structures, ranging from one monosaccharide to systems counting hundreds of units. These two characteristics make them largely invisible in structural biology studies. Here I will discuss with you how high-performance computing can provide this missing fundamental information, allowing us to understand glycan structure and recognition, and thus their many biological functions in health and disease.



NOTES

DEEP LEARNING BASED PREDICTION OF PROTEIN-CARBOHYDRATE INTERFACES

ARIA GHEERAERT[a], RAVY LEON FOUN LIN[a], THOMAS BAILLY[a], YANI REN[a], YANN VANDER MEERSCHE[a], GABRIEL CRETIN[a], JEAN-CHRISTOPHE GELLY[a], TATIANA GALOCHKINA[a]

[a] Université Paris Cité and Université des Antilles and Université de la Réunion, INSERM, BIGR, F-75014 Paris, France

Background. Protein-carbohydrate interactions are found in multiple essential biological pathways, notably cell adhesion and host-pathogen infection. Unfortunately, due to the chemical and structural variability of carbohydrates and the low affinities of protein-carbohydrate interactions, experimental resolution of protein-carbohydrate interfaces are particularly challenging and protein-carbohydrate interfaces are largely underrepresented in the Protein Data Bank. Therefore, there is an interest in developing accurate theoretical tools able to predict which parts of a protein are susceptible to carbohydrate binding.

Objectives. The aim of this study is to predict carbohydrate binding residues in proteins using deep-learning models.

Methods. We have developed two models: the first uses only information inferred from the protein sequence, while the second also incorporates structural information. In the two models, amino acid features are generated using embeddings derived from the pre-trained protein language model ESM-2[1]. In the second model, if a structure or a model of the protein is provided, the protein is represented as an amino acid contact graph, i.e. a graph whose vertices are amino acids which are related if the two amino acids satisfy a geometric contact condition. In the first model, we use a convolutional neural network to predict if a residue binds to carbohydrates, while for the latter, we benefit from the data structure to use a graph convolutional neural network.

Results. Both of our models outperform the existing carbohydrate binding residue[2] and non-specific binding residue prediction tools[3]. Moreover, using these models, we were able to detect carbohydrate binding residues missed by other methods in some biologically relevant proteins.

Conclusions. The methodology developed here can assist in understanding fundamental biochemical mechanisms, which may also have a major impact in drug design.

References

[1] Zeming Lin et al. Evolutionary-scale prediction of atomic level protein structure with a language model. bioRxiv, 2022.

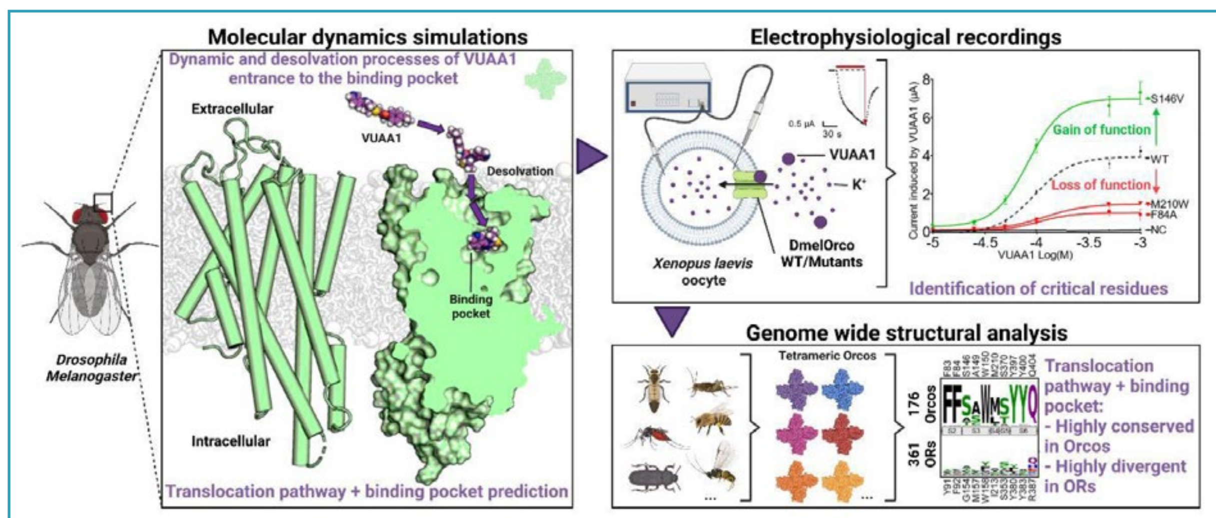
[2] Masaki Banno et al. Development of a sugar binding residue prediction system from protein sequences using support vector machine. Comput. Biol. Chem., 66:36–43, 2017.

[3] Maria Littmann et al. Protein embeddings and deep learning predict binding residues for various ligand classes. Sci. Rep., 11(1):1–15, 2021

LIGAND OF THE CONSERVED INSECT ODORANT RECEPTOR CO-RECEPTOR REACHES ITS BINDING SITE THROUGH A DYNAMIC TRANSLOCATION PATHWAY

**JODY PACALON[a,1], GUILLAUME AUDIC[b,1], JUSTINE MAGNAT[b]
MANON PHILIP[b], JÉRÔME GOLEBIOWSKI[c]
CHRISTOPHE J. MOREAU[b,*], JÉRÉMIE TOPIN[a,*]**

[a] Université Côte d'Azur, Institut de Chimie de Nice UMR7272, CNRS, France
[b] Univ. Grenoble Alpes, CNRS, CEA, IBS, F-38000 Grenoble, France
[c] Department of Brain & Cognitive Sciences, DGIST, 333, Techno JungAng, Daero, HyeongPoong Myeon, Daegu, 711-873, Republic of Korea
[*] Jérémie Topin, 28 avenue Valrose, 06108 Nice France, +33 (0)4 89 15 01 32, jeremie.topin@univ-cotedazur.fr; Christophe Moreau, 71, avenue des Martyrs CS10090 F-38044 Grenoble cedex 9 France, +33 457 428 579, Christophe.moreau@ibs.fr.



Insects are of major importance for our society, being both beneficial for agriculture and detrimental for human health as pathogen vectors. Olfaction is an essential sense for insects, notably for food and host seeking. In numerous insects, the olfactory receptor family forms a unique class of heteromeric cation channels with two subunits that evolved in opposite directions. The signal-generating subunit (Orco) is extremely

conserved across species, while the odorant-binding subunit (OR) diverged to recognize specific ligands present in the environment of the insect. Despite this divergent evolution, ORs have remarkably preserved their ability to interact with Orcos, even from different species. Due to its high degree of conservation, Orco is an attractive target for the development of repellents with very broad-spectrum effects. Recent advances have revealed the homomeric structures of both an «ancient» OR and an Orco without the OR subunits. Unexpectedly, these structures in apo or ligand-bound states did not reveal the pathway taken by the ligands between the extracellular space and the deep internal cavities that have been identified as ligand binding sites in MhOR5. Using a combination of dynamic simulations and structure-function approaches, this article highlights: i) the original molecular entry mechanism of a ligand (VUAA1) into an Orco, which involves a process of dehydration of the compounds; and ii) the ligand binding site of VUAA1 in the Orco. These mechanisms are potentially common to a very large variety of insect species including winged insects.

IONS-RNA/DNA INTERACTIONS: A SIMULATION CHALLENGE

JULIE PUYO-FOURTINE[a] , ELISE DUBOUÉ-DIJON[a]
[a] Laboratoire de Biochimie Théorique, UPR 9080 IBPC, 13 rue Pierre et Marie Curie
puyo@ibpc.fr

Ions are present in biological systems and are involved in numerous biological processes, such as RNA folding, enzyme-based rRNA maturation or the catalytic reaction of ribozymes. Properly describing the interactions between ions (Mg2+, Ca+, Na+) and nucleic acids Is thus key to study such processes. However, it is very challenging for simulations and these interactions are not properly captured by «standard» force fields, which do not account for electronic polarization [1], leading to an overestimation of the binding energies between ions and interacting oxygen atoms.

To overcome this, force fields taking into account electronic polarization have been developed. They fall into two families: polarizable force fields, such as AMOEBA or Drude explicitly model electronic polarization ; others only implicitly account for polarization through either pair-specific Lennard-Jones or scaled charges (Electronic Continuum Correction) models.

In this work, we first investigate how different force fields capture the interactions between a model system for nucleic acid backbone, dimethyl phosphate (DMP), and magnesium and calcium ions [2]. We show that implicitly polarized force field are very promising, providing a much improved interaction with no additional computational cost [3]. Second, we focus on a major limitation of ECC that requires a rescaling of the charges of the RNA and/or DNA [4-5]. Hence, on model sequences, we examine the impact of such rescaling on the nucleic acid conformation, with the goal to develop a scaled charge nucleic acid force field.

References

[1] Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. J. Comput. Chem. 2004
[2] J.Puyo-Fourtine, M.Juillé, J.Hénin, C.Clavaguéra, E.Duboué-Dijon. JPCB. 2022
[3] Cruz-León, S.; Grotz, K.; Schwierz, N. J. Chem. Phys. 2021
[4] Duboué-Dijon, E.; Javanainen, M.; Delcroix, P.; Jungwirth, P.; Martinez-Seara, H. J. Chem. Phys. 2020
[5] Leontyev, I. V.; Stuchebrukhov, A. J. Chem. Phys. 2009

EXPLORING THE RELATIONSHIP BETWEEN MICROTUBULE ARCHITECTURE AND MECHANICS THROUGH NETWORK-BASED MULTISCALE ANALYSIS OF TUBULIN DYNAMICS

MARCO CANNARIATO[a], ERIC A. ZIZZI[a], MARCO A. DERIU[a]
[a] PolitoBIOMed Lab, Department of Mechanical and Aerospace Engineering, Politecnico di Torino, Turin, Italy

The study of microtubule (MT) mechanics and the pathways involved in the transfer of vibrations between tubulins is crucial to understand how MTs are stabilized within the cell. Indeed, the hierarchical organization of MTs is the basis of their stability, mechanics, and function[1]. MTs mechanics may depend on tubulin interactions, protofilament (PF) number, or the presence of stabilizers such as the anti-cancer drug Taxol. Therefore, this work aims at shedding light on the relationship between MT architecture and mechanics, considering the effect of Taxol and the number of PFs. We employed a multiscale approach integrating all-atom molecular dynamics (MD) simulations, Protein Structure Networks (PSN), and Normal Mode Analysis (NMA) on Elastic Network Models (ENM). All-atom structures of the dimers, at different PF numbers and in presence of Taxol, were simulated in systems representative of the MT wall. Through the PSN, we studied the propagation of vibrations between tubulins within the MT. MTs of lengths between 250 nm and 400 nm were built by fitting tubulin conformations from MD simulations onto an Electron Microscopy Density (EMD) map and replicating them axially with an experimental step[2]. Mechanical properties were then derived by MT vibrational frequencies from NMA of ENM, where residues were connected if closer than 1.2 nm[3]. The results highlight that β-tubulin drives the transfer of vibrations between PFs, information relatable to previous evidence1, and that such communication is altered in presence of Taxol. Moreover, Taxol stabilizes the MT by

reducing the fluctuations of the tubulin dimer bending angle. At a higher scale, our results revealed that Taxol influences mainly the shear modulus, reflecting the atomistic result that Taxol alters the inter-PFs interaction. Moreover, MTs with different PF numbers were characterized by changes in mechanical properties, with the persistence length being linearly correlated with the number of PFs. Finally, we observed remarkable differences comparing the mechanics of MTs built with tubulins derived from simulations both at the proper wall curvature and at the 13PF curvature, highlighting the importance of the multiscale approach coupling MD with ENM. Taken together, we provide additional information to understand the nanomechanics of MT and how it can be altered to achieve its stabilization, which could be exploited for the development of biomimetic materials and to guide the design of new MT-stabilizing drugs.

References

[1] Janke et al, The tubulin code and its role in controlling microtubule properties and functions, Nat Rev Mol Cell Biol, 2020, 21:307-326.

[2] Sui et al, Structural Basis of Interprotofilament Interaction and Lateral Deformation of Microtubules, Structure, 2010, 18:1022-1031.

[3] Deriu et al, Anisotropic Elastic Network Modelling of entire Microtubules, Biophysical Journal, 2010, 99:2190-2199

MODEL SUGARS RIGHT: IMPROVED INTERACTIONS OF GLYCANS IN ALL-ATOM MOLECULAR DYNAMICS SIMULATIONS

DENYS BIRIUKOV[a,b], MIGUEL RIOPEDRE FERNÁNDEZ[b], HECTOR MARTINEZ-SEARA[b]

[a] CEITEC – Central European Institute of Technology, Masaryk University, Kamenice 5, 625 00 Brno, Czech Republic

[b] Institute of Organic Chemistry and Biochemistry of the Czech Academy of Sciences, Flemingovo namesti 2, 160 00 Prague, Czech Republic

Glycans are ubiquitously present in various organisms and tissues, particularly in the cellular environment of mammals and bacteria. They naturally occur as building blocks in larger biomolecules, such as glycolipids and glycoproteins, and also constitute huge polysaccharides called glycosaminoglycans. Glycosaminoglycans (GAGs), negatively charged sugar polymers located above the cell membrane, form the scaffold of the extracellular matrix. Due to numerous carboxyl and sulfonate groups present along polysaccharide chains, the biological function and binding affinities of GAGs are presumably closely connected to electrostatic interactions with other biomolecules present in the extracellular space, such as ions, peptides, proteins, and lipids. Moreover, given the high occurrence of GAGs and glycans in general, novel drugs can be specifically designed to target unique glycan sequences of pathogens.

Molecular dynamics (MD) simulations are an excellent tool to rationalize the molecular mechanisms of how glycans recognize other molecules, particularly via mostly uncharacterized sulfation-specific interactions. However, most all-atom MD force fields are known to suffer from the inconsistent description of electrostatic interactions due to missing electronic polarizability. This work summarizes our recent efforts to improve MD models of glycans using the "charge-scaling" approach, which accounts for the electronic polarization in a mean-field way[1]. We first compared available MD force fields for glycans and functional groups common for GAGs and elucidated their main problems. Then, we developed implicitly polarizable, charge-scaled models that outperform other force fields in the description of electrostatic interactions avoiding unphysical overbinding and overestimated contact ion pairing. Finally, using newly derived models, we investigated sugar-ion and sugar-peptide interactions as exemplifying electrostatic interactions. On the one hand, we revealed that short arginine-containing peptide sequences have a higher affinity for hyaluronan (the only non-sulfated GAG) compared to lysine-containing peptides, in agreement with the NMR experiments[2]. On the other hand, we showed that solvent-shared ion pairing is the dominant binding mode of calcium cations to sulfonates, the prevailing functional group of sulfated GAGs, which was supported by ab initio molecular dynamics simulations.

Our results highlight that currently available MD force fields for glycans can be notably improved, especially if electrostatic interactions are of interest. Such improvement can be achieved using the bottom-up approach when the principal building blocks are optimized first, and then large-scale biosimulations involving glycan-containing molecules are carried out. The accurate simulation models for glycans are vital for basic research and potential medical applications, including designing novel antibacterial and antiviral therapeutics.

References

[1] Nencini, R.; Tempra, C.; Biriukov, D.; Polák, J.; Ondo, D.; Heyda, J.; Ollila, S. O.; Javanainen, M.; Martinez-Seara, H. Prosecco: polarization reintroduced by optimal scaling of electronic continuum correction origin in MD simulations, Biophys. J. 2022, 121, 157a.

[2] Fernández, M. R.; Biriukov, D.; Dračínský, M.; Martinez-Seara, H. Molecular fingerprint in hyaluronan-peptide interactions: Cooperation of electrostatics and side-chain specificity in enhanced arginine effect, Biophys. J. 2023, 122, 20a-21a.

HGSCORE: ARTIFICIAL INTELLIGENCE FOR STRUCTURE-BASED MOLECULAR DOCKING SCORING

**KEVIN CRAMPON[a,b,c], ALEXIS GIORKALLOS[a], XAVIER VIGOUROUX[a]
LUIZ ANGELO STEFFENEL[b], STEPHANIE BAUD[c]**

[a] Center for Excellence in Advanced Computing, Atos SE, Echirolles, France
[b] LICIIS, Université de Reims Champagne-Ardenne, Reims, France
[c] UMR CNRS/URCA 7369 MEDyC, Université de Reims Champagne-Ardenne, Reims, France

Protein-ligand interactions are involved in many biological processes, such as protein inhibition or the appearance of ligand side effects. Experimental methods are expensive and time-consuming. To reduce both, industry and researchers use in silico methods as filters to preselect only molecules of interest for further experiments.

One way to do this is to use structure-based molecular docking, a highly complex process, combining spatial and conformational exploration and complex affinity assessment by predicting the free energy of binding. Although molecular docking simulations are less time-consuming than experimental methods, the computational time is not negligible. That is why numerous searches aim at reducing this time by using artificial intelligence (AI) and especially deep learning (DL). In addition, numerous other searches aim to improve molecular docking results by using deep learning[1].

Indeed, even if, scoring is not the most time-consuming part, a better scoring function will allow better molecular docking simulations. Here, we present a new deep learning scoring function: HGScore. This method, designed to study a protein-ligand complex uses the power of the graph representation and graph convolutional neural network (GCNN)[2]. Thus, each complex is represented by a heterogeneous graph (a graph composed of two sub-graphs with an additional set of edges connecting them) of the complex's heavy atoms. In addition, all nodes and edges are augmented by a set of physicochemical features: for instance, the atom type, its hybridization for the nodes; the type of covalent bond, and the nature of the molecular interaction for the nodes.

Our method has the originality of splitting the learning regarding the type of binding (covalent protein-protein binding, covalent ligand-ligand binding, atomic protein-ligand interaction, and ligand-protein atomic interaction) by extending the AttentiveFP[3] model. Then we treat the two molecules with another GCNN allowing to produce two distinct fingerprints (two 1D vectors). Finally, our model uses these two fingerprints to produce an affinity score.

Trained on the PDBBind dataset[4] (a dataset of experimental structures and protein-ligand affinities) and tested on the CASF 2016 benchmark[5], our method achieves a good level of performance, outperforming the classical scoring function (GoldScore or ChemScore) and is among the best AI methods.

References

[1] Crampon, K.; Giorkallos, A.; Deldossi, M.; Baud, S.; Steffenel, L. A. Machine-Learning Methods for Ligand–Protein Molecular Docking. Drug Discov. Today 2022, 27 (1), 151–164. https://doi.org/10.1016/j.drudis.2021.09.007.

[2] Fey, M.; Lenssen, J. E. Fast Graph Representation Learning with PyTorch Geometric; arXiv:1903.02428; arXiv, 2019. http://arxiv.org/abs/1903.02428 (accessed 2022-05-16).

[3] Xiong, Z.; Wang, D.; Liu, X.; Zhong, F.; Wan, X.; Li, X.; Li, Z.; Luo, X.; Chen, K.; Jiang, H.; Zheng, M. Pushing the Boundaries of Molecular Representation for Drug Discovery with the Graph Attention Mechanism. J. Med. Chem. 2020, 63 (16), 8749–8760. https://doi.org/10.1021/acs.jmedchem.9b00959.

[4] Liu, Z.; Su, M.; Han, L.; Liu, J.; Yang, Q.; Li, Y.; Wang, R. Forging the Basis for Developing Protein–Ligand Interaction Scoring Functions. Acc. Chem. Res. 2017, 50 (2), 302–309. https://doi.org/10.1021/acs.accounts.6b00491.

ROUND TABLE

OPPORTUNITIES FOR YOUNG MODELERS IN COMPANIES

A DYNAMIC ROUND TABLE DISCUSSION ON THE OPPORTUNITIES AVAILABLE TO YOUNG MODELERS WITHIN COMPANIES.

INVITED SPEAKERS

ALAIN MARTY
Carbios (Scientific Director, CSO)

CHRISTOPHE BOLDRON
Evotec (Vice-President, Head of Molecular Architects Toulouse)

JEAN-PHILIP PIQUEMAL
Qubit Pharmaceuticals (CSO, Co-founder)

KATIA DEKIMECHE & **DAVID RINALDO**
Schrodinger (Director & Applications Scientists)

JELENA VUČINIĆ

MOLECULAR MODELING AND ARTIFICIAL INTELLIGENCE FOR COMPUTATIONAL PROTEIN DESIGN: CONCEPTION OF OPTIMIZED ENZYMES AND NANOBODIES

Computational Protein Design (CPD) plays a critical role in advancing the field of protein engineering: it accelerates the delivery of novel proteins that display high specificity, high efficiency and better stability. In its essence, CPD is an optimization problem: using an all-atom energy function and a reliable search method, CPD targets the identification of amino acid sequences that fold into a target structure and ultimately perform a desired function. During my PhD, I proposed the Positive Multistate Protein Design method (POMPd) which simultaneously considers several conformational states of the protein during design. POMPd was validated on environmental, health and fundamental biology applications.

In this talk I will present how we successfully designed an enzyme widely used in industrial bio-refinery processes by improving its thermal stability and catalytic activity but also a stable stable cysteine-free nanobody scaffold that could be expressed as an intrabody and be stable in the reducing cytoplasmic environment. I will also briefly show how our method was used to reverse-engineer an ancestor of RNA polymerase, a work which provides insights on the origins of life.

MERVEILLE EGUIDA

COMPARISON OF PROTEIN CAVITIES BY POINT CLOUD PROCESSING: PRINCIPLES AND APPLICATIONS IN DRUG DESIGN

Protein cavities are the heart of molecular interactions that trigger and regulate biological processes in living organisms. Supported by the constant augmentation of characterized pockets in three-dimensional (3D) protein structures, methods to assess the similarity between protein cavities have multiple applications in drug design but face many challenges.

We herein propose new methods based on 3D image processing to compare global and subtle patterns in different protein (sub-) pockets. Through prospective applications validated by in vitro biological experiments, we showed how these methods can be used to predict a secondary binding target at the proteome scale, to design target-focused libraries for efficient small molecule hit identification, even in the absence of known binders, while opening perspectives for pharmacophore elaboration and alternative virtual screening strategies.

POSTER SESSION

I

Mon. 15/05

18h-20h

Poster **n°1** to **n°59**

LIGAND FLIP-FLOP ON A GPCR SURFACE

Cristina Gil Herrero[a b]
Sebastian Thallmair[a]

[a] Frankfurt Institute for Advanced Studies, Ruth-Moufang-Straße 1, 60438 Frankfurt am Main (Germany)
[b] Goethe University of Frankfurt, Frankfurt am Main (Germany)

G-protein-coupled receptors (GPCRs) are the main family of transmembrane proteins that regulate most of our physiological responses to external stimuli[1 2]. Therefore, they play a crucial role as therapeutic targets, being the most targeted protein family by approved drugs[3]. Salmeterol[4] and salbutamol[5] are two small-molecule drugs[6], which are involved in the treatment of pulmonary diseases by triggering the activation of the GPCR β2-adrenergic receptor (β2AR).

In this study, we used coarse-grained *Martini 3*[7] molecular dynamics simulations to explore the behaviour of the two agonists in the presence of their target protein, β2AR. Firstly, we performed simulations of the parametrized ligands together with a pure POPC membrane and then, also including the protein model.

Our results show that the ligands exploit the protein surface to change between membrane leaflets and thus, to pass through the membrane more easily. Even though leaflet changes, so-called "flip-flops", already take place in the absence of the protein, the hydrophilic substituents at the common ring structure of salmeterol and salbutamol provide an unfavourable situation for lipid membrane crossing. We found that the flip-flop frequency raises considerably in the presence of the protein, being 70% the flip-flops on the β2AR surface for salmeterol and 50% for salbutamol, in unbiased trajectories.

To shed light on this phenomenon, we analyzed the ligand flip-flop location on the β2AR surface, finding that the most populated transmembrane domain by the ligands is different from the one where lipid flip-flop is reported in experiments[8].

In addition, umbrella sampling calculations of the ligand flip-flops were performed along the most populated path, using as initial configurations snapshots extracted from free simulations. The resultant potentials of mean force (PMFs) exhibit a significant drop in the energy barrier *of more than 15 kJ/mol with slight differences in both flip-flop directions and between both drugs. Moreover, the free energy calculations were replicated exchanging the ligands to get their PMFs when describing the other ligand's flip-flop path. Although some distinctions are still found between the two ligands, very similar trends are depicted when the same flip-flop pathway is followed, indicating that the path is the most relevant feature influencing the energetics of the flip-flop.

Our findings provide insights into the mechanism of ligand permeation and could contribute to the design of more effective drugs that require cell permeation.

References

[1] D. M. Rosenbaum, S. G. Rasmussen, and B. K. Kobilka, *"The structure and function of g-protein-coupled receptors,"* Nature, vol. 459, no. 7245, pp. 356–363, 2009.

[2] F. Yang, S. Ling, Y. Zhou, Y. Zhang, P. Lv, S. Liu, W. Fang, W. Sun, L. A. Hu, L. Zhang, et al., *"Different conformational responses of the β2-adrenergic receptor-Gs complex upon binding of the partial agonist salbutamol or the full agonist isoprenaline,"* National Science Review, vol. 8, no. 9, 2021.

[3] Sriram, Krishna, and Paul A. Insel. *"G protein-coupled receptors as targets for approved drugs: how many targets and how many drugs?."* Molecular pharmacology 93.4 (2018): 251-258.

[4] M. Johnson, P. Butchers, R. Coleman, A. Nials, P. Strong, M. Summer, C. Vardey, and C. Whelan, *"The pharmacology of salmeterol,"* vol. 52, pp. 2131–2143, jan 1993.

[5] Price, Allan H., and Stephen P. Clissold. *"Salbutamol in the 1980s: a reappraisal of its clinical efficacy.»"*Drugs 38.1 (1989): 77-122.

[6] Lötvall, J. *"Pharmacological similarities and differences between β2-agonists."* Respiratory medicine 95 (2001): S7-S11.

[7] P. C. Souza, R. Alessandri, J. Barnoud, S. Thallmair, I. Faustino, F. Grünewald, I. Patmanidis, H. Abdizadeh, B. M. Bruininks, T. A. Wassenaar, et al., *"Martini 3: a general purpose force field for coarse-grained molecular dynamics,"* Nature methods, vol. 18, no. 4, pp. 382–388, 2021.

[8] Khelashvili, George, and Anant K. Menon. *"Phospholipid Scrambling by G Protein–Coupled Receptors."* Annual Review of Biophysics 51 (2022): 39-61.

APPLICATION OF BAYESIAN OPTIMIZATION IN DRUG DISCOVERY

COLLIANDRE LIONEL[1]
MULLER CHRISTOPHE[1]
LONG JINTAO[2]
HRISTOZOV DIMITAR[2]
BODKIN MICHAEL[2]

[1] Evotec SAS, Campus Curie, 195 route d'Espagne, 31100 Toulouse, France
[2] 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. This is usually achieved through the finding of hit compounds that are latter sequentially optimized toward the expected profile inside a Design-Make-Test-Analyze (DMTA) cycle. The main question in this trial-and-error process is "which compound should I make and test next?". For any project, the answer to this question depends on the knowledge that has been acquired so far and on the allocated resources.

Bayesian optimization (BO) is an efficient method to perform a "black box" optimization of an unknown objective function [2]. It refers to a sequential strategy that look for the global optimum of the objective function. It is driven by the gathering of new information that will help improve the current knowledge to be used in the next step. The heart of BO is a Machine Learning (ML) model that is trained with the data acquired so far. Then, at each step, new points are selected such as to provide the maximum amount of information over the objective landscape. The objective function is evaluated at these new points and the results are fed back to the ML model. Hence, the process can choose the next points 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. Most molecular descriptors only support 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 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 guide the design cycle in an active learning loop by scoring a pre-defined list of molecules and/or generating new molecules by decoding the optimal descriptor vector found at each step.

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

References

[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

[7] Winter R. et al. Learning continuous and data-driven molecular descriptors by translating equivalent chemical representations, 2019, Chem. Sci., 10, 1692-1701

[8] Griffiths R. et Hernandez-Lobato J.M., Constrained Bayesian optimization for automatic chemical design using variational autoencoders, 2020, Chem. Sci., 11, 577-586

ANALYSIS AND PREDICTION OF PROTEIN-RNA INTERACTION NETWORKS

CARLA MARTINS[a]
JESSICA ANDREANI[a]

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

Protein-RNA interactions are central to a wide range of cellular functions. Nevertheless, the molecular mechanisms involved are still to be understood[1] and protein-RNA binding is less well understood than binding between two proteins. This is reflected by the number of structures present in the PDB which is about 110 000 for protein-protein interactions and about 4300 for protein-RNA interactions. Recently, an AI-based tool, AlphaFold[2], has revolutionized the prediction of protein structures and complexes with high reliability. However, this tool is not yet able to predict interactions between proteins and other molecules. This limitation, together with the small number of protein-RNA interaction structures, makes it necessary to study this type of interaction in order to better characterize and predict it. My work aims at a better molecular understanding of protein-RNA interactions in order to overcome all these limitations. We aim to exploit abundant large-scale omics data to leverage the small amount of available structural data. In the present study, I combined (1) "omics" data on protein-RNA interactions obtained by different techniques (*in vivo* and *in vitro*) and (2) structural data of complexes from PDB. From the omics data, I identified the most likely RNA binding motifs obtained for each method. In addition, we gathered all the information and combined it to analyze the consistency between the methods as well as to determine the different RNA binding motifs for each protein. The majority of the binding motifs match canonical motifs listed in the database EuRBPDB[3]. Overall, we observed that in vitro techniques generated clearer binding motifs. For this reason, I performed subsequent sequence and structure cross-analyses on a subset of data with motifs obtained by in vitro methods, which corresponds to nearly 100 proteins for which about 500 protein-RNA interface structures are available. For this subset, I am currently analyzing how closely the RNA sequence interacting with the protein in the 3D structure matches the omics binding motif. A next step in my work will be to extend the motif analysis to omics data including organisms other than humans. On the longer term, our team aims to develop a computational tool taking omics data into account to predict protein-RNA interaction structures.

References

[1] Helder S, Blythe AJ, Bond CS, Mackay JP. Determinants of affinity and specificity in RNA-binding proteins. Current Opinion in Structural Biology. 2016;38:83-91. doi:10.1016/j.sbi.2016.05.005

[2] Jumper J, Evans R, Pritzel A, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596(7873):583-589. doi:10.1038/s41586-021-03819-2

[3] Liao JY, Yang B, Zhang YC, et al. EuRBPDB: a comprehensive resource for annotation, functional and oncological investigation of eukaryotic RNA binding proteins (RBPs). Nucleic Acids Research. 2020;48(D1):D307-D313. doi:10.1093/nar/gkz823

DETAILED ANALYSIS OF A THERMOSTABLE PROTEIN-DNA COMPLEX: THE CASE OF SAC7D AS A PROTOTYPE FOR PROTEIN-DNA PROTECTION

ELENA ALVAREZ SANCHEZ [a,b]
SIMON HUET[b]
BERNARD OFFMANN[a]
STÉPHANE TÉLETCHÉA[a]
[a] address 1: Nantes Université, US2B, CNRS, UMR6286, F-44000 Nantes, France
[b] address 2: Affilogic, 24 rue de la Rainière, F-44000 Nantes, France

Sac7d is a 7kDa protein belonging to the class of the small chromosomal proteins from archeon Sulfolobus acidocaldarius. Sac7d was discovered in 1974 in Yellowstone National Parks geysers, and studied extensively since then for its remarkable stability at large pH and temperature ranges. Sac7d binds to the DNA minor groove by raising its melting temperature, thus protecting DNA from this extreme conditions [1 2 3]. In this study, we analyzed Sac7d-DNA complex using 1µs molecular dynamics simulations to determine which amino acids contributed most to DNA binding. Energy interaction decomposition of the interface was performed using Molecular Mechanics Generalized Born Surface Area (MM/GBSA). We determined that more than 10 amino acids were critical for Sac7d/DNA recognition. The individual contribution of each amino acids to the binding interface was in agreement with previous explorations of the interface. We provide a novel in-depth focus on the DNA energetics consequence of sac7d tethering, with local DNA transitions occuring within the molecular dynamics simulations.

References

[1] Robinson H, Gao YG, McCrary BS, Edmondson SP, Shriver JW, Wang AH. The hyperthermophile chromosomal protein Sac7d sharply kinks DNA. Nature. 1998 Mar 12;392(6672):202-5. doi: 10.1038/32455. PMID: 9515968.

[2] Singh RK, Mukherjee A. Molecular Mechanism of Dual Intercalation in Sac7d-DNA Complexation. J Phys Chem B. 2022 Mar 3;126(8):1682-1690. doi: 10.1021/acs.jpcb.1c09355. Epub 2022 Feb 20. PMID: 35184562.

[3] Zacharias M. Atomic Resolution Insight into Sac7d Protein Binding to DNA and Associated Global Changes by Molecular Dynamics Simulations. Angew Chem Int Ed Engl. 2019 Apr 23;58(18):5967-5972. doi: 10.1002/anie.201900935. Epub 2019 Mar 26. PMID: 30767392.

MOLECULAR MODELING OF NOX5 PROTEIN: IMPACT OF THE INITIAL STRUCTURE AND MEMBRANE COMPOSITION

BAPTISTE ETCHEVERRY[a]
AURÉLIEN DE LA LANDE[a]
MARC BAADEN[b]
FABIEN CAILLIEZ[a]

[a] Institut de Chimie Physique, Université Paris-Saclay, CNRS (UMR 8000), 15 avenue Jean Perrin 91405, Orsay, France.
[b] CNRS, Université de Paris, UPR 9080, Laboratoire de Biochimie Théorique, 13 rue Pierre et Marie Curie, F-75005, Paris, France.

Electron Transfer within proteins is an essential process for cellular activity. NADPH oxidases are transmembrane proteins the main function of which is the production of reactive oxygen species in many organisms. In the NOX5 isoform, reduction of dioxygen to superoxide ion is carried out, after several steps of electron transfer between redox cofactors (two hemes and a flavin) across the cell membrane. A previous work in our group[1] using molecular dynamics simulations has addressed the issue of inter-heme electron transfer in NOX5 protein, based on the first experimental structure of a NOX protein.[2] Our objective now is to study with the use of theoretical chemistry methods the behavior of NOX5 protein and the successive electron transfer steps responsible for its activity.

New experimental structures of NOX proteins [3,4,5] have recently been determined using cryo-electron microscopy. These structures exhibit quite large deviation from the previous model, which might have an impact on the electron transfer process. In this work we will present molecular dynamics studies of NOX5 protein, modeled in a biological environment (membrane, counter-ions, water) with the use of molecular force fields. The building of the models of NOX5 will be detailed, as well as the analysis of trajectories of several hundreds of nanoseconds, in each redox state involved in the transmembrane electron transfer steps. We will explore the role of the initial atomic positions (based on the 2 types of experimental data) and of the membrane composition.

The structural analysis of these trajectories will be described, as well as the first calculations of electron transfer parameters in the framework of Marcus theory.[6] Free energy ΔG of electron transfer and reorganization energy will be shown. Comparison of the various simulations will bring insight on the influence of the different simulation conditions on structural and dynamical properties of the protein and on the electron transfer.

References

[1] Wu X, Hénin J, Baciou L, Baaden M, Cailliez F, de la Lande A. Mechanistic Insights on Heme-to-Heme Transmembrane Electron Transfer Within NADPH Oxydases From Atomistic Simulations. Front Chem. 9:650651 (2021). https://doi.org/10.3389/fchem.2021.650651.

[2] Magnani F, Nenci S, Millana Fananas E, Ceccon M, Romero E, Fraaije MW, Mattevi A. Crystal structures and atomic model of NADPH oxidase. Proc Natl Acad Sci U S A. 114(26):6764-6769 (2017). https://doi.org/10.1073/pnas.1702293114.

[3] Sun, J. Structures of mouse DUOX1–DUOXA1 provide mechanistic insights into enzyme activation and regulation. Nat Struct Mol Biol 27, 1086–1093 (2020). https://doi.org/10.1038/s41594-020-0501-x

[4] Wu, JX., Liu, R., Song, K. et al. Structures of human dual oxidase 1 complex in low-calcium and high-calcium states. Nat Commun 12, 155(2021). https://doi.org/10.1038/s41467-020-20466-9

[5] Noreng, S., Ota, N., Sun, Y. et al. Structure of the core human NADPH oxidase NOX2. Nat Commun 13, 6079 (2022). https://doi.org/10.1038/s41467-022-33711-0

[6] de la Lande A, Cailliez F, R Salahub D. Simulating Enzyme Reactivity. RSC Theorical and Computational Chemistry Series. Chapter 4 (2017) https://doi.org/10.1039/9781782626831

STUDY OF THE I431V MUTATION OF DHPS IN PLASMODIUM FALCIPARUM

EMILIE GUÉMAS[a,b]
SANDIE MÉNARD[b]
ANTOINE BERRY[b]
MARIE BRUT[a]

[a] LAAS-CNRS, Université de Toulouse, CNRS, UPS, Toulouse, France.
[b] Institut Toulousain des Maladies Infectieuses et Inflammatoires (Infinity), Université Toulouse, CNRS UMR5051, INSERM UMR1291, UPS, Toulouse, France.

Malaria remains a major public health problem. In addition to protective measures against the mosquito vector, the sulfadoxine-pyrimethamine (SP) combination is used throughout Africa for the intermittent preventive treatment of malaria in pregnancy (IPTp) and the seasonal malaria chemoprevention (SMC) to prevent malaria in children under 5 years. However, the efficacy of this chemoprophylaxis is threatened by the emergence of mutations in the *Plasmodium falciparum* dihydrofolate reductase (*pfdhfr*) and dihydropteroate synthase (*pfdhps*) genes, which code for the enzymes targeted by pyrimethamine and sulfadoxine, respectively. A new mutation in PfDHPS, I431V, was identified in Nigeria in 2007[1] and in Cameroon in 2010[2]. This mutation was mainly found in association with 4 other *Pf*DHPS mutations forming a vagKgs quintuple mutant. The mechanisms of resistance to sulfadoxine are very poorly studied.

We obtained a complete PfDHPS WT structure from a PfDHPS crystal structure (PDB ID: 6JWQ)[3] by filling the missing parts with DHPS structure of *P. falciparum* obtained by homology[4] or that of *P. vivax* (PIDB ID: 5Z79)[5]. Then, 5 PfDHPS mutants of interest, with various combinations of mutations were generated, forming 6 research systems. First, molecular dynamics (MD) simulations were performed with AMBER software, using the ff14SB, gaff and tip3p force fields, for a period of 200 ns. This allow us to

study the conformational properties and dynamic behaviour of the enzyme-substrate complexes with natural substrate (pABA) and inhibitor (sulfadoxine). These molecular dynamics trajectories were analysed to compare the substrate binding free energies, which suggest that the vagKgs quintuple mutant has an increased affinity for pABA. During MD simulations, sulfadoxine was released from the active site in all 6 systems.

Based on these structures obtained from the MD simulations, we will perform hybrid quantum mechanics/molecular mechanics (QM/MM) calculations to refine the understanding of the effects of mutation on active site structure and stability, solvent accessibility and charge distribution.

References

[1] Sutherland CJ, Fifer H, Pearce RJ, et al. Novel pfdhps haplotypes among imported cases of Plasmodium falciparum malaria in the United Kingdom. Antimicrob Agents Chemother 2009; 53: 3405–10.

[2] Chauvin P, Menard S, Iriart X, et al. Prevalence of Plasmodium falciparum parasites resistant to sulfadoxine/pyrimethamine in pregnant women in Yaoundé, Cameroon: emergence of highly resistant pfdhfr/pfdhps alleles. J Antimicrob Chemother 2015; 70: 2566–71.

[3] Chitnumsub P, Jaruwat A, Talawanich Y, et al. The structure of Plasmodium falciparum hydroxymethyldihydropterin pyrophosphokinase-dihydropteroate synthase reveals the basis of sulfa resistance. FEBS J 2020; 287: 3273–97.

[4] Oguike MC, Falade CO, Shu E, et al. Molecular determinants of sulfadoxine-pyrimethamine resistance in Plasmodium falciparum in Nigeria and the regional emergence of dhps 431V. Int J Parasitol Drugs Drug Resist 2016; 6: 220–9.

[5] Yogavel M, Nettleship JE, Sharma A, et al. Structure of 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase-dihydropteroate synthase from Plasmodium vivax sheds light on drug resistance. J Biol Chem 2018; 293: 14962–72.

COMPUTATIONAL STUDIES TOWARDS ENZYMATIC DEGRADATION OF POLYURETHANE

AGATA RACZYŃSKA[a,b,c]
RAJENDRA SHARMA[a]
JÉRÉMY ESQUE[a]
ARTUR GÓRA[b]
ISABELLE ANDRÉ[a]

[a] Toulouse Biotechnology Institute, TBI, Université de Toulouse, CNRS, INRAE, INSA, Toulouse, France. 135, avenue de Rangueil, F-31077 Toulouse Cedex 04, France
[b] Tunneling Group, Biotechnology Centre, Silesian University of Technology, ul. Krzywoustego 8, 44-100 Gliwice, Poland
[c] Faculty of Chemistry, Silesian University of Technology, ul. Strzody 9, 44-100 Gliwice, Poland

Synthetic polymers, commonly known as plastics, are durable materials that possess many desirable features, but their high resistance to biodegradability, once considered an advantage, is now one of the main causes of pollution in terrestrial and aquatic environments. Production of plastics have outgrown most man-made materials; however, the management of their end-of-life is very limited [1].

Biodegradation uses the ability and function of microorganisms to convert organic substrates, such as synthetic polymers, into small, low molecular weight fragments that can then be degraded into carbon dioxide and water [2]. However, because of the short time frame since the introduction of plastics and the widespread presence of other carbon sources, in the course of natural evolution there was not a sufficiently strong pressure aimed at the creation of effective enzymes that degrade synthetic polymers [3]. Instead, redesigned, and improved enzymes represent a promising machinery for treating waste.

In the presented study we have aimed to find and learn about enzymes capable of degrading polyurethanes (PUR), which are a group of commonly used synthetic polymers. We have focused on studying the degradation of a model PUR, Impranil DLN, produced by Covestro. We applied modern computational techniques, including molecular docking, molecular dynamics simulations, protein-ligand interaction analysis and computational enzyme design.

The results allowed us to learn about the molecular aspects responsible for the recognition and binding of polymer chains by enzymes. We were also able to map in detail the amino acids which participate in the binding of PUR fragments to the enzyme surface, as well as those that interfere in the adsorption process. The knowledge gained in this study is then used to propose mutations on the surface of the selected PUR-biodegrading enzyme. Such mutations aim to enhance the model PUR's affinity to the binding site, and thus hopefully increase enzyme's activity. Increasing the affinity of ligands by modifying the surface of the enzyme is an important step in the design of enzymes capable of degrading plastics, as the adsorption process of synthetic polymers is often the rate limiting step of the entire process of enzymatic degradation.

Acknowledgement: *This work is financed by Diamond Grant No. 0169/DIA/2020/49, Campus France and INSA Toulouse. This work was granted access to the HPC resources of the Computing meso-center of Région Midi-Pyrénées (CALMIP, Toulouse, France).*

References

[1] Geyer R, Jambeck JR, Law KL. Production, use, and fate of all plastics ever made. Sci Adv. 2017;3(7):3–8.

[2] Gomiero JAGE-A. Biological Degradation of Polymers in the Environment. In Rijeka: IntechOpen; 2019. p. Ch. 5. Available from: https://doi.org/10.5772/intechopen.85124

[3] Danso D, Chow J, Streita WR. Plastics: Environmental and biotechnological perspectives on microbial degradation. Appl Environ Microbiol. 2019;85(19):1–14.

CP-DFTB/MM SIMULATIONS OF TYROSINE-TYROSINE PCET IN RNR-INSPIRED MODEL SYSTEMS

KATHARINA SPIES[a,b]
NATACHA GILLET[b]
MARCUS ELSTNER[a]
[a] Institute of Physical Chemistry, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany
[b] Univ Lyon, ENS de Lyon, CNRS UMR 5182, Laboratoire de Chimie, F69342 Lyon, France

Proton-coupled electron transfer (PCET) plays an important role in diverse biological processes, involving organic or organo-metallic cofactors. Prominent examples from nature are the electron transport chain in Photosystem II which is essential for the production of ATP or the 32 Å long radical pathway based on tyrosines oxidation in the enzyme Ribonucleotide Reductase (RNR). An extended sampling method was developed in our group in which coupled-pertubed equations are implemented into the computational favorable density functional tight binding method (CP-DFTB).[1] All prior computational studies on PCET rely on multicale simulations with computational high costly QM methods as DFT, we here propose an alternative workflow using CP-DFTB which allows us to obtain a detailed insight into PCET reactions on longer timescales. CP-DFTB enables the use of Mulliken charges as reaction coordinates to calculate free energies of chemical reactions like PCET, accessing thermodynamic and kinetic properties as well as the principal reaction mechanism. The method has been successfully applied to QM and QM/MM setups including two tyrosine side chains in water.[2] We aim to investigate PCET in biological systems, therefore we start testing the CP-DFTB method on small model systems: β-hairpin peptides[3] and α-helical proteins[4], that are inspired by RNR and have been used for simulating PCET reactions before.[5,6] One residue in each model system was mutated so that two tyrosines are located in near vicinity, one being fully reduced and the other in oxidized and deprotonated state. MD and QM/MM simulations were performed on length scales suitable to the experimental lifetimes of the radicals[7,8] and during the simulations the secondary structure of the model systems were preserved. We present here our first CP-DFTB/MM results on a realistic biological system.

References

[1] Natacha Gillet, Marcus Elstner, and Tomáš Kubař. "Coupled-perturbed DFTB-QM/MM metadynam ics: Application to proton-coupled electron transfer". J. Chem. Phys. 2018. P. 072328.

[2] Denis Maag et al. "Mechanism of proton-coupled electron transfer described with QM/MM implementation of coupled-perturbed density-functional tight-binding". J. Chem. Phys. 2023.

[3] Robin Sibert et al. "Proton-Coupled Electron Transfer in a Biomimetic Peptide as a Model of Enzyme Regulatory Mechanisms". J. Am. Chem. Soc. 2007. Pp. 4393–4400.

[4] Bruce W. Berry, Melissa C. Martınez-Rivera, and Cecilia Tommos. "Reversible voltammograms and a Pourbaix diagram for a protein tyrosine radical". Proc. Natl. Acad. Sci. U.S.A. 2012. Pp. 9739–9743.

[5] Astrid Nilsen-Moe et al. "Proton-coupled electron transfer from tyrosine in the interior of a de novo protein: mechanisms and primary proton acceptor". J. Am. Chem. Soc. 2020. Pp. 11550–11559.

[6] Tyler G McCaslin et al. "Structure and Function of Tryptophan–Tyrosine Dyads in Biomimetic β Hairpins". J. Phys. Chem. B. 2019. Pp. 2780–2791.

[7] Starla D Glover et al. "Photochemical tyrosine oxidation in the structurally well-defined α3Y protein: proton-coupled electron transfer and a long-lived tyrosine radical". J. Am. Chem. Soc. 2014. Pp. 14039–14051.

[8] Cynthia V Pagba et al. "Proton-Coupled Electron Transfer and a Tyrosine–Histidine Pair in a Photosystem II-Inspired β-Hairpin Maquette: Kinetics on the Picosecond Time Scale". J. Phys. Chem. B. 2016. Pp. 1259–1272.

COLLECTIVITY IN SELF-ASSEMBLED PROTEIN FILAMENTS INVOLVED IN HOMOLOGOUS RECOMBINATION

AFRA SABEI [a,b]
MARA PRENTISS [c]
CHANTAL PRÉVOST [a,b]

[a] CNRS, Université Paris-Cité, UPR 9080, Laboratoire de Biochimie Théorique, 13 rue Pierre et Marie Curie, F-75005 Paris, France

[b] Institut de Biologie Physico-Chimique, Fondation Edmond de Rothschild, PSL Research University, Paris, France

[c] Department of Physics, Harvard University, Cambridge, MA 02138, USA

Inside the helical nucleofilaments of recombinase proteins belonging to the RecA family, homologous recombination, a crucial procedure for DNA repair, takes place.

These nucleofilaments are created by polymerizing RecA proteins on single-stranded DNA (ssDNA) and then incorporate double-stranded DNA (dsDNA) from the cell's genetic material in a dynamic, multi-step process involving homology search and pairing exchange with the ssDNA.[1]

The three DNA strands that are simultaneously attached to the filament are extremely tight and their differential tension plays a significant role in the strand exchange driving force. The group of Mara Prentiss in Harvard university(USA) recently showed that the length of inserted DNA affects the kinetics of reverse strand exchange, DNA unbinding, or filament disintegration [2]. This prompted us to study the structural and dynamical properties of RecA nucleoprotein filaments hosting varying length of strand exchange product, using all-atom molecular dynamics simulations. From these simulations, we analyzed how the length of inserted DNA influences the network of protein-DNA contacts inside the filament, and the relative positions of the inserted DNA strands. We looked at collective interactions inside the filaments and examined filament deviations from a relaxed shape using the PTools/Heligeom package created in LBT and dedicated to helical assemblies [3].

Our simulation results demonstrate that inter-phosphate distances vary over the bound DNA length and that the bound DNA strand structure is not uniform. In addition, we discovered that the stress diminishes when the entering and exiting dsDNA regions in B-form are absent. Our results imply that binding more than one DNA turn generates local curvature and deformation in the RecA nucleoprotein filament. The length-dependent changes observed in the structure, position, and interactions of thethree DNA strands attached to the filament may result from these departures from a relaxed geometry, which may have an effect on crucial homologous recombination events. In conclusion, our study elucidates how collective interactions within the RecA nucleoprotein filament are influenced by the length of inserted DNA, resulting in helical nucleoprotein filaments with a non-homogeneous internal structure. These discoveries have significant ramifications for our comprehension of the mechanism underlying homologous recombination and could be leveraged to build novel approaches to DNA repair, genetic diseases and cancer.

References

[1] M. Prentiss, C. Prévost, C. Danilowicz. Structure/function relationships in RecA protein-mediated homology recognition and strand exchange.
Critical Reviews in Biochemistry and Molecular Biology. 2015, 50, 453-76.

[2] C. Danilowicz, E. Vietorisz, V. Godoy-Carter, C. Prévost and M. Prentiss. Influences of ssDNA-RecA filament length on the fidelity of homologous recombination. J Mol Biol 2021, 433, 167143

[3] B. Boyer, J. Ezelin, P. Poulain, A. Saladin, M. Zacharias, C. H. Robert and C. Prévost. An integrative approach to the study of filamentous oligomeric assemblies, with application to RecA. PLoS ONE 2015, 10, e0116414

TOWARD THE DESIGN OF A NOVEL TYPE II B-RAF PARADOX BREAKER INHIBITOR

ROHIT ARORA[a], JOANNES T.M. LINDERS[b]
SAMIA ACI-SÈCHE[a], THOMAS VERHEIJEN[b]
ERIKA VAN HEERDE[b], DIRK BREHMER[b]
APIRAT CHAIKUAD[c,d], STEFAN KNAPP[c,d]
PASCAL BONNET[a]

[a] Institut de Chimie Organique et Analytique, UMR CNRS-Université d'Orléans 7311, Université d'Orléans BP 6759, 45067 Orléans Cedex 2, France

[b] Janssen Research and Development, a division of Janssen Pharmaceutica N.V., Turnhoutseweg 30, Beerse 2340, Belgium

[c] Structural Genomic Consortium, University of Oxford, Oxford OX3 7DQ, UK

[d] Goethe-University, Institute for Pharmaceutical Chemistry, Max-von Laue Str. 9, 60438, Frankfurt am Main, Germany

Protein Kinases (PK) are involved in a number of distinct signaling pathways and are crucial in several aspects of cellular recognition. Among the 518 PKs identified in the human kinome, B-Raf, a serine/threonine PK, is an important constituent of the Ras/Raf/MEK/ERK signal transduction pathway (mitogen-activated protein kinase or MAPK signaling cascade) and participates in cell proliferation and survival. The V600E missense mutation in B-Raf kinase leads to an anomalous regulation of the MAPK pathway which is linked to augmented cell proliferation and tumorigenesis. ATP-competitive B-Raf inhibitors, including vemurafenib, have demonstrated both blocking of the MAPK signaling pathway and tumour response in cells expressing B-RafV600E. However, these compounds paradoxically activate the MAPK pathway in cells expressing wild-type B-Raf. This paradox can promote cellular proliferation and induce cutaneous squamous cell carcinomas and keratoacanthomas. Hence, these ATP-competitive inhibitors, called paradox inducers, can either inhibit or paradoxically activate the MAPK pathway depending on the B-Raf kinases, wild-type or V600E mutant, predominantly expressed in the cell lines.

Next-generation B-Raf inhibitors, named paradox breakers, have been recently developped to inhibit B-Raf^{V600E} activity without paradoxically affecting MAPK pathway in wild-type B-Raf cells. Paradox breakers also overcome several known mechanisms of resistance to first-generation B-Raf inhibitors.

We analyzed four B-Raf kinase inhibitors, 2 paradoxical inducers and 2 breakers, to unravel the biological mechanism at a structural, conformational and cellular level[1,2]. From these studies, we identified key structural components potentially involved on the paradoxical effect. In this work, we focus on the design of a newly paradoxical breaker inhibitor using structural information of both Type I and Type II kinase inhibitors. Crystal structure validates predicted binding mode and cellular data confirm paradoxical breaker behavior of the newly designed compound[3].

References

[1] Arora R, Di Michele M, Stes E, Vandermarliere E, Martens L, Gevaert K, Van Heerde E, Linders JT, Brehmer D, Jacoby E, Bonnet P, Structural investigation of B-Raf paradox breaker and inducer inhibitors. J. Med. Chem., 2015, 1818-1831.

[2] Di Michele M, Stes E, Vandermarliere E, Arora R, Astorga-Wells J, Vandenbussche J, van Heerde E, Zubarev R, Bonnet P, Linders JT, Jacoby E, Limited Proteolysis Combined with Stable Isotope Labeling Reveals Conformational Changes in Protein (Pseudo) kinases upon Binding Small Molecules. J. Proteome Res., 2015, 4179-4193.

[3] Arora R, Linders JT, Aci-Sèche S, Verheyen T, Van Heerde E, Brehmer D, Chaikuad A, Knapp S, Bonnet P, Design, synthesis and characterisation of a novel type II B-RAF paradox breaker inhibitor, Eur. J. Med. Chem., 2023, 115231

DEVELOPMENT OF IN SILICO METHODS FOR STRUCTURE-BASED RNA DRUG DESIGN

L. ZIANI[a]
S. PASQUALI[a]
D. FLATTERS[a]
A.-C. CAMPROUX[a]
[a] Universite Paris Cité, BFA, UMR 8251, CNRS, ERL U1133, Inserm, F-75013 Paris, France

Funding: ANR MERLIN "Multiscale Exploration of RNA polymorphism for drug design »

The importance of studying RNA molecules has been highlighted by the recent pandemic, with the SARS-CoV-2 featuring an RNA-based genome and a replication mechanism controlled by non-coding (nc) RNA[1]. Unlike RNA genome, ncRNAs are less prone to mutation due to their fundamental function, making them interesting therapeutic targets2. However, targeting them with small molecules (SM) remains very challenging because of their flexible and dynamic nature, where multiple structures can be adopted by the same sequence, a property known as structural polymorphism.

The main objective of this study is to explore and optimize emerging in silico drug design tools for RNA to take into account this structural polymorphism. A first step involves the characterization of RNA-SM interactions which requires the detection of possible RNA binding sites. We're currently exploring the increasing experimental RNA structural databases available and focusing on HARIBOSS[3] as the most complete and accessible database, containing 746 RNA-SM complexes. We'll use these complexes for testing and comparing several machine learning prediction methods for binding site identification specifically developed for RNAs such as BiteNetN[4] and RLBIND[5] (convolutional neural network) or RNASite[6] (random forest). Different docking methods developed for RNAs (RLDOCK[7], MORDOR[8]) or for both proteins and RNAs (AutoDock[9], DOCK6[10]) will also be tested and compared on these data. Following the

exploration of the performance of the existing methods (both for binding site extraction and docking approaches) and their limits, we will select optimal methods and combinations adapted to RNA and its structural polymorphism.

References

[1] AJiao AL at al, RNA-mediated gene activation, Epigenetics 9, 2013, 27-36, doi.org/10.4161/epi.26942

[2] Hargrove A, Small molecule–RNA targeting: starting with the fundamentals, Chemical communications, 2020, 94: 14744-56, doi.org/10.1039/D0CC06796B

[3] Panei F P, Torchet R, Ménager H, Gkeka P, Bonomi M, HARIBOSS: a curated database of RNA-small molecules structures to aid rational drug design, Bioinformatics, Volume 38, Issue 17, 2022, 4185–4193, doi.org/10.1093/bioinformatics/btac483

[4] Kozlovskii I, Popov P, Structure-based deep learning for binding site detection in nucleic acid macromolecules, NAR Genomics and Bioinformatics, Volume 3, Issue 4, 2021, lqab111, doi.org/10.1093/nargab/lqab111

[5] Wang K, Zhou R, Wu Y, Li M, RLBind: a deep learning method to predict RNA–ligand binding sites, Briefings in Bioinformatics, Volume 24, Issue 1, 2023, doi.org/10.1093/bib/bbac486

[6] Su H, Peng Z, Tang J, Recognition of small molecule-RNA binding sites using RNA sequence and structure, Bioinformatics 37(1), 1-7, 2021, doi.org/10.1093/bioinformatics/btaa1092

[7] Sun L Z, Jiang Y, Zhou Y, Chen S J, RLDOCK: A New Method for Predicting RNA–Ligand Interactions, J. Chem. Theory Comput., 16, 11, 7173-7183, 2020, doi.org/10.1021/acs.jctc.0c00798

[8] Guilbert, C, James, T L. Docking to RNA via root-mean-square-deviation-driven energy minimization with flexible ligands and flexible targets, J. Chem. Inf. Model, 2008, 48, 1257–1268, doi: 10.1021/ci8000327

[9] Morris G M, Goodsell D S, et al., Automated Docking Using a Lamarckian Genetic Algorithm and and Empirical Binding Free Energy Function, J. Computational Chemistry, 1998, 19: 1639-1662, doi.org/10.1002/(SICI)1096-987X(19981115)19:14<1639::AID-JCC10>3.0.CO;2-B

[10] Allen W J, Balius T E, Mukherjee S, et al., DOCK 6: Impact of new features and current docking performance, Journal of Computational Chemistry, 2015, Volume 36, Issue 15, 1132-1156, doi.org/10.1002/jcc.23905

MODELING STRUCTURES OF INTRINSICALLY DISORDERED PROTEIN-PROTEIN INTERACTIONS: THE CASE STUDY OF THE CHCHD4-AIF COMPLEX

TÂP HA-DUONG [a]
LIUBA MAZZANTI [a]
BRUNO FIGADERE [a]
CATHERINE BRENNER [b]

[a] BioCIS, CNRS UMR 8076, Université Paris-Saclay, 91400 Orsay, France
[b] Université Paris-Saclay, CNRS, Institut Gustave Roussy, Aspects métaboliques et systémiques de l'oncogénèse pour de nouvelles approches thérapeutiques, 94805 Villejuif, France

Intrinsically disordered proteins (IDPs) are characterized by one or several regions which lack stable secondary and tertiary structure in their unbound state under physiological conditions. They frequently play crucial roles in the regulation of many biological processes and, to exert their functions, interact with several biomolecular partners. Due to their very high flexibility, the determination of their bound and unbound structures is a difficult task, to which molecular modeling can usefully contribute and provide valuable information.

In the presented study, we were interested in the case of the mitochondrial CHCHD4-AIF complex which controls the import and the proper folding of defined cysteine containing proteins in the mitochondrial intermembrane space[1,2] and which was recently demonstrated to play a key role in tumorigenesis in lung cancer[3]. Unfortunately, the three-dimensional structure of the CHCHD4-AIF complex is still unknown, impeding the structure-based rational design of inhibiting ligands. The tertiary structure of the full-length human AIF protein was resolved by X-ray crystallography (PDB ID : 1M6I)[4]. An experimental 3D structure of human CHCHD4 is also available in the PDB, but only its folded CX9CCX9C domain (residues 45-109) could be resolved (PDB ID : 2K3J)[5]. However, Hangen et al. demonstrated that this CHCHD4 domain does not participate in the interaction with AIF but that it is rather its intrinsically disordered N-terminal segment of 27 residues which is involved in binding AIF[1].

To predict the structure of IDP-protein complex, we have set up an efficient protocol in three steps[6]: (i) Enhanced molecular dynamics (MD) simulations of the disordered region to sample its conformational ensemble. (ii) Identification of the transient conformations with secondary structures and docking of these conformations into the protein partner. (iii) MD simulations of the most promising complexes to refine their structures and quantify their binding free energies. We report here the preliminary results of this approach to the case of the CHCHD4-AIF complex.

References

[1] Hangen et al. Interaction between AIF and CHCHD4 Regulates Respiratory Chain Biogenesis. Mol. Cell 2015, 58, 1001.
[2] Reinhardt et al. AIF meets the CHCHD4/Mia40-dependent mitochondrial import pathway. Biochim. Biophys. Acta - Molecular Basis of Disease 2020, 1866, 165746.
[3] Rao et al. AIF-regulated oxidative phosphorylation supports lung cancer development. Cell Res. 2019, 29, 579.
[4] Ye et al. DNA binding is required for the apoptogenic action of apoptosis inducing factor. Nat. Struct. Mol. Biol. 2002, 9, 680.
[5] Banci et al. MIA40 is an oxidoreductase that catalyzes oxidative protein folding in mitochondria. Nat. Struct. Mol. Biol. 2009, 16, 198.
[6] Chan-Yao-Chong et al. Investigation into Early Steps of Actin Recognition by the Intrinsically Disordered NWASP Domain V. Int. J. Mol. Sci. 2019, 20, 4493.

MODELLING BOUND AND UNBOUND RNA STRUCTURES USING MOLECULAR DYNAMICS SIMULATIONS TO UNRAVEL THE RELATIONSHIP BETWEEN STRUCTURE, FLEXIBILITY AND CHEMICAL REACTIVITY

CÉCILIA HOGNON[a]
NOLWENN VIGNERON[a]
AFRA SABEI[a]
ELISA FREZZA[a]
[a] Cibles Thérapeutiques et Conception de Médicaments – Université Paris Cité : UMR8038, CNRS – France

Ribonucleic acid (RNA) molecules are involved in most steps of the genetic expression including catalysis of central cellular functions. RNA functions crucially rely on both the specific tridimensional (3D) folding of the molecule, which in turn depends on the sequence and on how nucleobases pair through hydrogen bonds (secondary structure)[1], and its conformation. This relationship is even more crucial for protein-RNA complexes. Hence, determination of RNA tridimensional structures is fundamental for understanding their function. However, obtaining high-resolution 3D structures via X-ray crystallography and NMR is still a challenge.[2] To overcome the lack of 3D structures, in the last decades, several low-resolution techniques have been developed, like chemical probing, whose data have been integrated in the prediction of secondary (2D) and 3D RNA structures with different levels of detail.[3-5] In particular, the SHAPE (Selective 2 Hydroxyl Acylation analysed by Primer Extension) technology provides quantitative reactivity information for each nucleotide and has become the most popular among these techniques since it does not depend on the nature of the nucleotide unlike other chemical probing techniques and is amenable to high-throughput protocols. The probes are small-molecule electrophiles like 1- methyl-7-nitroisatoic anhydride (1M7) that acylate the 2'-hydroxyl group to form a 2'-O-adduct.[6] Although this approach is very popular and it is known that the SHAPE reaction is dependent on the local structural properties of each nucleotide, it has not yet been understood why different reactivities can be obtained for the same nucleotide depending on the probe used and several questions associated with the relationship between structure, conformation, flexibility, and reactivity are still open. To overcome this with the aim to use SHAPE data to predict bound and unbound RNA structures, in the last years, we have performed all-atom molecular dynamics simulations on a set of RNA molecules (unbound and bound to a protein or a ligand) for which their tridimensional structures are available and SHAPE data are accessible in the literature or have been obtained in our wet lab. We analyzed the correlations between different geometrical parameters and the chemical reactivity. For the bound RNAs we also characterized the interfaces and their formation in relationship with the change of SHAPE reactivity. Our investigations confirm that SHAPE reactivity is guided by the local flexibility of the different chemical moieties and the ribose plays a crucial role and suggest that a multiscale approach based on different length scales seems to be necessary to understand and integrate chemical probe data in relation with RNA flexibility and structure.

References

[1] D. M. Lilley, Catalytic mechanisms of nucleolytic ribozymes, Biochemical Society Transactions, Biblio., 2017, 45, 683–691.
[2] B. Coimbatore Narayanan et al, Nucleic Acid Database, Nucleic Acids Research, Biblio., 2013, 42, D114– D122.
[3] K. M. Weeks, RNA structure analysis by chemical probing, Current Opinion in Structural Biology, Biblio., 2010, 20, 295–304.
[4] C. E. Hajdin, et al, SHAPE-directed RNA secondary structure modeling, Proceedings of the National Academy of Sciences, Biblio., 2013, 110, 5498–5503.
[5] D. Mitchell, et al, Probing RNA structure in vivo, Current Opinion in Structural Biology, Biblio., 2019, 59, 151–158.
[6] J. L. McGinnis, et al, Ribosome RNA assembly intermediates, Biochemistry, Biblio., 2014, 53, 3237–3247.

INTEGRATING COMPUTATIONAL APPROACHES TO GUIDE THE IN VITRO ENGINEERING OF BVMO ENZYMES.

ALEXANDRE G. DE BREVERN [a]
JOSEPH REBEHMED [b]

[a] INSERM UMR 1134, DSIMB, Université Paris Cité, France
[b] Department of Computer Science and Mathematics, Lebanese American University, Lebanon

Baeyer-Villiger monooxygenases (BVMOs) are enzymes that convert linear and cyclic ketones into their corresponding esters and lactones, which are essential products for the food, cosmetic and pharmaceutical industries. This biochemical reaction offers advantages over its chemical counterpart, such as improved regio- and enantio-selectivity, and is consistent with the principles of sustainable and green chemistry. Type I BVMOs are composed of a single polypeptide chain and require the co-factors FAD and NADPH to carry out their catalytic function.

The first goal of this study was to identify new type I BVMOs in the databases in order to broaden the range of enzymes available for biocatalysis, and to examine the amino acid conservation patterns in the evolutionary history of this enzyme family. We found that amino acids in the vicinity of the FAD and NADPH cofactors exhibited high levels of conservation, and the details of the interactions were highlighted. Interestingly, residues located at the enzyme binding site displayed lower levels of conservation, and this could potentially explain the selectivity and specificity of the various members of this family for different ligands.

In the second part of this study, the focus was on improving the catalytic activity of BVMO4 from Dietzia sp. D5. Although this enzyme demonstrated good thermostability, it was not effective at catalyzing the conversion of cyclohexanone to caprolactone. To address this issue, threading was employed to generate a structural model of the enzyme, which was then combined with above phylogenetic results and literature information to identify the hotspots whose mutation would likely affect substrate specificity. Site saturation mutagenesis was performed on 12 selected sites, resulting in the screening of 528 mutants. Approximately one-fourth of the screened mutants exhibited more than a 50% increase in cyclohexanone oxidation activity. The top performing mutants, namely Y499I, Y499F, and Y499L, showed about a 12-fold increase in caprolactone production compared to the wild type BVMO4, while retaining its thermostability.

References

J. Rebehmed; Evolution study of the Baeyer-Villiger monooxygenases enzyme family: Functional importance of the highly conserved residues; Biochimie 2013, 95(7):1394-402.

S. Bisangi; Enhancing the Activity of a Dietzia sp. D5 Baeyer‐Villiger Monooxygenase towards Cyclohexanone by Saturation Mutagenesis; ChemistrySelect 2017, 2(24):7169-7177.

COMPARISON OF UBIQUITIN CHAINS USING MOLECULAR DYNAMICS

KONSTANTIN OEXLE [a]
KEVIN SAWADE [a]
CHRISTINE PETER [a]
[a] Department of Chemistry, University of Konstanz, Konstanz, Germany

Understanding Ubiquitin signaling is crucial because of its role in modulating protein functions. Ubiquitin chains are multidomain proteins with different signal behavior depending on their structural characteristics size and linkage site.[1] Understanding the conformational space of polyubiquitin chains via computational studies clarifies the structural characteristics. System size and metastable states impede atomistic simulations. Therefore, coarse graining is applied to simplify the system. The developed back-mapping based sampling (BMBS) produces an atomistic ensemble that can be projected into the two-dimensional coarse grained landscape. This sampling strategy has recently proven to work for more complex systems. Together with the sampling a conformational clustering has been introduced for further information about the structural composition of the 2D landscape.[2] In this work, the canonical chains of di- and tetra Ubiquitin (K48, K63) are compared by using a suitable set of Collective Variables (CV). Furthermore, conformational clustering and BMBS will be adapted for these larger systems.

References

[1] Masato Akutsu, Ivan Dikic, Anja Bremm; Ubiquitin chain diversity at a glance. J Cell Sci, 2016; 129 (5): 875–880

[2] Hunkler et al, Generating a conformational landscape of ubiquitin chains at atomistic resolution by back-mapping based sampling , Frontiers in Chemistry; 10 2023

TOWARDS THE MECHANISTIC UNDERSTANDING OF A PROTEIN MUSCLE KINASE WITH MULTISCALE MOLECULAR DYNAMICS SIMULATIONS

MADLEN MALCHAREK [a]
CHRISTOPH GLOBISCH[a]
CHRISTINE PETER[a]
[a] Department of Chemistry, University of Konstanz, Universitätsstraße 10, 78464 Konstanz, Germany

Protein kinases of the titin-like family located in muscle fibers are postulated to act as mechanosensors. They are suggested to perceive the muscle sarcomere's mechanical stretch and trigger signal pathways in response. In general, external triggers can affect protein structure or protein domain interactions, thus affecting the accessibility or shape of interaction sites. However, the exact mechanism of the mechanosensor activity of these protein kinases is not fully understood and appears to vary within the family. In this project we are working towards the understanding of the molecular mechanism of a muscle protein kinase of the titin-like family. The molecular structure of the kinase makes the intramolecular domain-domain interaction particularly worthy of investigation.

Molecular dynamics simulations can provide valuable atomistic level information on such a system. Since the system of interest is rather large and the relevant timescales long, we use multiscale approaches as they allow precise analysis on suitable time scales. We performed initial classical atomistic and coarse-grained simulations, from which we developed a customized analysis approach using among other methods machine learning techniques. Particular interest has so far been focused on the study of domain-domain interactions of the protein, hence residuewise minimum distances were calculated and evaluated using dimensionality reduction and clustering algorithms. This gave us first insights into the interaction landscape of the domains of the kinase and provides a good starting point for extended sampling of the system under different conditions.

INVESTIGATE THE MECHANISMS OF ONSET FOR ALSIN-RELATED PATHOLOGY BY MEANS OF MOLECULAR MODELLING

MARCELLO MICELI[a]
MARCO CANNARIATO[a]
MARCO AGOSTINO DERIU[a]
[a] Polito BIO Med Lab, Department of Mechanical and Aerospace Engineering, Politecnico di Torino, Turin, Italy

Alsin is a multidomain protein of 1657 amino acids, the mutation of which has been linked as responsible for the onset of the Infantile onset Ascending Hereditary spastic paralysis (IAHSP) which is a rare disease characterized by motor neuron degeneration. Alsin consists of four structured domains, namely the N-terminal RLD domain, the DH/PD, the MORN and the C-terminal VPS9 domain[1]. Alsin domains are responsible for a set of functions such as the process of self-tetramerization, selective recognition of Phosphatidylinositol phosphate (PIP) for the localization on membrane ruffle and the triggering of guanine nucleotide exchange in RAC and RAB GTPases, while the whole protein is found to be involved in the vesicular traffic[2]. Currently the atomistic structure of Alsin is unknown, indeed no experimental atomistic structure is available nor for the entire protein neither for some fragment. The absence of information at the molecular level is a limit for the comprehension of molecular mechanisms linked to physiological pathways and the understanding of how aberrant expressions of the Alsin protein are linked to the onset of the pathology. This work aims at developing an atomistic model for Alsin domains by classical homology modelling, I-Tasser methodology and Alphafoldv2. For those models, the Alsin structure-function relationships have been investigated by means of molecular dynamics simulations. In greater detail, this work focused on: i) exploring the capability of the RLD domain to interact with membranes, proposing a putative binding pocket and a mechanism of selective recognition of PIP[3]; ii) investigating the conformational dynamics of the DH/PH domain, as a consequence of its interaction with RAC GTPases, highlighting the stabilizing effect of this interaction on the DH/PH conformational dynamics[4]; iii) propose a possible mechanisms of self-interaction driven by the MORN domain; iv) integrating experimental and computational studies on VPS9 domain to investigate how single point mutations may be related to the onset of IAHSP[5]. A domain-focused computational investigation of Alsin will pave the way to develop a model for the whole Alsin protein, linking together the information coming from the single domain study and then providing information to build and investigate the homo-tetrameric form. This computational model will help to understand, at a molecular scale, how Alsin domains are involved and act in the process of vesicular trafficking and how the aberrant expression could interfere with the protein function, resulting in the onset of the IAHSP.

References

[1] Sato K et al Altered oligomeric states in pathogenic ALS2 variants associated with juvenile motor neuron diseases cause loss of ALS2-mediated endosomal function S. J Biol Chem.,2018, 293(44),17135-17153

[2] Miceli M, et al, ALS2-Related Motor Neuron Diseases: From Symptoms to Molecules Biology, 2022 11(1):77

[3] Cannariato et al, In silico investigation of Alsin RLD conformational dynamics and phosphoinositides binding mechanism, PLOS ONE, 2022, 17(7): e0270955.

[4] Cannariato M. Miceli M. et al, Prediction of Protein–Protein Interactions Between Alsin DH/PH and Rac1 and Resulting Protein Dynamics, Frontiers in Molecular Neuroscience, 2022, (14) .

[5] Miceli M. et al, Elucidating molecular connection between IAHSP onset and Alsin protein by means of homology modelling and molecular dynamics, Biomedical Science and Engineering, 2021, (5):s1

UNDERSTANDING THE SIGNAL TRANSDUCTION MEDIATED BY MINCLE RECEPTOR

MAXIME NORIEGA[a], PASCAL DEMANGE[a], ADRIEN SCHAHL[b], GUILLAUME FERRÉ[a], OLIVIER SAUREL[a], MATTHIEU CHAVENT[b], ANDREW ATKINSON[a], GEORGES CZAPLICKI[a]

[a] Integrative Biological NMR group, Institut de Pharmacologie et de Biologie Structurale (IPBS), Toulouse

[b] Multiscale Computational Immunology group, Institut de Pharmacologie et de Biologie Structurale (IPBS), Toulouse

Mincle is a c-type lectin receptor that act as sensor of pathogen-associated molecular patterns. Mincle play key roles in the innate immune system. Mincle is a transmembrane receptor that consist of a short cytosolic N-terminal sequence and an extracellular domain including a C-terminal carbohydrate recognition domain (CRD). The extracellular and cytosolic domains are connected by a single-pass transmembrane domain. The extracellular binding of trehalose dimycolate, a glycolipid from mycobacteria, to human Mincle leads to intracellular activation of NF-κB via the Syk-Card9-Bcl10-Malt1 pathway[1,2]. Signal transduction requires FcRγ, a single-pass transmembrane protein bearing cytosolic ITAM motifs that are phosphorylated by Src family kinases in the first intracellular step of this signalling pathway[3].

The aim of the project is to establish a molecular description of the signal transduction mediated by the Mincle receptor in the context of infection by *Mycobacterium tuberculosis*. To achieve this, different constructs will be used. Firstly, the CRD domain alone will be expressed and isotopically labeled with ^{15}N to perform NMR experiments aimed at understanding its dynamics as well as its interaction with glycolipids. Next, the complete protein in the presence or absence of its partner FcRγ will be expressed in a membrane context to understand how these two proteins associate and to study the structural rearrangements that may occur following recognition of the mycobacterial glycolipid. To do this, the use of nanodiscs to better mimic the membrane environment will be employed.

In a first step, we have employed expression of the extracellular domain of human Mincle in Origami strain, enabling the production of isotopically labelled protein for structural studies by NMR. We are carrying out ligand interaction studies and characterisation of the dynamic properties of Mincle through ^{15}N relaxation measurements. In parallel, a theoretical approach employing a multi-scale molecular dynamics method is used. To capture various conformations of the protein, an adaptive sampling protocol has been employed in combination with the Amoeba polarizable force field. Additionally, coarse-grained models are used, exploiting recent refinements to the Martini3 coarse-grained force field that enable it to adequately study protein-ligand interactions. This allows us to simulate the system for longer duration, with the hope of capturing protein/ligand interaction events.

The philosophy behind this project is to implement an integrative approach involving these multiscale simulations and a range of biophysical techniques in order to understand the molecular mechanisms underlying this signalling event and to contribute to the development of new therapies against tuberculosis.

References

[1] Matsumoto, M., Tanaka, T., Kaisho, T., Sanjo, H., Copeland, N. G., Gilbert, D. J., ... & Akira, S. (1999). A novel LPS-inducible C-type lectin is a transcriptional target of NF-IL6 in macrophages. The Journal of immunology, 163(9), 5039-5048.

[2] Matsunaga, I., & Moody, D. B. (2009). Mincle is a long sought receptor for mycobacterial cord factor. Journal of Experimental Medicine, 206(13), 2865- 2868.

[3] Yamasaki, S., Ishikawa, E., Sakuma, M., Hara, H., Ogata, K., & Saito, T. (2008). Mincle is an ITAM-coupled activating receptor that senses damaged cells. Nature immunology, 9(10), 1179-1188.

BOTTOM-UP COARSE-GRAINED MODEL OF TiO_2 – PHOSHOLIPID INTERACTIONS

MIKHAIL IVANOV
ALEXANDER LYUBARTSEV
Department of Materials and Environmental Chemistry, Stockholm University, Stockholm

Titanium dioxide (TiO_2) nanomaterials are ubiquitous in modern applications like solar cells, sunscreens, paints and self-cleaning coatings, yet their possible adverse environmental and health effects are not fully understood[1]. It has been shown that in biological systems, the exposure of TiO_2 nanoparticles may lead to the formation of lipid-nanomaterial aggregates[2], which means that simulating the lipid-nanomaterial interfaces could help predict the behavior of nanomaterials in biological environments. One possibility is to simulate the system of interest at atomistic resolution using classical molecular dynamics. However, a typical size of TiO_2 nanoparticles exceeds 10 nm, so a coarse-grained model is a more viable option.

In this work, we aim to develop a coarse-grained model of TiO_2 nanomaterials in contact with common cell membrane phospholipids using the Inverse Monte Carlo method[3], implemented in the MagiC software package[4]. The method finds a set of effective potentials (in tabular form) for all of the coarse-grained interaction sites in the system that reproduces the structure of the system obtained at atomistic resolution. The obtained TiO_2-lipid effective potentials are then exported to LAMMPS[5] to run coarse-grained molecular dynamics simulations.

We have optimized seven different TiO_2-lipid coarse-grained models, including four different low-energy TiO_2 surfaces – anatase (101), anatase (100), rutile (110) and rutile (101) in contact with DMPC (1,2-dimyristoyl-sn-glycero-3-phosphocholine) or POPE (1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine) phospholipids in the presence of Na+ and Cl- ions at 0.15 M concentration in implicit water. Our coarsegrained models show a good agreement with the atomistic reference data when comparing structural properties, such as the radial distribution functions and number

density profiles of lipid headgroup atoms, while achieving at least an order of magnitude speed-up compared to atomistic simulations. Furthermore, we test our models by running large-scale simulations of anatase spherical nanoparticles of variable size (r = 2, 5, 10 nm) in contact with POPE lipids. Our simulations show that the lipid adsorption increases with the nanoparticle size.

References

[1] Jamie R Lead, et al., Environ. Toxicol. Chem., 2018, 37(8):2029–2063
[2] Kokot H. et al., Adv. Mater., 2020, 2003913
[3] A.P. Lyubartsev, A. Laaksonen, Phys Rev. 1995, E. 52, 3730–3737
[4] A. Mirzoev, A.P.Lyubartsev, J. Chem. Theory and Computations, 2013, 9(3), 1512–1520
[5] Plimpton, S, J. Comput. Phys., 1995, 117:1–19

TUBULIN SIMULATIONS ON THE COARSE-GRAINED LEVEL

POLINA TETERINA[a]
CHRISTOPH GLOBISCH[a]
CHRISTINE PETER[a]
[a] Department of Chemistry, University of Konstanz, Konstanz, Germany

Microtubules, the cytoskeleton filaments of cells are widely used as cancer therapeutics targets due to their vital role in the cell division process. They are comprised of α- and β-tubulin, with the latter containing the binding site for taxol, a microtubule-stabilizing drug from the taxane diterpenes family. In order to perform their functions microtubules are able to rapidly switch between growth and shrinkage in the process that is referred to as "dynamic instability". The mechanisms underlying the dynamic instability of microtubules are still a subject of intense research interest as well as the exact mechanism of taxol's action. Recent studies that use cryo-EM and computational modeling approaches provide new perspectives on the intricacies of microtubule dynamics. In this work, we aim to complement the emerging views by simulating microtubule systems on the coarse-grained level. We will establish a coarse-grained model of tubulin subunits with the ligands using the Martini3 force field that provides a necessary degree of chemical specificity. Conformational changes will be characterized by a combination of various dimensionality reduction and clustering methods. We anticipate contributing not only to a better understanding of microtubule systems but to uncovering the potential of coarse-grained computational modeling as well.

COMPLEXITY OF CYTOCHROME C-CAGED NANO-STRUCTURED HYDROTROPES: FUNCTIONAL DYNAMICS AND BOOSTED STABILITY IN HARSH ENVIRONMENTS

DHEERAJ KUMAR SARKAR[a,b], PRANAV BHARADWAJ[c], MEENA BISHT[d], SACHIN M. SHET[c], S.K. NATARAJ[c], VEERESH LOKESH[d] GREGORY FRANKLIN[d], JAN BREZOVSKY[a,b], DIBYENDU MONDAL[c,d]

[a] Laboratory of Biomolecular Interactions and Transport, Institute of Molecular Biology and Biotechnology, Faculty of Biology, Adam Mickiewicz University, Poznan, Poland

[b] International Institute of Molecular and Cell Biology in Warsaw, Poland

[c] Centre for Nano and Material Sciences, Jain University, Bangalore, India

[d] Institute of Plant Genetics, Polish Academy of Sciences, Poznan, Poland

Background: Amphiphilic (self-assembled) nanostructured solvents have a wide range of applications in biocatalysis and energy-efficient bioprocesses. Due to the complex physicochemical properties and interactions, these solvents have gained much interest in recent years in the area of protein packaging.[1]

Objectives: In the current study, we have investigated the molecular roles of hydrotropes nanostructures caging a Cytochrome c (Cyt c) enzyme in maintaining its stability and activity. We have systematically evaluated the utility of adenosine-5'-triphosphate (ATP) and choline salicylate ionic liquid (IL) based mixed hydrotropes towards improved packing and thermal stability of this enzyme.

Methods: We applied high-throughput molecular dynamics (HTMD) [2] and Bayesian Markov state models [3] to generate the long-live metastable states, which were inferred to investigate the molecular insights governing the Cyt C structure and dynamics at room (300 K) and elevated temperature (363.15 K) under the influence of ATP and IL at a molecular concentration of 5 mM and 300 mg/mL, respectively.

Results: We investigated four molecular systems with Cyt c in the explicit water, ATP, IL and ATP+IL respectively in both 300 K and 363.15 K. At 300 K, we could observe ATP and IL could individually and synergistically enhance the dynamics of Ω 40-54 functional loop region facilitating reversible openings. Metastable states resolved indicated a much faster opening and slower closing of this functional loop. At 363.15 K, no significant effects on Ω 40-54 functional loop region were observed, however, the thermal stability of Ω 70-85 region was significantly enhanced. We also observed higher conformational instability of Cyt c in water than in the presence of hydrotropes. ATP with IL was observed to stabilize the closed state of Cyt c more and could provide thermal stability by frequently interacting with Lys and Arg residues.

Conclusions: In the current study, we explored the utility of ATP and IL-based nano-structured hydrotropes while caging the protein retaining high-temperature biocatalysis. Reversible binding of the ATP and IL with the loop Ω 40-54 indicated higher activity of Cyt c at 300 K while showing higher thermal tolerance for the region Ω 70-85 at 363.15 K. Overall, both ATP and IL-based hydrotropes could specifically have polar interactions with Lys and Arg in maintaining the activity and thermal stability of Cyt c and resulting higher peroxidase activity.

This research was supported by POWER project POWR.03.02.00-00-I006/17. The computations were performed at the Poznan Supercomputing and Networking Center.

References

[1] Bui-Le L, Clarke CJ, Bröhl A, Brogan APS, Arpino JAJ, Polizzi KM, Hallett JP. Revealing the complexity of ionic liquid-protein interactions through a multi-technique investigation. Commun Chem. 2020 May 6;3(1):55.

[2] Doerr S, Harvey MJ, Noé F, De Fabritiis G. HTMD: High-Throughput Molecular Dynamics for Molecular Discovery. J Chem Theory Comput. 2016 Apr 12;12(4):1845-52.

[3] Scherer MK, Trendelkamp-Schroer B, Paul F, Pérez-Hernández G, Hoffmann M, Plattner N, Wehmeyer C, Prinz JH, Noé F. PyEMMA 2: A Software Package for Estimation, Validation, and Analysis of Markov Models. J Chem Theory Comput. 2015 Nov 10;11(11):5525-42.

INFLUENCE OF THE POLY PROLINE REGION INSERTIONS AND DUPLICATIONS ON THE DOMAINS CONTACTS OF THE HEPATITIS E VIRUS ORF1 POLYPROTEIN.

NICOLAS JEANNE[a,b,d], SÉBASTIEN LHOMME[a,b]
MARIE BRUT[c], JACQUES IZOPET[a,b]

[a] Virology laboratory, Federative Institute of Biology, CHU Purpan, Toulouse, France
[b] INSERM Infinity, viral infection team: persistence, host response and patho-physiology, CHU Purpan, Toulouse, France
[c] LAAS-CNRS, Toulouse University, CNRS, UPS, Toulouse, France
[d] LAAS-CNRS, Toulouse University, CNRS, Toulouse, France

Hepatitis E Virus (HEV) is the leading cause of worldwide viral hepatitis. The HEV genome is a single-stranded, positive-sense RNA of about 7.2 kb, organized into three Open Reading Frames (ORF). ORF1 codes for the non-structural polyprotein (~1700 residues), which contains the main enzymatic domains required for viral replication, Methyltransferase Y domain (MetY), Macrodomain, Helicase, Polymerase, as well as other regions whose functions are not well understood, such as the Poly Proline Region (PPR) and the Metal Binding Domain.

Research by my team has shown that the PPR is a site for host sequence insertions or HEV sequence duplications. The insertions provide a replicative advantage to the virus that carries them [1,2]. These inserted or duplicated sequences are not functional and can be fragments of genes or introns. The amino acid composition of the PPR, including insertions or duplications, indicates a rise of the net charge (growth of the number of positively charged amino acids), as well as an increase in the number of post-translational modification sites such as ubiquitinations, acetylations, and phosphorylations.

The literature is not clear on whether this polyprotein is cleaved, but recent studies seem to indicate that it is not [3,4].One question of this study is to define whether insertions or duplications play a role in the structure flexibility of the polyprotein and modify the interactions between the different domains of ORF1.

Three-dimensional models were produced using Alphafold2 [5] from sets of ORF1 polyprotein sequences with and without host genomic insertions or virus duplications. From these models, molecular dynamics experiments were performed. We have developed specialized tools to inspect the hydrogen bonds during the entire molecular dynamics for each type of ORF1 set, and visualize the domain contacts affected by the presence and absence of insertion or duplication events.

We have focused on the contacts between the PPR, where the genomic events occur, and the rest of ORF1. Only the contacts between residues distant from at least ten residues have been taken into account.

Preliminary results show that insertion and duplication events increase the contacts between the PPR and the Y region of the MetY, the PPR, and the RNA-dependent RNA polymerase. On the other hand, there is a decrease in the contacts between PPR and the Macrodomain.

The polymerase is involved in the replication process, as is the Y region of the MetY, on which experimental mutations have demonstrated a reduction in replication effectiveness [6]. The role of the Macrodomain remains unclear.

References

[1] Lhomme S. et al. Characterization of the Polyproline Region of the Hepatitis E Virus in Immunocompromised Patients. J Virol. 2014; 88(20):12017-12025. doi:10.1128/JVI.01625-14.

[2] Lhomme S. et al. Insertions and Duplications in the Polyproline Region of the Hepatitis E Virus. Front Microbiol. 2020; 11:1-10. doi:10.3389/fmicb.2020.00001.

[3] Goulet A. et al. Structure Prediction and Analysis of Hepatitis E Virus Non-Structural Proteins from the Replication and Transcription Machinery by AlphaFold2. Viruses. 2022; 14(7):1537. doi:10.3390/v14071537.

[4] Fieulaine S.. et al. De novo modelling of HEV replication polyprotein: Five-domain breakdown and involvement of flexibility in functional regulation. Virology. 2023; 578:128-140. doi:10.1016/j.virol.2022.12.002.

[5] Jumper J. et al. Highly accurate protein structure prediction with AlphaFold. Nature 2021; 596(7873):583-589. doi:10.1038/s41586-021-03819-2.

[6] Parvez MK. Mutational analysis of hepatitis E virus ORF1 «Y-domain»:Effects on RNA replication and virion infectivity. World Journal of Gastroenterology 2017; 23(4):590-602. doi:10.3748/wjg.v23.i4.590.

FUNCTIONAL MODES AND THERMAL B-FACTOR PREDICTIONS FOR MULTIGENIC STRUCTURAL ANALYSIS PREDICTED FROM ALPHAFOLD

NICOLAS PETIOT[a]
PATRICK SENET[a]
MATHIEU SCHWARTZ[b]
FABRICE NEIERS[b]
ADRIEN NICOLAÏ[a] adrien.nicolai@u-bourgogne.fr

[a] Laboratoire Interdisciplinaire Carnot de Bourgogne, UMR 6303 CNRS-Université de Bourgogne Franche-Comté, 21078 Dijon, France.

[b] Flavour Perception: Molecular Mechanisms (Flavours), Université de Bourgogne Franche-Comté, INRAE, CNRS, 21000 Dijon, France.

Glutathione Transferase (GST) is a superfamily of enzymes that are generally homodimeric structures and multigenic in numerous organisms. They are involved in detoxification process as well as in chemoperception in mammals and insects [1]. Their main function is to catalyze the conjugation of reduced glutathione (GSH) to xenobiotic electrophilic centers. In their catalytic cycle, the GSH usually binds in a specific set of amino-acids called G-site and the hydrophobic xenobiotic in the so-called H-site [2]. Interactions between insects and plant's chemicals lead to a major driving force in herbivorous insect evolution, hence this encourages the study of insect GSTs to understand how spontaneous mutations modify the stability, selectivity and the catalytic efficiencies of this enzyme superfamily. The goal of the present work is to analyze how the amino-acid sequence modifies the structure and flexibility of an ensemble of variants of GST from class δ and ε of *Drosophilia melanogaster*'s. We modeled 25 GST structures (11 in class δ and 14 in class ε). First, from the amino acid sequences, we used Multiple Sequence Alignment (MSA) to describe the binding sites and the dimerization interface. Second, we computed the 3D atomic structures of each GST using the AlphaFold program developed by DeepMind [3]. It allows, based on the protein sequence and deep learning techniques, to achieve very accurate predictions on the corresponding three-dimensional structures. When available, the AlphaFold structures were compared with experimental ones from the Protein Data Bank (PDB). Third, we used the Anisotropic Network Model (ANM) [4] to compute: i) the thermal factors of all GSTs and ii) the correlation between residue motions governed by the functional (acoustical) modes predicted by ANM. Comparison between the 25 GSTs will be presented and rationalized. This *in silico* methodology will be extended to direct the design of new enzymes that will be synthesized, purified, and resolved experimentally.

References

[1] Mathieu Schwartz, Valentin Boichot, Stéphane Fraichard, Mariam Muradova, Patrick Senet, Adrien Nicolaï, Frederic Lirussi, Mathilde Bas, Francis Canon, Jean-Marie Heydel, et al. Role of insect and mammal glutathione transferases in chemoperception. Biomolecules, 13(2):322, 2023.

[2] Adrien Nicolaï, Nicolas Petiot, Paul Grassein, Patrice Delarue, Fabrice Neiers, and Patrick Senet. Free-energy landscape analysis of protein-ligand binding: the case of human glutathione transferase A1. Applied Sciences, 12(16):8196, 2022.

[3] John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Zidek, Anna Potapenko, et al. Highly accurate protein structure prediction with alphafold. Nature, 596(7873):583–589, 2021.

[4] Ali Rana Atilgan, SR Durell, Robert L Jernigan, Melik C Demirel, O Keskin, and Ivet Bahar. Anisotropy of fluctuation dynamics of proteins with an elastic network model. Biophysical Journal, 80(1):505–515, 2001.

COARSE-GRAINED MODELING USING NEURAL NETWORKS TRAINED ON STRUCTURAL DATA

MAKSIM POSYSOEV [a]
MIKHAIL IVANOV [a]
ALEXANDER LYUBARTSEV [a]
[a] Department of Materials and Environmental Chemistry, Stockholm University, Stockholm, Sweden

Multiscale modeling is important in biomolecular and material science, but requires consideration of a wide range of spatial and time scales. Coarse-graining, which aims to reduce less significant degrees of freedom while retaining important ones, is a commonly used approach in developing multiscale models. Empirical parameterization of interaction potentials is frequently used to create coarse-grained models. However, determining the optimal functional form of effective interaction potentials between coarse-grained sites is a challenge. We propose a data-driven approach that employs machine learning and artificial neural networks (ANNs) to develop coarse-grained force fields based on system structure. The approach combines concepts from the ML/ANN force field and structure-based coarse-graining using the inverse Monte Carlo method[1]. Our research illustrates the approach on the Lennard-Jones fluid and a basic model of methanol-water mixtures. The approach has the potential to generate more transferable coarse-grained models, which addresses a critical problem in multiscale modeling.

This study presents a method for predicting the total energy of an interacting atom using the Behler-Parrinello[2] neural network architecture. The proposed method also integrates concepts from the Inverse Monte Carlo method1, utilizing statisticalmechanical relationships to connect changes in interaction parameters with changes in the radial distribution function. The training procedure aims to replicate the radial distribution function rather than predicting the energy directly, providing a new approach to the inverse problem of determining interatomic potentials from structural data.

In this study, a methanol-water mixture is used as the test system to train a neural network model using a range of concentrations. Subsequently, alternative concentrations were chosen for validation purposes, allowing for comparison with the Inverse Monte Carlo approach. The method exhibits adaptability to varying concentrations and the ability to predict the behavior of systems not included in the training data while preserving structural properties, as indicated by the radial distribution function (RDF).

References

[1] Lyubartsev, A. P.; Laaksonen, A. Calculation of Effective Interaction Potentials From Radial Distribution Functions: A Reverse Monte Carlo Approach. Phys. Rev. E 1995, 52, 3730.

[2] Behler, J.; Parrinello, M. Generalized Neural-Network Representation of High- Dimensional Potential-Energy Surfaces. Phys. Rev. Lett. 2007, 98, 146401.

EXPLORING THE CONFORMATIONAL LANDSCAPE OF THE GHRELIN RECEPTOR (GHSR) WITH BIASED MOLECULAR DYNAMICS SIMULATIONS

RITA ANN ROESSNER, MAXIME LOUET, NICOLAS FLOQUET

Institut des Biomolecules Max Mousseron, UMR5247, CNRS, Université De Montpellier, ENSCM, 1919 Route de Mende, Montpellier, Cedex 05 34095, France

G protein-coupled receptors (GPCRs) are the largest family of membrane receptors and, as such, are responsible for the majority of cellular responses to a variety of different stimuli, including hormones, neurotransmitters and external substances[1]. The key role of GPCRs in understanding molecular determinants of a wide range of human diseases is emphasized by considering that one third of approved pharmaceutical drugs targets them[2]. The Growth Hormone Secretagogue Receptor (GHSR), shortly named Ghrelin receptor belongs to the subfamily of class A GPCRs and controls growth hormone secretion, food intake, and reward-seeking behaviors upon interaction with the neuro-endocrine peptide hormone Ghrelin[3]. The conformational behavior of GHSR upon partners binding is one transversal research axis of our laboratory for many years[4,5]. While it is widely recognized that the activity of GPCRs involves a range of structurally and functionally distinct states in equilibrium, it remains unclear how ligands and/or protein partners can affect these equilibria[6]. The objective of my PhD is to shed light on the modulation of the conformational landscape of GHSR by small molecular ligands (agonists, inverse agonists and antagonists) and ultimately by its preferred protein partners (Gproteins, Arrestin). Taking advantage of the huge amount of structural data in this field, we performed a principal component analysis (PCA) of all class-A GPCRs structures available in the Protein Data Bank (PDB). Based on the assumption that the collective motions among this family are conserved we used the amplitudes of displacement along the first eight eigenvectors as collective variables for well-tempered metadynamics simulations combined with a multiple walkers / replica exchange scheme. We show that 1 μs of accumulative simulation time along each principal component is sufficient to exhaustively sample the collective motions of the receptor in the eight-dimensional space with good reproducibility of the potential of mean forces (PMF). The models resulting from these simulations were in good agreement with experimental data previously obtained in our team and/or described in the literature[7,8]. Our results suggest that biased MD simulations in combination with PCA can be exploited to efficiently sample the collective motions of GHSR. We will use the most representative conformations of the receptor as starting points for standard (unbiased) MD simulations whose *a posteriori* analysis using a Markov state model will provide insights on the dynamical transitions within the conformational landscape of GHSR depending on bound ligand(s)/ protein partner(s).

References

[1] Alberts B, Bray D, Hopkin K, Johnson AD, Lewis J, Raff M, Essential Cell Biology. Garland Science, 2015, 863 p.

[2] Sriram K, Insel PA. G Protein-Coupled Receptors as Targets for Approved Drugs: How Many Targets and How Many Drugs? Mol Pharmacol. 2018, 251–8.

[3] Sivertsen B, Holliday N, Madsen AN, Holst B. Functionally biased signalling properties of 7TM receptors - opportunities for drug development for the ghrelin receptor, Br J Pharmacol., 2013, 1349–62.

[4] Floquet N, M'Kadmi C, Perahia D, Gagne D, Bergé G, Marie J, Activation of the ghrelin receptor is described by a privileged collective motion: a model for constitutive and agonist-induced activation of a sub-class A GPCR, J Mol Biol. 2010, 769–84.

[5] Damian M, Mary S, Maingot M, M'Kadmi C, Gagne D, Leyris JP, Ghrelin receptor conformational dynamics regulate the transition from a preassembled to an active receptor:Gq complex. Proc Natl Acad Sci U S A, 2015, 1601–6.

[6] Casiraghi M, Damian M, Lescop E, Point E, Moncoq K, Morellet N, Functional Modulation of a G Protein-Coupled Receptor Conformational Landscape in a Lipid Bilayer, J Am Chem Soc., 2016, 11170–5.

[7] Louet M, Casiraghi M, Damian M, Costa MG, Renault P, Gomes AA, Concerted conformational dynamics and water movements in the ghrelin G protein-coupled receptor, eLife, 2021, 63201.

[8] Ferré G, Louet M, Saurel O, Delort B, Czaplicki G, M'Kadmi C,. Structure and dynamics of G protein-coupled receptor–bound ghrelin reveal the critical role of the octanoyl chain, Proc Natl Acad Sci, 2019, 17525–30.

TRACE AMINE ASSOCIATED RECEPTORS (TAARS) RESPONSE TO AMINES ARE LARGELY AFFECTED BY SEQUENCE VARIANTS.

JODY PACALON[a]
CHRISTINE BELLOIR[b]
SÉBASTIEN FIORUCCI[a]
LOÏC BRIAND[b]
JÉRÉMIE TOPIN[a]

[a] Université Côte d'Azur (UCA), UMR7272, Faculté des sciences Parc Valrose, 28 avenue Valrose, 06000, Nice (France)
[b] Centre des Sciences du Goût et de l'Alimentation, CNRS, INRAE, Institut Agro, Université Bourgogne Franche-Comté, F-21000 Dijon, France

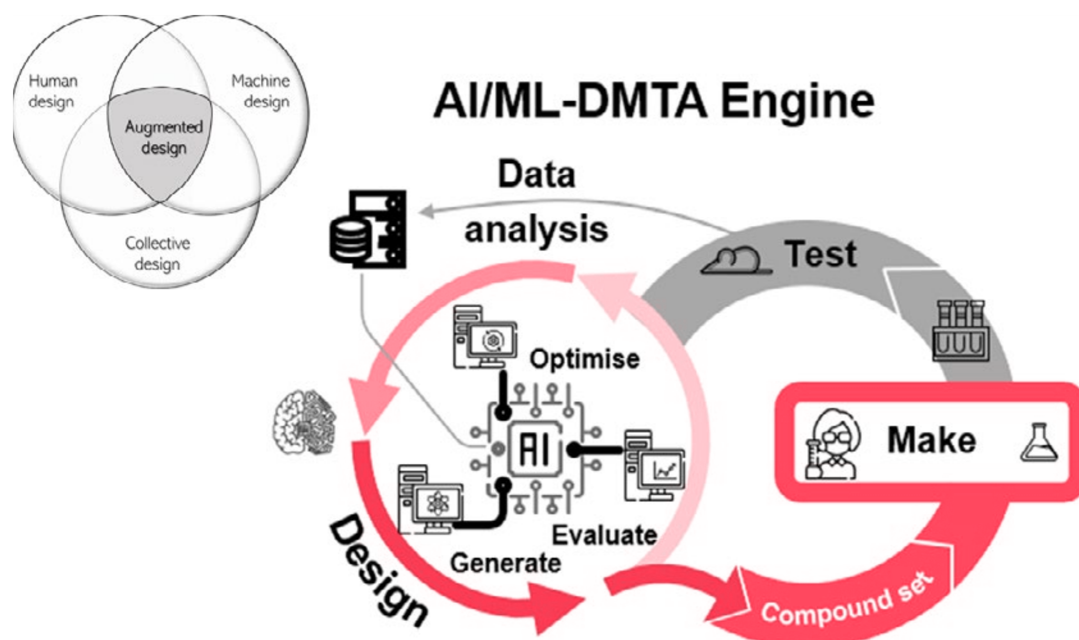
Volatile amines are recognized by a family of chemosensory receptors: the Trace Amine Associated Receptors (TAARs). Compared to regular olfactory receptors, TAARs are few (6 receptors expressed in the olfactory epithelium) and highly conserved. Thus, polymorphisms in this family can drastically alter our perception of amine compounds. A joint approach of numerical simulations and in vitro experiments has revealed the activation mechanisms of hTAAR5. hTAAR5-S95P is a polymorphism found at high frequency in Nordic countries. People with this mutation have their perception of trimethylamine affected, making them less able to perceive the rotten fish smell caused by this molecule. Our 3D model captures both the inability of hTAAR5-S95P to be activated by trimethylamine (TMA) in vitro, and the activation of the receptor by different agonists. Long-scale molecular dynamics simulations of the system bound to ligands with different efficacies are performed and recover that the receptor is activated only when stimulated by agonists, capturing the features of a prototypical active state of GPCR. 2 specific features of the TAAR family were studied.

EVOTEC MOLECULAR ARCHITECTS WAYS OF WORKING

MARTINA BERTAZZO[a], PATRICIA ANGELLOZ-NICOUD [a] FRANÇOIS CARTIER[a], LAURIANNE DAVID[a], CHRISTOPHE MULLER[b] LIONEL COLLIANDRE[b], LAËTITIA BREUIL[a], BRICE SAUTIER[a] MARTIN KOTEV[a], OBDULIA RABAL-GRACIA[a], SYLVIE GOMEZ[a] JÉRÔME MENEYROL[a], CHRISTOPHE BOLDRON[a]

[a] Evotec SE, Molecular Architects, Campus Curie, 195 Route d'Espagne, 31036 Toulouse, France
[b] Evotec SE, in silico R&D, Campus Curie, 195 Route d'Espagne, 31036 Toulouse, France

Drug design is evolving thanks to the significant increase in the amount of data and computing power as well as the development of cutting-edge in silico technologies including AI/ML. To transform these major developments into drug research accelerator, new scientific profiles and ways of working must emerge to encode the large range of information from literature including known ligands, activity data, advanced drugs, clinical data, PDB structures, chemogenomics, etc.... At Evotec, we have created the Molecular Architects group which is composed of Computational Chemists and Drug Hunters working in synergy towards the identification of drugs for patients in need. Molecular Architects are trained and skilled to progress compounds using augmented design and rapid data-rich DMTA* cycles.



In this poster, we will present the Molecular Architects ways of working in close collaboration with *in-silico* R&D developers and synthetic chemists, all leveraging the Evotec drug discovery testing platform.

* DMTA: Design, Make, Test, Analyse

CONFORMATIONAL SAMPLING OF CYCLIC PEPTIDES WITH REMD AND SIMULATED TEMPERING

DIRK STRATMANN[a,b] , SAMUEL MURAIL[b]
JAYSEN SAWMYNADEN[a], MAUD JUSOT[a], FABIO PIETRUCCI[a]
JACQUES CHOMILIER[a], PIERRE TUFFÉRY[b]

[a] IMPMC, Sorbonne Université, Paris
[b] BFA, Université Paris Cité

Cyclic peptides are an important class of pharmaceutical drugs[1]. The exploration of their 3D conformations is important to rationally design novel drug compounds. Here we used molecular dynamics simulations to explore their conformational landscape[2,3]. Our test set consisted in nine head-to-tail cyclized peptides of 7 to 10 residues with a mixture of L and D amino acids for which NMR structures were available in the PDB. Free energy maps have been obtained using replica-exchange MD (REMD) in implicit solvent and simulated tempering (ST) in both implicit and explicit solvent and with several force fields. For most of the cyclic peptides the maps converge, but not for all test cases. The use of an explicit solvent did not improve the results, neither the use of a more recent force field. These results show that the use of a standard GBSA implicit solvent is still a good option for conformational sampling at least for cyclic peptides. This choice reduces the computational costs dramatically, especially for REMD simulations, as the number of replicas has to be ten times higher for a standard explicit solvent REMD simulation. In times of climat change and a necessary drastic reduction in energy consumption, it is very important to use computational resources efficiently.

References

[1] C. Lamers, “Overcoming the shortcomings of peptide-based therapeutics,” Future Drug Discov., vol. 4, no. 2, p. FDD75, Jun. 2022, doi: 10.4155/fdd-2022-0005.

[2] M. Jusot, D. Stratmann, M. Vaisset, J. Chomilier, and J. Cortés, “Exhaustive Exploration of the Conformational Landscape of Small Cyclic Peptides Using a Robotics Approach,” J. Chem. Inf. Model., vol. 58, no. 11, pp. 2355–2368, Nov. 2018, doi: 10.1021/acs.jcim.8b00375.

[3] J. Damjanovic, J. Miao, H. Huang, and Y.-S. Lin, “Elucidating Solution Structures of Cyclic Peptides Using Molecular Dynamics Simulations,” Chem. Rev., vol. 121, no. 4, pp. 2292–2324, Feb. 2021, doi: 10.1021/acs.chemrev.0c01087.

ALPHAFOLD-ASSISTED STRUCTURAL INSIGHTS INTO HUMAN NO-SYNTHASES REVEALING THE MECHANISM OF DIMERS

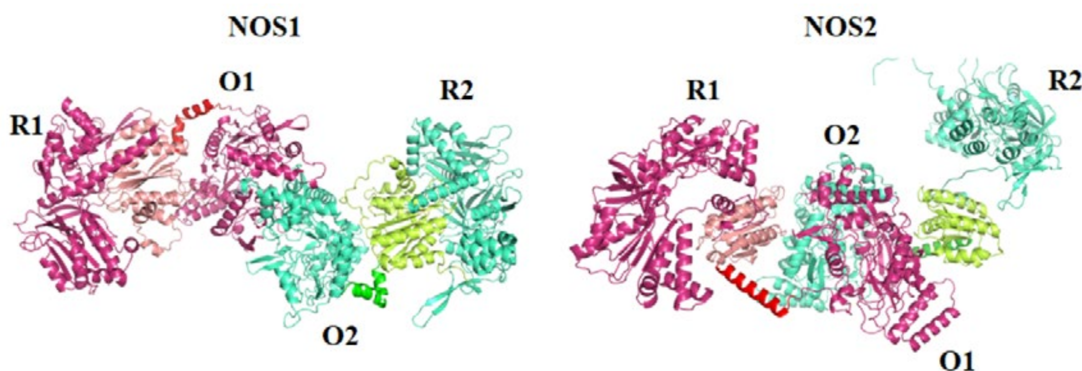
FRANÇOIS ANDRE, SATYA PRAKASH TRIPATHI, JÉRÔME SANTOLINI
I2BC (Institut de Biologie Intégrative de la Cellule), Laboratoire Stress Oxydant et Détoxication, CEA/CNRS/Université Paris-Saclay

Nitric oxide synthase (NOS) enzymes are complex redox proteins that catalyze the oxidation of L-arginine to citrulline resulting into the production of Nitric Oxide, essential secondary messenger in human physiological processes. The three isoforms of NOS in human are either inducible (NOS2) or constitutive (neuronal NOS1 and endothelial NOS3). They differ mainly in terms of their dependency for activation by Ca2+-dependent Calmodulin (CaM) binding. They share a good degree of sequence similarity, but their electron transfer regulation differs.

In NOSs the catalysis is compartmentalized in two principal domains, i.e., oxygenase and reductase domains, within and between which electron transfer occurs. Although the available information suggests that the bioactive conformation of NOS is a dimeric arrangement, complete structural information is lacking, limiting the enzyme's functional characterization. In the present work, for the first time, the Alphafold assisted complete dimeric structures of human NOS1, NOS2, and NOS3 isoforms, both with and without bound CaM, were used to explain the differences in electron regulation processes.

The generated AF models of dimers were thoroughly analyzed to identify the most plausible solution for the NOS structure. The predicted monomeric human NOS structures are available in Alphafold structural database, but not yet the dimeric conformations, while the dimer is the only one active form for NO synthesis. Due to the size of the isoforms dimers (between 2300 and 2860 residues), we benchmarked AlphaFold-based predictions using the available crystal structures of oxygenase domains and partial reductase domains. After assessing the prediction efficacy for NOS monomeric domains, we performed the prediction of dimeric and bioactive conformations of NOSs using Alphafold multimeric module. The biological assembly of enzymes with and without CaM were also prepared, as it contributes critically to the functional differences between isoforms. The different redox cofactors were then inserted, and molecular dynamics study performed to generate the optimized enzyme-cofactor-substrate complexes.

The dimeric arrangement of inducible NOS (NOS2) adopted an open confirmation whereas the constitutive NOSs adopted closed conformation because of the different orientation of oxygenase and reductase domain in their monomeric structure. The opened conformation in NOS2 facilitated the crisscross assembly of the monomers in the dimeric arrangement, where the reductase domain from one monomer (R1) forms a functional interface with oxygenase domain of the other monomer (O2), and vice versa (Figure). This arrangement could not be attained in NOS1 and NOS3 due to a closed conformation of each monomer. This is correlated to the reported functional roles of regulatory elements (AI –also called auto inhibitory region-, beta finger, C-terminal tail). The newly obtained structural information regarding the quaternary arrangement of the various isoforms support mechanistic understanding of the functional aspects.



IN SILLICO STUDIES FOR THE DEVELOPMENT OF A NON-COVALENT INHIBITOR OF UBE2N

JANA SOPKOVA-DE OLIVEIRA SANTOS[a], CÔME GHADI[a] CHARLINE KIEFFER[a], LOUIS-BASTIEN WEISWALD[b] LEONIE IBAZIZENE[b] ANNE SOPHIE VOISIN-CHIRET[a] MATTHIEU MERYET-FIGUIERE[b]

[a] CERMN (UR 4258), Université de Caen Normandie, UNICAEN, F-14032 Caen, France

[b] ANTICIPE (U1086), UNICAEN, INSERM, F-14032 Caen, France

Ovarian Cancer (OC) can be caused by malignant degeneration of the epithelial cells covering the surface of the ovary. OC are arduous to detect and often diagnosed late because they are for a long time. In the world, it accounts for 230,000 new cases and 150,000 deaths each year. They can be first treated by surgery. When surgery is impossible, chemotherapies are applied.[1] However, chemoresistances occur frequently, that is why there is an urgent need to develop molecules with a new mechanism of action. Recent studies have shown that ubiquitin E2 N conjugating enzyme (UBE2N, also called Ubc13) is a potential therapeutic target for several types of cancers.[2]

UBE2N is an E2 ubiquitin ligase that plays a role in response to DNA breaks. To do it, UBE2N interacts with Mms2 in the nucleus and Uev1A in cytoplasm. The UBE2N/Mms2 complex forms a thioester bond between the ubiquitin molecule and cysteine residue 87 of the UBE2N active site, which catalyses the formation of ubiquitin chains.[3] Finally, UBE2N plays a key role in DNA repair pathways.[4]

A flexible loop (residues 114-124) modulates the active site of UBE2N located around Cys87. The displacement of the loop generates an open or closed conformation and thus regulates the volume of the cavity to accommodate the ligand.[4] Currently, best UBE2N inhibitors described in the literature (UC-764864/65, NSC697923 and BAY 11-7082) binds covalently on UBE2N active site cysteine.[5] More, several performed in vitro screenings allowed to identify some non-covalent inhibitors of UBE2N.[6]

Aim of our study is to develop new non-covalent inhibitors of UBE2N. For this end, we apply an in silico strategy. By analysing the available 3D structures of UBE2N co-crystallised with various protein partners, we selected two pockets of interest for the design of competitive inhibitors. The first one is located near Cys87, targeted by covalent inhibitors, and the second one is on the binding surface with Mms2 or Uev1a. We performed in silico screening of the CERMN's chemolibrary (19,000 molecules) using docking approach on these two sites to select potential new non-covalent inhibitors. Three PDB structures of UBE2N (PDB code: 4ONM, 6UMP, 3HCU) representing a conformational variety of the flexible loop were used during the screenings. Then, molecular dynamics were applied to check the stability of the interaction between the selected compounds from screenings and UBE2N. They allowed to select some interesting candidates according to the interaction energy and the stability of the complex.

References

[1] Reid, B. M., Permuth, J. B., & Sellers, T. A. (2017). Epidemiology of ovarian cancer: a review. Cancer biology & medicine, 14(1), 9.

[2] Wu, X., Zhang, W., Font-Burgada, J., Palmer, T., Hamil, A. S., Biswas, S. K., ... & Karin, M. (2014). Ubiquitin-conjugating enzyme Ubc13 controls breast cancer metastasis through a TAK1-p38 MAP kinase cascade. Proceedings of the National Academy of Sciences, 111(38), 13870-13875.

[3] Rout, M. K., Hodge, C. D., Markin, C. J., Xu, X., Glover, J. M., Xiao, W., & Spyracopoulos, L. (2014). Stochastic gate dynamics regulate the catalytic activity of ubiquitination enzymes. Journal of the American Chemical Society, 136(50), 17446-17458.

[4] Hodge, C. D., Spyracopoulos, L., & Glover, J. M. (2016). Ubc13: the Lys63 ubiquitin chain building machine. Oncotarget, 7(39), 64471.

[5] Hodge, C. D., Edwards, R. A., Markin, C. J., McDonald, D., Pulvino, M., Huen, M. S., ... & Glover, J. M. (2015). Covalent inhibition of Ubc13 affects ubiquitin signaling and reveals active site elements important for targeting. ACS chemical biology, 10(7), 1718-1728.

[6] Ardecky, R., Madiraju, C., Matsuzawa, S. I., Zou, J., Ganji, S., Pass, I., ... & Reed, J. C. (2014). Selective UBC 13 Inhibitors. Probe Reports from the NIH Molecular Libraries Program [Internet].

INVESTIGATION ON THE PROTEIN DYNAMICS AT EXTREME TEMPERATURES

BEATRICE CAVIGLIA[A,C,D]
SUPERVISORS: JUDITH PETERS 1[B], ALESSANDRO PACIARONI 2[C] FABIO STERPONE 3[D]

[a] Université de Paris Cité, 45 Rue des Saints-Pères, Paris, France
[b] Université Grenoble Alpes, CNRS, Laboratoire Interdisciplinaire de Physique, 38400Saint-Martin-d'Héres, France
[c] Università degli Studi di Perugia, Dipartimento di Fisica e Geologia, Via A. Pascoli, 06123Perugia PG, Italy
[d] Laboratoire de Biochimie Théorique (UPR9080), CNRS, Université de Paris Cité, 13 Rue Pierre et Marie Curie, 75005Paris, France

Life has adapted to extreme conditions on Earth. One of the most striking evidences of adaptation to extreme environments are bacteria that are capable of thriving in a vast temperature range, from below 0° C in glacial waters to above 100° C in deep-sea hydrothermal vents. It is known that the individual molecular components of these organisms, exhibit enhanced stability and resistance to the temperature stress. A research focus lies on the proteins, which are the most abundant and less stable macromolecules in the cell. However, the link among the individual protein stability, and the process of cell death caused by the raise of temperature is not yet clear. Understanding the biophysical determinants of cellular thermostability would be fundamental from a theoretical, biotechnological and clinical prospective [1], [2]. In a theoretical work by Dill and his co-workers it has been proposed that the temperature induced cell death follows a collective unfolding of the proteome, with the entire set of proteins approaching their individual melting in a narrow temperature range [3].

However, this picture has been recently challenged by further experimental and simulation work [4], [5]. Experimentally, it was possible to monitor the amount of proteins actually unfolding at the cell death temperature. These recent studies showed that it is only a small fraction of proteins that unfold at the cell death temperature. This was further confirmed by Sterpone and his co-workers, who investigated the dynamical profile of the proteins in the E. Coli, a mesophilic bacterium. Using a combination of Neutron-scattering experiments and Molecular dynamics simulations they showed that a dynamical catastrophe occurs at the cell death temperature which is caused by only around 10% of the proteins unfolding [6]. The goal of this project is to use coarse-grained and all-atom Molecular Dynamics simulations and Neutron Scattering experiments to further investigate the dynamics of the proteins in the case of extremophilic bacterium and to reveal whether it is a small set of proteins that unfold as it has been shown for the mesophile. The bacteria selected for this investigation are the P. Arcticus, a psychrophile, and the A. Aeolicus, a hyperthermophile. Different dynamical aspects that have been studied so far using MD simulations, such as the global diffusion coefficient and the mean square displacement of the proteins selected for these two bacteria, will be presented.

References

[1] Dill, K. A. , Ghosh, K. , Schmit, J. D. (2011) 'Physical limits of cells and proteomes' PNAS 108,17876

[2] Coffey, D. S. , Getzenberg, R. H. , DeWeese, T. L. (2006) 'Hyperthermic biology and cancer therapies: A hypothesis for the Lance Armstrong effect' JAMA 296, 445.

[3] Polizzi, K. M., Bommarius, A. S. , Broering, J. M. , Chaparro-Riggers, J. F. (2007) 'Stability of Biocatalysts' Curr. Opin. Chem. Biol. 11, 220

[4] Leuenberger, P. et al. (2017) 'Cell-wide analysis of protein thermal unfolding reveals determinants of thermostability', Science, 355, 6327

[5] Jarazab, A. et al. (2020) 'Meltome atlas—thermal proteome stability across the tree of life', Nature Methods, 17, 495.

[6] Daniele Di Bari, Stepan Timr, Marianne Guiral, Marie-Thérèse Giudici-Orticoni, Tilo Seydel, Christian Beck, Caterina Petrillo, Philippe Derreumaux, Simone Melchionna, Fabio Sterpone, Judith Peters, and Alessandro Paciaroni (2023) 'Diffusive Dynamics of Bacterial Proteome as a Proxy of Cell Death', ACS Central Science 9 (1), 93-102

MACHINE LEARNING-BASED PREDICTION OF AMPC β-LACTAMASE INHIBITION: FROM KNOWN MOLECULES TO DESIGN NEW ANTIMICROBIAL AGENTS.

MARION SISQUELLAS[ab] , YOUCEF BAGDAD[a]
KÉVIN CARIOU[c], MÉLANIE ETHEVE-QUELQUEJEU[d],
MICHEL ARTHUR[e], MARIA A. MITEVA[a]

[a] CiTCoM UMR8038, Inserm U1268, 4 Av de l'Observatoire, 75006 Paris, France
[b] Institut Cochin Inserm U1016, 22 Rue Mechain 75014 Paris, France
[c] Chimie ParisTech, PSL University, CNRS, 11 Rue Pierre et Marie Curie, 75005 Paris, France
[d] CNRS UMR 8601, LCBPT, 45 rue des Saints Pères, 75270 Paris, France
[e] UMRS 1138-Centre de Recherche des Cordeliers, 15 rue de l'école de médecine, 75006 Paris, France

Antimicrobial resistance is a major problem that has been growing steadily in recent years, causing millions of deaths[1]. The emergence of multi-drug resistance (MDR) is particularly found among Entero-bacteriaceae such as Escherichia coli (E. coli). E. coli causes serious infections and have multiple resistance mechanisms, the most common being extended-spectrum β-lactamase (ESBL) and AmpC β-lactamase production[2,3]. One of the main mechanisms underlying resistance to β-lactam antibiotics are the AmpC β-lactamases[4]. In this study, we employ in silico approaches to identify new inhibitors of AmpC β-lactamase. First, we collected 384223 compounds experimentally tested on E. coli AmpC. The curation of these data has led to 891 inhibitors and 81720 non-inhibitors of AmpC β-lactamase. We used these compounds to develop new classification machine learning (ML) models to predict putative inhibitors of this enzyme. Then, we used generative adversarial networks (GAN) to develop generative models in order to design new molecules capable to inhibit AmpC β-lactamase.

References

[1] Antimicrobial Resistance Collaborators, «Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis», Lancet Lond. Engl., vol. 399, no 10325, p. 629 655, févr. 2022, doi: 10.1016/S0140-6736(21)02724-0.

[2] T. D. Institute, « WHO global priority pathogens list of antibiotic-resistant bacteria », 2 décembre 2021. https://www.doherty.edu.au/newsevents/news/who-global-priority-pathogens-list-of-antibiotic-resistant-bacteria/.

[3] C. A. Michael, D. Dominey-Howes, et M. Labbate, « The Antimicrobial Resistance Crisis: Causes, Consequences, and Management », Front. Public Health, vol. 2, 2014, https://www.frontiersin.org/articles/10.3389/fpubh.2014.00145

[4] S. Lee et al., « Prevalence, Characteristics and Clonal Distribution of Extended-Spectrum β-Lactamase- and AmpC β-Lactamase-Producing Escherichia coli Following the Swine Production Stages, and Potential Risks to Humans », Front. Microbiol., vol. 12, https://www.frontiersin.org/articles/10.3389/fmicb.2021.710747

DEVELOPMENT OF A PREDICTIVE TOOL TO ASSIST THE DIAGNOSIS IN MYOPATHIES RYR1 VARIANTS

N. DEBBAH[1,2,3], A. FRENOY[4], M. LEGER[4], I. MARTY[4], V. JACQUEMOND[5] K. JAGLA[6], J. FAURE[1], A. THOMAS[3], J. RENDU[1]

[1] Univ. Grenoble Alpes, Inserm, U1216, Chu Grenoble Alpes - La Tronche
[2] CNRS Alpes (laboratoire TIMC, UMR5525) - La Tronche
[3] Departement de Pharmacochimie Moléculaire, UMR CNRS 5063 - Saint-Martin-D'hères
[4] Univ. Grenoble Alpes, Inserm, U1216, Chu Grenoble Alpes, Grenoble Institut Neurosciences - La Tronche
[5] Institut Neuromyogène (INMG) /CNRS Umr CNRS 5310, INSERM U1217 - Lyon
[6] Institut Génétique Reproduction Et Développement IGRED-Umr Inserm 1103-CNRS 6293-UCA – Clermont Ferrand

Background: The human RyR1 is a Calcium Channel protein belonging to the Ryanodine Receptor family. Traversing in the sarcoplasmic membrane, It is folded as a homotetramer of 5038 residues organized in 20 differents domains. It allows the exit of Calcium under the effect of an Excitation potential and therefore is essential in the process of excitation-contraction for skeletal muscle fibers.

Mutations in the human RyR1 are implicated in a large panel of neuromuscular disorders ranging from fetal akinesia to per anesthesic malignant hyperthermia susceptibility[1]. More than 2840 variants of the RyR1 gene have been identified, many with unknown clinical significance, which leads to a diagnosis deadlock for medical experts. The patients have to overcome many medical test that can last for long periods before receiving a clear diagnosis.

Recent advances in Cryo-ElectroMicroscopy allowed an increasing number of RyR1 structure, notably in the Oryctolagus cuniculus or the Sus Scrofa[2] with different physiological modulators and in different states such as closed, inactivated or open conformations.

Objective: A reliable classification of RYR1 variants of unknown significance is thus needed; the objective of our project is to create an efficient predictive pipeline to help medical diagnosis of RyR1 related myopathies.

Mehods: To reach this aim, we built a transdisciplinary project involving multiple partners. We combine homology modeling[3], computational biology with an AI approach, understanding in fundamental biology and clinical genetic data to build a predictive tool for diagnosis that will be validated by functionnal study.

Results: So far, we have

- Built models of human RyR1 protein, in both in its closed and open conformations using as templates the extremely close rabbit RYR1 structure.
- Modeled 2840 Single amino-acid variants of human RyR1 in their local environment selected from *Clinvar*[4] and *Gnomad*[5] clinical databases. Variants are either of unknown significance, benigns or contains a pathogenic mutation.
- Developed a series of physico-chemical, structural descriptors (e.g. distance to residue, proximity to known binding sites) as well as evolutionary conservation descriptors and local energy variation.
- Created a first dataset with these descriptors to be used for subsequent various Machine Learning based methods.
- Obtained preliminary results tackling our first dataset with supervised methods.

Conclusion: The development of our predictive tool relies on a novel integrative approach combining genetics, molecular modeling and AI methodology. This mutidisciplnary project started in October 2021, the first step of modeling have been validated. Our goal is to propose this work as an approach for clinician to perform diagnosis on RyR1 variants.

References

[1] Lawal TA, Todd JJ, Witherspoon JW, Bönnemann CG, Dowling JJ, Hamilton SL, Meilleur KG, Dirksen RT. Ryanodine receptor 1-related disorders: an historical perspective and proposal for a unified nomenclature. Skelet Muscle. 2020 Nov 16;10(1):32. doi: 10.1186/s13395-020-00243-4. PMID: 33190635; PMCID: PMC7667763.

[2] Woll KA, Van Petegem F. Calcium-release channels: structure and function of IP3 receptors and ryanodine receptors. Physiol Rev. 2022 Jan 1;102(1):209-268. doi: 10.1152/physrev.00033.2020. Epub 2021 Jul 19. PMID: 34280054.

[3] Jacobson, M. P.; Pincus, D. L.; Rapp, C. S.; Day, T. J. F.; Honig, B.; Shaw, D. E.; Friesner, R. A.: A Hierarchical Approach to All-Atom Protein Loop Prediction, Proteins: Structure, Function and Bioinformatics. 2004; 55: 351-367.

[4] Landrum MJ, Lee JM, Benson M, Brown GR, Chao C, Chitipiralla S, Gu B, Hart J, Hoffman D, Jang W, Karapetyan K, Katz K, Liu C, Maddipatla Z, Malheiro A, McDaniel K, Ovetsky M, Riley G, Zhou G, Holmes JB, Kattman BL, Maglott DR. ClinVar: improving access to variant interpretations and supporting evidence. Nucleic Acids Res. 2018 Jan 4;46(D1):D1062-D1067. doi: 10.1093/nar/gkx1153. PMID: 29165669; PMCID: PMC5753237.

[5] Gudmundsson S, Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, Collins RL, Laricchia KM, Ganna A, Birnbaum DP, Gauthier LD, Brand H, Solomonson M, Watts NA, Rhodes D, Singer-Berk M, England EM, Seaby EG, Kosmicki JA, Walters RK, Tashman K, Farjoun Y, Banks E, Poterba T, Wang A, Seed C, Whiffin N, Chong JX, Samocha KE, Pierce-Hoffman E, Zappala Z, O'Donnell-Luria AH, Minikel EV, Weisburd B, Lek M, Ware JS, Vittal C, Armean IM, Bergelson L, Cibulskis K, Connolly KM, Covarrubias M, Donnelly S, Ferriera S, Gabriel S, Gentry J, Gupta N, Jeandet T, Kaplan D, Llanwarne C, Munshi R, Novod S, Petrillo N, Roazen D, Ruano-Rubio V, Saltzman A, Schleicher M, Soto J, Tibbetts K, Tolonen C, Wade G, Talkowski ME; Genome Aggregation Database Consortium; Neale BM, Daly MJ, MacArthur DG. Addendum: The mutational constraint spectrum quantified from variation in 141,456 humans. Nature. 2021 Sep;597(7874):E3-E4. doi: 10.1038/s41586-021-03758-y. Erratum for: Nature. 2020 May;581(7809):434-443. PMID: 34373650; PMCID: PMC8410591.

TOWARDS MARKOV STATE MODELS OF CHEMICALLY DRIVEN NON-EQUILIBRIUM STEADY STATES

EMANUELE ZIPPO[a]
LUKAS STELZL[a,b]
[a] Johannes Gutenberg University Mainz
[b] Institute of Molecular Biology (IMB) Mainz

Non-equilibrium steady states (NESS) are a fundamental concept in biophysics, as many biological systems operate far from thermodynamic equilibrium, driven by external forces or internal gradients. An example of chemically driven NESS would be the system composed of an enzyme and a substrate that can be converted to product by burning a chemical fuel, such as ATPs. In this work, we investigate the effect of phosphorylation in the liquid-liquid phase separation of the intrinsically disordered protein TDP-43. In particular, we show how the enzyme CK1δ can dissolve a TDP-43 droplet through phosphorylation. We address the problem of simulating chemical reactions using a mixture of coarse-grained implicit solvent molecular dynamics model and Monte Carlo steps. This kind of system also has biological relevance as TDP-43 aggregates may play a role in the development of neurodegenerative diseases, such as ALS, FTD or Alzheimer's disease. We also focus on the use of Markov state models (MSMs) in non-equilibrium processes. MSMs are a powerful tool for studying complex dynamics, as they need only information about local equilibrium within conformational macro-states to compute transition rates. Our aim is to build MSMs from simulation trajectories in order to get a coarse-graining in time and an interpretable representation of the dynamics of our system

COMPUTATIONAL STUDIES OF AND STRUCTURAL INSIGHTS INTO FREE FATTY ACID RECEPTOR 4 (FFAR4) IN THE TREATMENT OF OBESITY AND RELATED DISEASES

NICOLAS RENAULT[a], GUILLAUME PATIENT[a]
NAÏM KHAN[b], AMAURY FARCE[a]

[a] U1286 INFINITE Inserm, Université de Lille, CHU LILLE, Lille, France
[b] U1231 Inserm, AgroSup, Université de Bourgogne, Dijon, France

The G protein-coupled Receptors (GPCR) are important drug targets representing 35 % of total approved druggable targets. These receptors are transmembrane proteins involved in cellular signaling of external stimuli (hormones, neurotransmitters, ions…) and induce the activation of the associated G protein via GDP/GTP exchange. In the last 20 years, deorphanized GPCRs have been highlighted as being active in lipid metabolism: Free Fatty Acid Receptors (FFAR).[1] Studies have shown that modulating the activity of GPR 120 (FFAR4), mostly expressed in the tongue and intestine, has an impact on the lipid intake (feeding) and absorption in mice thus representing a novel approach to promote treatments in preventing obesity and cardiovascular diseases.[2,4] Complementary to classic pharmaceutical chemistry studies in collaboration with the team of Pr. Naim Khan at University of Bourgogne, in silico methods have been applied to the study of the structural features of the protein.

The lack of structural information on GPR 120 and limited drug-testing contribute to applying in silico methods to determine structural features and dynamic interactions of the protein. Homology modelling of different states of GPR 120 (inactive, intermediate, and active state) has been performed by screening the references of the PDB database and numerous templates of Class A rhodopsin-like GPCR proteins have been selected. Associated G protein has been modelized for the active state of the receptor and dynamic studies of the receptor in a membrane resulted in a refined bundle of transmembrane helixes (TM). Identification of the active site and ligand interaction enabled preliminary docking studies of TUG-891, a selective potent agonist of GPR 120, and highlighted ionic interactions with residue ARG 99 labelled as an orthosteric binding site of GPR 120 for agonist compounds. Antagonist docking pinpointed the role of the side-chain conformation of TYR 299 in the active site.

Docking studies of 3 agonist ligands (TUG-891, TUG-1197 and linolenic acid) have been performed and results have been refined using molecular dynamics with AMBER 14IPQ forcefield to consider ionic interactions between ARG99 in the binding pocket and the carboxylic acid group of the ligands. Triplicate experiments of molecular dynamics for the different systems protein-ligand enabled the identification of the most stable binding mode and enabled us to determine pharmacophoric constraints for a High-Throughput Virtual Screening.

References

[1] Hara T, Kimura I, Inoue D, Ichimura A, Hirasawa A. Free Fatty Acid Receptors and Their Role in Regulation of Energy Metabolism. In: Nilius B, Amara SG, Lill R, Offermanns S, Gudermann T, Petersen OH, et al., éditeurs. Reviews of Physiology, Biochemistry and Pharmacology, Vol 164 [Internet]. Cham: Springer International Publishing; 2013 [cité 17 nov 2021]. p. 77-116. (Reviews of Physiology, Biochemistry and Pharmacology; vol. 164). Disponible sur: http://link.springer.com/10.1007/112_2013_13

[2] Murtaza B, Hichami A, Khan AS, Shimpukade B, Ulven T, Ozdener MH, et al. Novel GPR120 agonist TUG891 modulates fat taste perception and preference and activates tongue-brain-gut axis in mice. J Lipid Res. févr 2020;61(2):133-42.

[3] Martin C, Passilly-Degrace P, Gaillard D, Merlin JF, Chevrot M, Besnard P. The Lipid-Sensor Candidates CD36 and GPR120 Are Differentially Regulated by Dietary Lipids in Mouse Taste Buds: Impact on Spontaneous Fat Preference. Blanc S, éditeur. PLoS ONE. 25 août 2011;6(8):e24014.

[4] Khan AS, Hichami A, Murtaza B, Louillat-Habermeyer ML, Ramseyer C, Azadi M, et al. Novel Fat Taste Receptor Agonists Curtail Progressive Weight Gain in Obese Male Mice. Cell Mol Gastroenterol Hepatol. 2023;15(3):633-63.

SUBTLE SEQUENCES CHANGES IN THE CODONS OF TP53 MAY CAUSE DNA REPAIR RATE DIFFERENCES: A MOLECULAR DYNAMICS STUDY

JAMES DAVIES[1]
PROFESSOR JAMIE PLATTS[2]
DR GEORGINA MENZIES[1]

[1] Molecular Bioscience Division, School of Biosciences, Cardiff, Wales
[2] School of Chemistry, Cardiff University, Cardiff, Wales

Somatic mutations in the tumour suppressor gene TP53 occur in around 50% of human cancers and are considered the most common mutational event among lung cancer subtypes[1]. In smoking-associated lung cancers, the occurrence of characteristic mutations at hotspot sites, e.g., codons 157, 158, 196 ,245, 248, 273, and 282 has been linked to the binding of carcinogenic chemicals in tobacco smoke including benzo[a]pyrene (B[a]P)[2]. The ability of B[a]P to bind to DNA, however, extents to numerous additional sites that are incapable of initiating carcinogenesis. This suggests that the rate at which these carcinogens are removed is likely to govern an individuals associated cancer risk. Interestingly, the efficiency of DNA repair processes is shown to vary substantially according to the position of damage within the genome, thereby implicating DNA context as a potential modulator of chemical toxicity. By performing molecular modelling of B[a]P within the context of the TP53 gene, we aim to decipher what properties of the DNA context surrounding a B[a]P damage site impact the efficiency at which it is repaired. Using molecular modelling techniques, we will model B[a]P-damaged DNA sequences representative of TP53 hotspot and non-hotspot sites. This modelling will occur first individually to determine the degree of structural changes imposed on the DNA by B[a]P binding, and then in the presence of a DNA repair protein to evaluate the influence of this distortion on repair. Alongside this, we hope to develop an automated analytical pipeline that may be used to isolate the precise structural variables that drive poor repair among TP53 hotspot codons and therefore predict any statistical correlation between damage position and mutation probability. The results generated will provide crucial information surrounding the aetiology of this mutation pattern that plagues patient suffering from smoking-associated lung cancer.

References

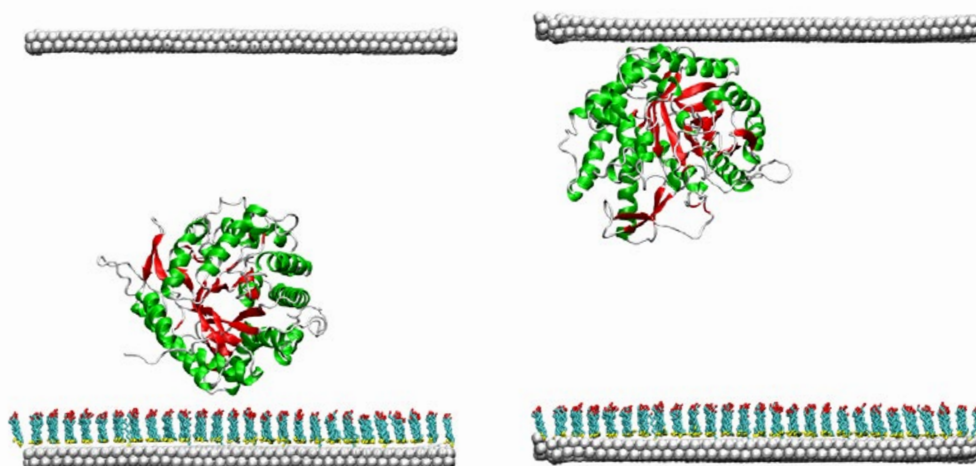
[1] Baugh, E., Ke, H., Levine, A. et al. Why are there hotspot mutations in the TP53 gene in human cancers?. Cell Death Differ 25, 154–160 (2018).

[2] Yu D, Berlin JA, Penning TM, Field J. Reactive oxygen species generated by PAH o-quinones cause change-in-function mutations in p53. Chem Res Toxicol. 2002 Jun;15(6):832-42. doi: 10.1021/tx010177m. PMID: 12067251.

BETWEEN TWO WALLS: MODELING THE ADSORPTION BEHAVIOR OF REDOX ENZYMES ON BARE AND SAM-FUNCTIONALISED GOLD SURFACES

SOPHIE SACQUIN-MORA[a,b], NICOLAS BOURASSIN[A,B]
FLORENT BARBAULT[C], MARC BAADEN[A,B]
[a] CNRS, Université de Paris, UPR 9080, Laboratoire de Biochimie Théorique, 13 rue Pierre et Marie Curie, 75005 Paris, France
[b] Institut de Biologie Physico-Chimique-Fondation Edmond de Rotschild, PSL Research University, Paris, France
[c] Université de Paris, ITODYS, CNRS, F-75006, Paris, France

The efficient immobilization of enzymes on surfaces remains a complex but central issue in the biomaterials field[1], which requires us to understand this process at the atomic level. Using a multi-scale approach combining all-atom molecular dynamics and coarse- grain Brownian dynamics simulations, we investigated the adsorption behavior of β-glucosidase A (βGA) on bare and SAM-functionalized gold surfaces[2]. We monitored the enzyme position and orientation during the MD trajectories, and measured the contacts it forms with both surfaces. While the adsorption process has little impact on the protein conformation, it can nonetheless perturb its mechanical properties and catalytic activity[3]. Our results show that compared to the SAMfunctionalized surface, the adsorption of βGA on bare gold is more stable, but also less specific, and more likely to disrupt the enzyme's function. This observation emphasizes the fact that the structural organization of proteins at the solid interface is a keypoint when designing devices based on enzyme immobilization, as one must find an acceptable stability-activity trade-off[4].



References

[1] Hitaishi, V.; Clement, R.; Bourassin, N.; Baaden, M.; de Poulpiquet, A.; Sacquin-Mora, S.; Ciaccafava, A.; Lojou, E., Controlling Redox Enzyme Orientation at Planar Electrodes. Catalysts 2018, 8 (5), 192.

[2] Bourassin, N.; Barbault, F.; Baaden, M.; Sacquin-Mora, S., Between two walls : Modeling the adsorption behavior of β- glucosidase on bare and SAM-functionalised gold surfaces, Langmuir 2022, 38, 1313–1323

[3] Bourassin, N.; Baaden, M.; Lojou, E.; Sacquin-Mora, S., Implicit Modeling of the Impact of Adsorption on Solid Surfaces for Protein Mechanics and Activity with a Coarse-Grained Representation. J Phys Chem B 2020, 124 (39), 8516-8523.

[4] Weltz, J. S.; Kienle, D. F.; Schwartz, D. K.; Kaar, J. L., Reduced Enzyme Dynamics upon Multipoint Covalent Immobilization Leads to Stability-Activity Trade-off. J. Am. Chem. Soc. 2020, 142 (7), 3463-3471.

MOLECULAR MODELLING OF PLASMODESMATA ORGANIZATION BY MCTP PROTEINS

SUJITH SRITHARAN[1] sritharan@ibpc.fr
ANTOINE TALY[1]
EMMANUELLE BAYER[2]
[1] Laboratoire de Biochimie Théorique, UPR9080, CNRS, Université Paris Cité, Paris, France
[2] Laboratoire de Biogenèse Membranaire, UMR5200, CNRS, Université de Bordeaux, Villenave d'Ornon, France

In plants, intercellular communication is primarily achieved through plasmodesmata. These membrane pores cross the cell wall and create symplastic continuity between cells[1]. Plasmodesms are crucial in coordinating developmental processes and defence mechanisms against pathogens[3]. They are also hijacked by viruses that can structurally modify them to propagate their viral genome from cell to cell.

Plasmodesms have a unique membrane organization: they are crossed by a «tube» of endoplasmic reticulum (ER), which is in intimate contact with the plasma membrane (PM), delimiting the pores. The two membranes are only a few nm apart (~10 nm) and connected by «tethers». The multiple C2 domains and transmembrane region protein (MCTP) family, critical regulators of cell-to-cell signalling in plants, act as ER-PM tethers, specifically at plasmodesmata[2,3]. However, the molecular mechanism and function of membrane tethering within plasmodesmata remain unknown. Furthermore, MCTP proteins are still poorly known at the level of the 3D structure.

Thus, we first generated structural models of A. thaliana MCTP by different prediction methods using deep learning like AlphaFold and RosettaFold [4,5]. Then, we compare them by computing the rsmd and contact maps. Further, the movement of transmembrane regions in the lipid bilayer was characterised by coarse-grained simulation using the MARTINI3 force field and principal components analysis. We were finally able to extract representative conformations from our simulations.

References

[1] K. Knox et al., « Putting the Squeeze on Plasmodesmata: A Role for Reticulons in Primary Plasmodesmata Formation », Plant Physiol., vol. 168, no 4, p. 1563-1572, août 2015, doi: 10.1104/pp.15.00668.

[2] « Multiple C2 domains and transmembrane region proteins (MCTPs) tether membranes at plasmodesmata », EMBO Rep., vol. 20, no 8, p. e47182, août 2019, doi: 10.15252/embr.201847182.

[3] J. D. Petit, Z. P. Li, W. J. Nicolas, M. S. Grison, et E. M. Bayer, « Dare to change, the dynamics behind plasmodesmata-mediated cell-to-cell communication », Curr. Opin. Plant Biol., vol. 53, p. 80-89, févr. 2020, doi: 10.1016/j.pbi.2019.10.009.

[4] M. Baek et al., « Accurate prediction of protein structures and interactions using a three-track neural network », Science, vol. 373, no 6557, p. 871-876, août 2021, doi: 10.1126/science.abj8754.

[5] J. Jumper et al., « Highly accurate protein structure prediction with AlphaFold », Nature, vol. 596, no 7873, Art. no 7873, août 2021, doi: 10.1038/s41586-021-03819-2.

MULTI-SCALE MODELING TO IDENTIFY POTENTIAL LIGANDS OF DCIR

MAGDALENA SZCZUKA[a]
MATTHIEU CHAVENT[a]
[a] Institut de Pharmacologie et de Biologie Structurale, Université de Toulouse, CNRS, Université Paul Sabatier, 31400 Toulouse, France

DCIR is a C-type lectin receptor expressed on the surface of dendritic cells[1]. It plays a role in recognizing and binding to specific molecules on pathogens, which helps trigger an immune response to fight off infections. DCIR has been shown to be involved in the immune response against a variety of pathogens, including viruses such as HIV, hepatitis C virus, and influenza virus, as well as bacteria such as Mycobacterium tuberculosis and Streptococcus pneumoniae. The identification of ligands that can selectively bind to DCIR can provide a better understanding of its function and potential therapeutic targets. Thus, the objective of this study is to identify potential ligands or molecular functional groups that can interact with DCIR using a multi-scale modeling approach.

To achieve this, a library of small molecules[2] is parsed, and MD simulations are performed using the Martini3 force field[3,4]. The use of beads from the Martini3 force field enables the treatment of functional groups of ligands separately, allowing for specific mapping of DCIR and identification of potential binding sites that may be specific to those functional groups. The developed workflow enables a comprehensive analysis of the simulation results, allowing for the identification of key binding sites and potential ligands of DCIR.

This study provides insights into the molecular interactions of DCIR and may lead to the development of new therapies that target this receptor. In perspective, this workflow could be applied to other receptor-ligand systems, enabling a more comprehensive understanding of molecular interactions.

References

[1] Nagae, M., Ikeda, A., Hanashima, S., Kojima, T., Matsumoto, N., Yamamoto, K. and Yamaguchi, Y. (2016), Crystal structure of human dendritic cell inhibitory receptor C-type lectin domain reveals the binding mode with N-glycan. FEBS Lett, 590: 1280-1288. https://doi.org/10.1002/1873-3468.12162

[2] Hilpert, C., Beranger, L., Souza, P.C.T., Vainikka, P.A., Nieto, V., Marrink, S.J., Monticelli, L.&; Launay, G. (2022) Facilitating CG simulations with MAD: the MArtini Database Server. BiorXiv. http://dx.doi.org/10.1101/2022.08.03.502585.

[3] Alessandri, R., Barnoud, J., Gertsen, A.S., Patmanidis, I., de Vries, A.H., Souza, P.C.T. and Marrink, S.J. (2022), Martini 3 Coarse-Grained Force Field: Small Molecules. Adv. Theory Simul., 5: 2100391. https://doi.org/10.1002/adts.202100391

[4] Souza, P.C.T., Alessandri, R., Barnoud, J. et al. Martini 3: a general purpose force field for coarse-grained molecular dynamics. Nat Methods 18, 382–388 (2021). https://doi.org/10.1038/s41592-021-01098-3

VIBRATIONAL SPECTROSCOPIC PROPERTIES OF CHROMOPHORES IN SOLUTION VIA TIME-SERIES ANALYSIS ASSISTED BY MACHINE LEARNING

ABIR KEBABSA[a], PHILIPPE P. LAINÉ[a]
FRANÇOIS MAUREL[a], ÉRIC BRÉMOND[a]

[a] Université Paris Cité, ITODYS, CNRS, F-75013 Paris, France

Infrared and Raman spectroscopies are governed by the absorption of resonant frequencies characterizing specific normal modes of vibration of a molecule. The identification of these normal modes along a molecular dynamics trajectory is commonly performed via time-series analysis. This numerical technique deciphers how a variable change over time. Applying it to the dipole moment or polarizability tensor time evolution results in the reconstruction of the corresponding vibrational spectrum of the molecule.[1]

Computing a property such as the polarizability tensor along a sufficiently long-time dynamics trajectory remains demanding both in terms of time and computational resources, especially when the property is derived from quantum mechanics (QM). In this poster, we show that by assisting a time-series analysis with machine learning techniques, we succeed to alleviate the computational cost and improve the dynamics sampling of a QM/MM trajectory. We specifically apply this protocol to the reconstruction of the Infrared and Raman spectra of a particular class of chromophore stable in water solution, the ADOTA+ triangulenium carbocation.[2]

References

[1] Thomas M., Brehem M., Fligg R., Vöhringer R., Kirchner B. Phys., Chem. Chem. Phys., 2013, 15, 6608.
[2] Barsuk I., Lainé P.P., Maurel F., Brémond E., Phys. Chem. Chem. Phys. 2020, 22, 20673.

MODELING VIBRATIONAL SPECTRA OF ORGANIC CHROMOPHORES IN SOLUTION: INSIGHTS FROM HYBRID QM/MM APPROACHES

ABIR KEBABSA[a], PHILIPPE P. LAINÉ[a], FRANÇOIS MAUREL[a], ÉRIC BRÉMOND[a]

[a] Université Paris Cité, ITODYS, CNRS, F-75013 Paris, France

As all spectroscopic properties, the vibrational response of a molecule is altered by the environment. The solvent molecular distribution around the solute strongly affects its vibrational motion and results in a modification of its spectroscopic signature. It becomes thus crucial to explicitly take into account the environment with a hybrid QM/MM approach while targeting the accurate modeling of such a property.

In this talk, we show that by performing a time-series analysis of the dipole moment and polarizability tensor along a QM/MM molecular dynamics trajectory, we successfully model the vibrational InfraRed and Raman spectra of an organic chromophore such as azadioxatriangulenium[1] (ADOTA+) in water solution.[2] By comparing them with InfraRed and Raman spectra derived from the gas phase or at static level, we especially measure which vibrational modes are prone to solvent effects and anharmonicity. Furthermore, we explore the potential of machine learning to enhance and accelerate the QM/MM simulation of vibrational spectra.

References

[1] I. Barsuk, P.P. Lainé, F. Maurel, É. Brémond Physical Chemistry and Chemical Physics 22 (2020) 20673-20684.

[2] M. Thomas, M. Brehm, R. Fligg, P. Vöhringerb and B. Kirchner Physical Chemistry Chemical Physics 15 (2013) 6608-6622.

HELIGEOM: A COMPUTATIONAL TOOL FOR EXPLORING POLYMORPHISM IN OLIGOMERIC PROTEIN ASSEMBLIES

CHANTAL PRÉVOST
CHARLES ROBERT
HUBERT SANTUZ
BENOIST LAURENT

Laboratoire de Biochimie Théorique, CNRS UPR9080 et Université de Paris, Institut de Biologie Physico-Chimique, Paris

Oligomeric protein assemblies are widely present in the cell. They can be found as filaments or rings and are involved in various cell functions ranging from scaffolding (support for cell walls, internal communication networks) to signal transduction or motors (cell mobility, maintenance of the genetic material, membrane crossing). These assemblies may modify their morphology as a response to perturbations such as the binding of external macromolecules or small molecules (ATP, drugs) or change in chemical environment (salt concentration, pH). Other factors contribute to morphology changes as well, including the application of mechanical forces such as torsion or stretching. The response of the oligomeric assembly to these external perturbations is often key to their function. In addition to major reorganization of monomer structures and interactions, even small changes in the interfaces can result in changes of the overall shape of the oligomer. We have developed Heligeom[1], an analysis and construction tool in the PTools library[2], as a simple way to explore the relationship between pairwise binding geometries between consecutive monomers and the global shape of the resulting assembly[3]. Examples of Heligeom applications in integrative modeling will be discussed.

References

[1] Boyer B, Ezelin J, Poulain P, Saladin A, Zacharias M, Robert CH, Prévost C. An integrative approach to the study of filamentous oligomeric assemblies, with application to RecA. PLoS ONE, 2015, 10-e0116414

[2] Saladin A, Fiorucci S, Poulain P, Prévost C, Zacharias M. PTools: an opensource molecular docking library. BMC Structural Biology, 2009, 9-27

[3] Boyer B, Laurent B, Robert CH, Prévost C. Modeling Perturbations in Protein Filaments at Micro and Meso Scales with NAMD and PTools/Heligeom. Bio-protocol, 2021, 11-e4097

MARVOK STATE MODELS TO STUDY SYNAPTIC NEUROTRANSMISSION AT THE MOLECULAR LEVEL

PAULA MILÁN RODRÍGUEZ
MARCO CECCHINI
Laboratoire d'Ingénierie des Fonctions Moléculaires, Université de Strasbourg

Glycine receptors (GlyRs) are pentameric ligand-gated ion channels (pLGICs) that convert a chemical signal, i.e the binding of neurotransmitter, into an ion flux through the postsynaptic membrane [1] . They mediate inhibitory neurotransmission in the spinal cord and brainstem and have been associated with a range of neurological disorders including Hiperekplexia, temporal lobe epilepsy and chronic pain [2]. Functional studies by patch-clamp electrophysiology have shown that synaptic receptors accomplish their function by switching between (at least) three distinct conformational states: a resting state, where the ion pore is closed; an active state stabilized by agonists, where the pore is open; and one or more desensitized states, where the channel shuts with the agonist still bound [3, 4].

From a fundamental perspective, previous modelling attempts have been reported to decipher the gating mechanism on receptor homologs from bacteria or lower invertebrates [5, 6, 7]. Unfortunately, these models are not easily transferable to a human receptor as the GlyR. In addition, none of them include the receptor desensitization nor the existence of pre-active intermediates [8], so that the gating mechanisms described so far are incomplete.

In this project, we propose to combine All-Atom (AA) Molecular Dynamics (MD) simulations with Markov State Modelling (MSM) to study GlyR gating mechanism. An MSM is a stochastic model that describes the dynamics of complex molecular systems[9]. It uses information generated by MD simulations to provide access to the structural (conformations), thermodynamic (probabilities) and kinetic properties (transition rates) of the system with atomic resolution. With this strategy we aim at constructing the first complete model of the gating mechanism on a human synaptic receptor. By now, we have produced 26 μs of MD simulations of the GlyR in all known conformational states. We have determined an ensemble of features capable of clustering the structures in biologically relevant states, and we have constructed a preliminary model. Our model distinguishes four relevant states: resting, pre-active, active and desensitized. We are currently working on sampling the transition between the open and desensitized state to construct a partial MSM describing the desensitization mechanism along with the energetics involved.

References

[1] M. Cecchini and J.-P. P. Changeux, "The nicotinic acetylcholine receptor and its prokaryotic homologues: Structure, conformational transitions and allosteric modulation," Neuropharmacology, vol. 96, no. PB, pp. 137–149, 2015.

[2] J. W. Lynch, Y. Zhang, S. Talwar, and A. Estrada-Mondragon, "Chapter eight - glycine receptor drug discovery," in Ion Channels DownUnder (D. P. Geraghty and L. D. Rash, eds.), vol. 79 of Advances in Pharmacology, pp. 225–253, Academic Press, 2017.

[3] B. Katz and S. Thesleff, "A study of 'desensitization' produced by acetylcholine at the motor end-plate," The Journal of physiology, vol. 138, pp. 63–80, 09 1957.

[4] M. Gielen and P.-J. Corringer, "The dual-gate model for pentameric ligand-gated ion channels activation and desensitization," The Journal of Physiology, vol. 596, 02 2018.

[5] N. Martin, S. Malik, N. Calimet, J.-P. Changeux, and M. Cecchini, "Un-gating and allosteric modulation of a pentameric ligand-gated ion channel captured by molecular dynamics," PLOS Computational Biology, vol. 13, p. e1005784, 10 2017.

[6] B. Lev, S. Murail, F. Poitevin, B. A. Cromer, M. Baaden, M. Delarue, and T. W. Allen, "String method solution of the gating pathways for a pentameric ligand-gated ion channel," Proceedings of the National Academy of Sciences, vol. 114, no. 21, pp. E4158–E4167, 2017.

[7] C. Bergh, S. A. Heusser, R. Howard, and E. Lindahl, "Markov state models of proton- and pore-dependent activation in a pentameric ligand-gated ion channel," eLife, vol. 10, p. e68369, oct 2021.

[8] S. Shi, S. N. Lefebvre, L. Peverini, A. Cerdan, P. Milan Rodriguez, M. Gielen, J.-P. Changeux, M. Cecchini, and P.-J. Corringer, "Illumination of a progressive allosteric mechanism mediating the glycine receptor activation," Nat comm, 2022.

[9] B. Husic and V. Pande, "Markov state models: From an art to a science," Journal of the American Chemical Society, vol. 140, 01 2018.

DECIPHERING THE ACTIVATION MECHANISM OF THE AT1 RECEPTOR USING CLASSICAL AND STEERED DYNAMIC SIMULATIONS: INFLUENCE OF ENVIRONMENTAL FACTORS.

RYM BEN BOUBAKER [a*]
DANIEL HENRION [a]
MARIE CHABBERT [a]

[a] CNRS UMR 6015 – INSERM U108, MITOVASC, University of Angers, FRANCE
[*] Present address : CNRS UMR 5203 – INSERM U1191, Institut de Génomique Fonctionnelle, Montpellier, FRANCE

Angiotensin II type 1 receptor (AT1) belongs to the GPCR superfamily. This receptor is strongly involved in cardiovascular homeostasis and blood pressure. In the absence of its natural ligand, AT1 can be activated by mechanical forces (shear forces or membrane stretching), but the molecular mechanisms of activation are still unknown.

To answer these questions, we analyzed the activation mechanism of the AT1 receptor under mechanical stress (positive membrane tension to mimic membrane stretching), in different lipidic environments by molecular dynamics simulations (MD). Classical MD simulations indicated that the AT1 receptor submitted to a positive surface tension in an anionic environment could adopt a non-canonical conformation. For further understanding of the activation mechanism, we set up steered MD simulations (SMD). They revealed mechanistic steps of AT1 activation and identified the same non-canonical conformation as an intermediate state in the activation process. We investigated the interactions stabilizing the inactive and intermediate conformations.

This study suggested that any disruption of these interactions by external factors might initiate AT1 receptor activation in the absence of ligand. These results show the impact of environmental factors on GPCR responses.

ON THE MOLECULAR ORIGINS OF THERMOPHORESIS IN DILUTE BINARY MIXTURES

MARIO ARAUJO-ROCHA
ALEJANDRO DIAZ-MARQUEZ
GUILLAUME STIRNEMANN
Laboratoire de Biochimie Théorique, CNRS UPR 9080, IBPC, 13 rue Pierre et Marie Curie, 75005, Paris, France

The issue of molecular confinement plays a critical role in the investigation of the genesis of life. How can relevant reactions occur at efficient rates when there are no cell membranes to confine the reagents? Thermophoresis is the phenomenon of mass transfer induced by the presence of a temperature gradient. Recent research indicates that thermophoretic effects may have been crucial in the selective concentration of biologically significant reagents on early Earth[1].

Although this phenomenon has been known for almost 150 years, a full molecular understanding is still missing, mostly because the tools developed for processes at thermodynamic equilibrium are not necessarily adequate to describe out-of-equilibrium phenomena.

In our work, we studied dilute binary mixtures of Lennard-Jones liquids with thermophobic (accumulating in the cold section) and thermophilic (accumulating in the hot section) solutes. The system choice is due to the ease of modifying the solute's characteristics by changing its mass or Lennard-Jones parameters, effectively modulating isotopic and chemical effects. These are simulated by running nonequilibrium molecular dynamics simulations of a box containing a heat source and a heat sink.

While previously-suggested explanations were found to be not adequate, we have found a strong correlation between the sign and amplitude of the particle accumulation, and the ballistic motion of the solutes, with thermophilic solutes showing faster motion than the solvent in general. In contrast, thermophobic solutes display slower ballistic motions. Furthermore, we have attempted to explain these results by analyzing the solute and solvent exchange kinetics over the temperature gradient.

References

[1] T. Matreux, K. Le Vay, A. Schmid, P. Aikkila, L. Belohlavek, A. Çalışkanoğlu, E. Salibi, A. Kühnlein, C. Springsklee, B. Scheu, D. Dingwell, D. Braun, H. Mutschler, C. Mast, Nat. Chem., 13, 1038-1045 (2021)

MODELING THE NADPH OXIDASE (NOX) COMPLEX

AYKAC FAS BURCU[a], ERARD MARIE[b],
NÜSSE OLIVER[b], SACQUIN-MORA SOPHIE[a], TALY ANTOINE[a]

[a] Laboratoire de Biochimie Théorique UPR 9080 CNRS, Institut de Biologie Physico-Chimique, Paris, France
[b] Laboratoire de Chimie Physique, CNRS, Université Paris-Sud, Université Paris-Saclay, Orsay, France

The phagocyte NADPH oxidase is a multicomponent enzyme that becomes active after the assembly of its cytosolic proteins p40phox, p47phox, p67phox; the small GTPase Rac, and transmembrane proteins p22phox and NOX2 (gp91phox). Upon activation, the cell signaling cascade triggers phosphorylations of the subunits of the cytosolic heterotrimeric complex, resulting in its conformational transition into the active state. The formation of the active complex also involves reorganizing the intramolecular interactions within the cytosolic complex and forming new interactions with the membrane subunits and anionic phospholipids in the membrane [1, 2]. The C-term of NOX2, as well as p47phox and p67phox, contain structured domains separated by intrinsically disordered regions (IDRs), and their flexibility is crucial for the activity of NOX. This work aims to model the NOX complex embedded in a physiologically relevant membrane and identify the protein-protein/protein-membrane interactions that form the basis of the oxidase assembly and functioning, including the modeling of the IDRs and the phosphorylations of its subunits, based on the NOX cytosolic complex [3] and the membrane components [4, 5]. Therefore, an integrative structural biology approach is followed, combining deep learning-based modeling approaches for multimeric complexes [6-9], all-atom, and coarse-grained conventional molecular dynamics simulations, supported by constraints from our experimental partners. Thus, this combined approach will help us identify the key residues responsible for the enzyme activity and, possibly, potential therapeutic targets [10] and will contribute to the understanding of the NADPH oxidase machinery at the atomic level.

References

[1] TVermot, A., Petit-Härtlein, I., Smith, S. M., Fieschi, F. NADPH Oxidases (NOX): An Overview from Discovery, Molecular Mechanisms to Physiology and Pathology. Antioxidants, 2021, 10(6), 890.

[3] Ziegler, C. S., Bouchab, L., Tramier, M., Durand, D., Fieschi, F., Dupré-Crochet, S., Mérola, F., Nüße, O, Erard, M. Quantitative live-cell imaging and 3D modeling reveal critical functional features in the cytosolic complex of phagocyte NADPH oxidase. JBC, 2019, 294(11), 3824-3836.

[4] Noreng, S., Ota, N., Sun, Y., Ho, H., Johnson, M., Arthur, C. P., ... & Koerber, J. T. Structure of the core human NADPH oxidase NOX2. Nature Communications, 2022, 13(1), 6079.

[5] Liu, R., Song, K., Wu, J. X., Geng, X. P., Zheng, L., Gao, X., ... & Chen, L. Structure of human phagocyte NADPH oxidase in the resting state. Elife, 2022, 11, e83743.

[6] Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., ... & Hassabis, D. Highly accurate protein structure prediction with AlphaFold. Nature, 2021, 596(7873), 583-589.

[7] Evans, R., O'Neill, M., Pritzel, A., Antropova, N., Senior, A. W., Green, T., ... & Hassabis, D. Protein complex prediction with AlphaFold-Multimer. bioRxiv, 2021.

[8] Johansson-Åkhe, I., & Wallner, B. Improving peptide-protein docking with AlphaFold-Multimer using forced sampling. Frontiers in Bioinformatics, 2022, 2, 85.

[9] Gao, M., Nakajima An, D., Parks, J. M., & Skolnick, J. AF2Complex predicts direct physical interactions in multimeric proteins with deep learning. Nature communications, 2022, 13(1), 1744.

[10] Diebold, B. A., Smith, S. M., Li, Y., & Lambeth, J. D. NOX2 as a target for drug development: indications, possible complications, and progress. Antioxidants & redox signaling, 2015, 23(5), 375-405.

EVALUATION OF ACTIVE LEARNING STRATEGIES FOR IMPROVING QSAR MODELING: A REGRESSION-BASED BENCHMARKING STUDY

PABLO MAS[a,b]
MARC BIANCIOTTO[b]
RODOLPHE VUILLEUMIER[a]

[a] PASTEUR, Département de chimie, École Normale Supérieure, PSL University, Sorbonne Université, CNRS, 75005 Paris, France
[b] Molecular Design Sciences - Integrated Drug Discovery, Sanofi R&D, 94400 Vitry-sur-Seine, France

Machine learning has become an essential tool for drug discovery, as it enables efficient analysis and prediction of biological activities and properties of compounds, facilitating the discovery of new drug candidates. In particular, active learning is a valuable approach for quantitative structure-activity relationship (QSAR) modeling, as it can considerably reduce the cost and time needed to develop accurate predictive models, allowing them to be trained with fewer labeled data points by selecting the most informative samples. In this work, we present a benchmarking study of various active learning strategies for regression-based QSAR modeling. We compare uncertainty-based, diversity-based, and random active learning strategies. Different regression uncertainty estimation methods are tested using conformal and quantile prediction. The evaluation is carried out on absorption, distribution, metabolism, and excretion (ADME) datasets, including properties such as cell effective permeability, lipophilicity, half-life, or cytochromes inhibition and range from a few hundred molecules to tens of thousands.

Active learning strategies are compared based on their ability to select informative compounds from a large pool of molecules to train gradient boosted algorithms using metrics such as root mean squared error (RMSE) and mean absolute error (MAE) of prediction. The benchmarking framework is implemented in Python, using datasets from the Therapeutics Data Commons[1] (TDC) library. To vectorize molecules in a machine-readable format we have used the RDKit[2] library and have optimized a combination of Extended-Connectivity Fingerprints (ECFP), MACCS keys, and RDKit2D descriptors. The MAPIE[3] library is used in combination with different machine learning libraries to perform conformal predictions with gradient boosted regressors[4,5].

We find that, while the performance of active learning strategies is dataset-dependent, uncertainty-based methods generally perform better than diversity-based and random sampling as they allow predictive models to obtain better scores with a lower amount of training data. The findings of this study can have practical implications in the field of QSAR modeling by reducing the cost and time required for developing accurate predictive models.

References

[1] Huang, K., Fu, T., Gao, W., Zhao, Y., Roohani, Y. H., Leskovec, J., Coley, C. W., Xiao, C., Sun, J., & Zitnik, M. (2021). Therapeutics Data Commons : Machine Learning Datasets and Tasks for Drug Discovery and Development. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2102.09548

[2] G. Landrum, RDKIT: Cheminformatics and machine-learning software in C++ and Python. https://www.rdkit.org/

[3] Taquet, V., Blot, V., Morzadec, T., Lacombe, L., & Brunel, N. (2022). MAPIE: an open-source library for distribution-free uncertainty quantification. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2207.12274

[4] Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T. (2017). LightGBM: a highly efficient gradient boosting decision tree. Neural Information Processing Systems, 30, 3149-3157.

[5] Chen, T., & Guestrin, C. (2016). XGBoost. arXiv (Cornell University). https://doi.org/10.1145/2939672.2939785

Pablo Mas and Marc Bianciotto are Sanofi employees and may hold shares and/or stock options in the company. Rodolphe Vuilleumier has nothing to disclose.

INVESTIGATING DESATURATION/ HYDROXYLATION SPECIFICITY OF THE CPFAH12 BIFUNCTIONAL ENZYME USING MULTI-SCALE MODELLING

ALBA NIN-HILL[a], JÉRÉMY ESQUE[a], SELMA BENGAOUER[a]
JÉRÉMY LE REUN[a], CARME ROVIRA[b,c]
FLORENCE BORDES[a], ISABELLE ANDRÉ[a]

[a] Toulouse Biotechnology Institute (TBI), Université de Toulouse, F-31077 Toulouse Cedex 04

[b] Departament d'Inorgànica i Orgànica and Institut de Química Teòrica i Computacional (IQTC). Universitat de Barcelona. Barcelona. Spain

[c] Institució Catalana de Recerca i Estudis Avançats (ICREA). Barcelona. Spain

Desaturation is one of the most common fatty acid (FA) modifications and it has a crucial impact on membrane fluidity, among other physiological needs. The membrane fatty acid desaturases, found in all organisms, catalyze this desaturation reaction on acyl chains that are esterified on a broad range of lipids (phospholipids with different head groups, sphingolipids,...) with a well-defined regioselectivity (double bond positioning on the FA)[1]. These membrane enzymes have a diiron cluster in their active site and need dioxygen and an electron transfer chain to activate their catalytic activity. Some desaturases of the Δ12-FAD family (the enzymes that act on the bond between carbons 12 and 13) have evolved to perform modifications other than desaturation, such as hydroxylation[2]. Some of the hydroxylated products, like ricinoleic acid, present some interesting properties for industrial applications. In this work, we will study *Cp*FAH12 enzyme; the only characterized fungal Δ12-hydroxylase from *Claviceps purpurea*[3]. This enzyme has been characterized as a bifunctional hydroxylase/desaturase enzyme that shares 86% of amino acid identity with the common related *Cp*FAD12, a monofunctional Δ12-desaturase. The main objective of this work is to investigate the molecular determinants responsible for the hydroxylation/desaturation specificity by analyzing both enzymes. One of the main drawbacks studying membrane-bound desaturases is the small number of characterized three-dimensional structures. We thus modelled their structures using Alphafold2 and refined them using classical MD and QM/MM simulations. Overall, our results enabled to predict the 3D structures of *Cp*FAH12 and *Cp*FAD12, their substrate binding modes, and the possible productive binding mode of O2 in their active site. On these structural grounds, we are now investigating the reaction mechanisms of hydroxylation and desaturation in *Cp*FAH12 by means of QM/MM metadynamics simulations. This will enable to improve our understanding of the catalytic mechanisms and hopefully will help to identify key molecular determinants involved in the specificity of reaction of this bifunctional enzyme and more precisely of the unusual hydroxylation reaction. This work fits in the scope of developing efficient hydroxylases for the diversification of lipid structures of great interest as synthons in oleochemistry.

Acknowledgements: *Alba Nin-Hill received a grant from Universitat de Barcelona, modalitat Margarita Salas. This work was granted access to the HPC resources of the Computing mesocenter of Région Midi-Pyrénées (CALMIP, Toulouse, France).*

References

[1] Shanklin J, Cahoon E. B. Desaturation and related modifications of fatty acids. Annu. Rev. Plant Physiol., 1998, 611-641

[2] Cao, S., Zhou, X. R., Wood, C. C., Green, A. G., Singh, S. P., Liu, L., Liu, Q. A large and functionally diverse family of Fad2 genes in safflower (Carthamus tinctorius L.). BMC Plant Biology, 2013, 1-18

[3] Kajikawa M, et al. Production of ricinoleic acid-containing monoestolide triacylglycerides in an oleaginous diatom, Chaetoceros gracilis. Sci. Rep., 2016, 36809

KOPIS: DEEP LEARNING BASED DOMAIN RECOGNITION

RAGOUSANDIRANE RADJASANDIRANE[a], GABRIEL CRETIN[a], YANN VANDER MEERSCHE[a], TATIANA GALOCHKINA[a], JEAN-CHRISTOPHE GELLY[a]

[a] Université Paris Cité and Université des Antilles and Université de la Réunion, INSERM, BIGR, F-75014 Paris, France

Introduction: Identification of protein domains, i.e. independent and compact regions structurally, functionally and/or evolutionarily relevant in proteins, is one of the essential steps of structure analysis. This step is even more crucial now with the availability of large databases of predicted structures from sequence-based deep learning methods (Alphafold2 [1,2] and ESM metagenomic atlas [3]). While most of current methods use classical partitioning algorithms based on distances between atoms, new faster and performant deep learning-based methods have been proposed. However, they do not allow to evaluate domain quality structurally, neither in terms of independence, nor consider alternative domain partitions like our successful SWORD/SWORD2 [4, 5] methods. In this context, we propose a new deep learningbased method named KOPIS, which offers matching performance while being faster than SWORD/SWORD2.

Material & Methods: The proposed strategy consists of two steps:
First, identification of point separation between domains. It is based on the contact probability map and a convolutional network of the ConvNext [6] type with dilation and generates, for each position along the sequence, the probability that the position constitutes a junction between domains. Second, based on the junctions determined in step 1, propose candidate domains to assess by another ConvNext network. This allows the identification of several structural domains and several architectural organizations, as well as identifying poorly ordered or unordered regions. The data used comes from domain assignments by the SWORD/SWORD2 algorithm [4,5] of a set from the Alphafold database filtered to minimize redundancy [2]. To finalize the models, a fine-tuning phase is carried out on the consensus assignments between the CATH [7] and SCOP [8] structural domain databanks. The learning and performance evaluation is classically carried out according to a crossvalidation procedure and constitution of independent datasets between learning, validation and test sets.

Results: On the classical benchmarks of the literature, we obtain similar performances to other reference methods such as SWORD/SWORD2, PDP [9], DomainParser [10], DDOMAIN [11]. Moreover, like SWORD/SWORD2, our method is able to identify alternative delineations which is unique among all other competing approaches. Finally, KOPIS is 10x faster than SWORD/SWORD2 which allows its systematic use on large protein structure datasets such as ESM Metagenomic Atlas and its more than 617 million structure models [3,12].

References

[1] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. Kohl, A. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, Nature, 596, 583-589 (2021)

[2] K. Tunyasuvunakool, J. Adler, Z. Wu, T. Green, M. Zielinski, A. Žídek, A. Bridgland, A. Cowie, C. Meyer, A. Laydon, S. Velankar, G. Kleywegt, A. Bateman, R. Evans, A. Pritzel, M. Figurnov, O. Ronneberger, R. Bates, S. Kohl, A. Potapenko, A. Ballard, B. Romera-Paredes, S. Nikolov, R. Jain, E. Clancy, D. Reiman, S. Petersen, A. Senior, K. Kavukcuoglu, E. Birney, P. Kohli, J. Jumper, D. Hassabis, Nature, 596, 590-596 (2021)

[3] Z. Lin, H. Akin, R. Rao, B. Hie, Z. Zhu, W. Lu, N. Smetanin, R. Verkuil, O. Kabeli, Y. Shmueli, A. dos Santos Costa, M. Fazel-Zarandi, T. Sercu, S. Candido, A. Rives, Evolutionary-scale prediction of atomic level protein structure with a language model, (2022)

[4] G. Cretin, T. Galochkina, Y. Vander Meersche, A. de Brevern, G. Postic, J. Gelly, Nucleic Acids Research, 50, W732-W738 (2022)

[5] G. Postic, Y. Ghouzam, R. Chebrek, J. Gelly, Sci. Adv., 3, (2017)

[6] Zhuang Liu, Hanzi Mao, Chao-Yuan Wu, Christoph Feichtenhofer, Trevor Darrell, & Saining Xie. (2022). A ConvNet for the 2020s.

[7] I. Sillitoe, N. Bordin, N. Dawson, V. Waman, P. Ashford, H. Scholes, C. Pang, L. Woodridge, C. Rauer, N. Sen, M. Abbasian, S. Le Cornu, S. Lam, K. Berka, I. Varekova, R. Svobodova, J. Lees, C. Orengo, Nucleic Acids Research, 49, D266-D273 (2020)

[8] J. Chandonia, L. Guan, S. Lin, C. Yu, N. Fox, S. Brenner, Nucleic Acids Research, 50, D553-D559 (2021)

[9] N. Alexandrov, I. Shindyalov, Bioinformatics, 19, 429-430 (2003)

[10] J. Guo, Nucleic Acids Research, 31, 944-952 (2003)

[11] H. Zhou, B. Xue, Y. Zhou, Protein Sci., 16, 947-955 (2007)

[12] ESM Metagenomic Atlas (https://esmatlas.com/)

CLASSIFICATION OF PROTEIN-CARBOHYDRATE INTERFACES USING UNSUPERVISED MACHINE LEARNING

ARIA GHEERAERT[a], RAVY LEON FOUN LIN[a], THOMAS BAILLY[a], YANI REN[a], FRÉDÉRIC GUYON[a], TATIANA GALOCHKINA[a]

[a] Université Paris Cité and Université des Antilles and Université de la Réunion, INSERM, BIGR, F-75014 Paris, France

Background. Carbohydrates play a crucial role in the majority of biological processes. Carbohydrates can act as a ligand in biochemical pathways, but also are present on the surface of all the living cells and are therefore involved in cellular recognition processes as well as in protein-protein interactions in case of glycosylation of one of the partners. Experimental characterization of protein-carbohydrate (PC) remains a technical challenge due to variability and flexibility of carbohydrate molecules, and the structural data on PC interfaces is scarce and underrepresented. Moreover, carbohydrate binding sites' (BS) diversity is very peculiar. Indeed, some similar protein regions can bind to different carbohydrates, while other BS targeting specific carbohydrates are found in proteins with very different folds[1]. Objectives. The goal of the current study was to provide a generalized view on different types of PC interfaces through the extensive analysis of the available structural data.

Methods. We have extracted all the carbohydrate-containing entries available in PDB and performed their annotation by merging all the information available in the biological databases. We have then identified all the available BS (more than 100k) and performed a pairwise comparison of the most representative BS using a graph-based score evaluating local distance distortion between different atom types[2]. Finally, we have performed their classification using unsupervised machine learning (ML) methods by applying clustering for the obtained affinity score matrix.

Results. We have proposed a new hierarchical classification of carbohydrate BS by their chemical and structural similarity. We show that most carbohydrate BS can be assigned to just a few categories: some of them being protein- or ligand-specific and others being more general, e.g. including proteins with very low homology relations. All the protein and carbohydrate annotations are now assembled in a new database, providing the user with possibility to select PC complexes with similar interaction pattern as well as to evaluate the class of PC interactions for a new unassigned data.

Conclusions. Our result paves the way to the development of the carbohydrate BS computational prediction tools, which have the potential to significantly impact understanding of fundamental biological processes and provide new drug design strategies.

References

[1] Copoiu L, Torres PHM, Ascher DB, Blundell TL, Malhotra S. ProCarbDB: a database of carbohydrate-binding proteins. Nucleic Acids Res. 2020 Jan 8;48(D1):D368-D375.

[2] Rasolohery I, Moroy G, and Guyon F. PatchSearch: A Fast Computational Method for Off-Target Detection, Journal of Chemical Information and Modeling 2017 57 (4), 769-777

DEVELOPMENT OF LARGE SPECTRUM ANTIVIRAL DRUGS TARGETING RNA HELICASE A

L. MARTY
J.-M. PÉLOPONÈSE
L. BRIANT
M. BLAISE
L. CHALOIN
IRIM – CNRS – University of Montpellier

RNA helicases (RHA, also known as DEAD/H helicases) play an important role in unwinding double helix of DNA or RNA prior the transcription of cellular genes. In human cells, RHA interacts with RNA polymerase II, transcription factors and co-activators to modulate the transcription and the translation of selected genes. In parallel, many viruses have been shown to take benefit of this cellular function since RHA interacts with viral RNA and/or viral proteins allowing to favor their multiplication. Indeed, RHA has been identified as an important enzyme for the replication of RNA viruses such as HIV, HTLV-1, Dengue, Zika and SARS-CoV-2 viruses. Our project is devoted to understand the molecular mechanisms by which RHA is involved and required for viral replication and also to develop small molecules as antiviral drugs targeting this helicase in order to block the viral proliferation. We have identified by virtual screening heterocyclic compounds targeting the core domain of RHA (outside the ATP binding site) that are able to bind to RHA and to inhibit viral replication in cell-based assays. Several rounds of lead optimizations were carried out in order to improve the efficacy and the selectivity of these compounds and each series was further evaluated by molecular docking and molecular dynamics simulations to estimate their binding stability and key interactions. Correlations between *in silico* predictions and experimental studies allowed to determine the minimal structural motif required for the antiviral activity (structure-activity relationships) and crystallographic studies are in progress to go further in the development of more potent antiviral drugs.

QM/MM STUDY OF THE FORMATION OF CYCLOBUTANE PYRIMIDINE DIMERS INVOLVING TRIPLET THYMINE

LALEH ALLAHKARAM[a], NATACHA GILLET [a]
ELISE DUMONT [b,c]

[a] Laboratoire de Chimie, ENS de Lyon UMR 5182, 46 allée d'Italie, 69364 Lyon Cedex7, France
[b] Université Côte d'Azur, CNRS, Institut de Chimie de Nice, UMR 7272 – 06108 Nice, France
[c] Institute Universitaire de France, 5 rue Descartes, 75005 Paris, France

Exposure to sunlight radiations induced photodamage to biomolecules and triggered several well-characterized DNA photolesions despite the intrinsic photostability of DNA nucleobases. UV radiation absorption by DNA leads to genetic lesions that impact DNA transcription and replication. But the damage can be repaired by specifically designed proteins (photolyases, BER, and NER mechanisms) in living cells [1]. Among all DNA damage, thymine dimer formation is one of the predominant and most harmful types of photochemical damage [2]. It corresponds to a [2 + 2] photocycloaddition between two adjacent pyrimidine nucleobases [3].

In this project, we first investigate the stability of triplet thymine along a TTTT track [4] owing to QM/MM-MD simulations, where all four adjacent thymines are included in the QM subsystem. Thus, we could successfully analyze the energy alignment and spin localization of the lowest triplet state of thymine embedded in a B-DNA environment [4]. We then investigated the free energy along the reaction coordinate for the formation C6-C6 bond and C5-C5 between two adjacent pyrimidine bases (TT and CT). To do that, we compute the potential of mean force (PMF) by QM/MM using umbrella sampling. Our simulations account for embedding the B-DNA helix, whereas other studies in the literature have usually considered a model system featuring two π-stacked thymines [5]. Finally, we explored the extra-helical flipping of the CPD damage using biased methods and classical MDs. Altogether, our results provide an extensive description of the generation and behavior of CPD in double-strand DNA.

References

[1] Masson F, Laino T, Tavernelli I, Rothlisberger U and Hutter J. Computational Study of Thymine Dimer Radical Anion Splitting in the Self-Repair Process of Duplex DNA, 2008, 3443-3450.

[2] Mouret S, Baudouin C, Charveron M, Favier A, Cadet J and. Douki T. Cyclobutane pyrimidine dimers are predominant DNA lesions in whole human skin exposed to UVA radiation, 2006, 13765-13770.

[3] Rössle H, Friedrichs J and Fran I. The Formation of DNA Photodamage: The Role of Excition Localization, 2010, 2011-2015.

[4] Allahkaram L, Monari A and Dumont E. The Behavior of Triplet Thymine in a Model B-DNA Strand. Energetics and Spin Density Localization Revealed by ab initio Molecular Dynamics Simulations 2021, 633-639.

[5] Sanjuán D, González G, Ramirez I, Andres L and Merchan M. Molecular Basis of DNA Photodimerization: Intrinsic Production of Cyclobutane Cytosine Dimers, 2008, 10768-10779.

HUMAN DE NOVO DNA-METHYLTRANSFERASE SELECTIVITY IS REGULATED BY A BASESPECIFIC HYDROGEN BONDING NETWORK

AYŞE BERÇIN BARLAS[a,b], EZGI KARACA[a,b]

[a] Izmir Biomedicine and Genome Center, Computational Structural Biology Research Group, Izmir, Turkey

[b] Dokuz Eylul University, Izmir International Biomedicine and Genome Institute, Izmir, Turkey

In mammals, de novo DNA methylation is essential to embryonic development, reprogramming, and gene regulation [1,2]. The de novo DNA methylation is directed by DNMT3A and DNMT3B enzymes and mainly exerted on CpG islands [3,4]. Over their enzymatic domains, DNMT3A/B proteins share over 90% of sequence similarity. Even so, DNMT3A predominantly methylates the first cytosine in the CGC and CGT motifs, while DNMT3B prefers the CGG and CGA sequences [5,6]. To elucidate the mechanistic basis of these selective methylation profiles, we performed extensive molecular dynamics simulations of DNMT3A/B enzymes bound to all possible CGX[C/G/T/A] motifs. For this, we have used the crystal structure of DNMT3A-CGC [7] complex as template, and modified the DNA motif to CGG, CGT and CGA, then replaced the DNA in the DNMT3B-CGG [8] structure with these motifs. For each complex, we performed 2 μs long MD simulations using GROMACS 2020.4 software [9] under AMBER parmbsc1 force-field [10]. Afterwards, we calculated the differential base-specific hydrogen bonding profiles of the systems. As an outcome, we observed that DNMT3A/B sequence selectivity is regulated by an arginine-lysine substitution: the long and flexible side chain of DNMT3A-Arg836 stably reads guanine in the complementary strand of the CGC motif, where the less flexible DNMT3B-Lys777 forms selective hydrogen bonds with two consecutive guanines in the target strand in CGG motif. Interestingly, DNMT3B-Lys777 and DNMT3B-Asn779 assist the sequence selectivity in a cooperative manner, while in DNMT3A, Arg836 alone performs the same task. We also observed that when bound to their cognate sequences, the DNMT3A-DNA hydrogen bonding profile is significantly altered by single nucleotide substitution in other CpG motifs, with DNMT3B being less affected by it. We correlated this with the higher CpG specificity of DNMT3A than DNMT3B. Taken together, these findings not only provide the missing molecular links in the DNMT3A/B mechanism of action, but also reveal that the findings obtained by experimental analyzes are provided by molecular dynamics simulations.

References

[1] Dor, Y., & Cedar, H. Principles of DNA methylation and their implications for biology and medicine. The Lancet, 392(10149), 777–786 (2018).

[2] Skvortsova, K., Stirzaker, C., & Taberlay, P. The DNA methylation landscape in cancer. Essays in Biochemistry, 63(6), 797–811 (2019).

[3] Schmitz, R. J., Lewis, Z. A., & Goll, M. G. DNA Methylation: Shared and Divergent Features across Eukaryotes. Trends in Genetics, 35(11), 818–827 (2019).

[4] Vukic, M., & Daxinger, L. DNA methylation in disease: Immunodeficiency, Centromeric instability, Facial anomalies syndrome. Essays in Biochemistry, 63(6), 773–783 (2019).

[5] Mao, S. Q., Cuesta, S. M., Tannahill, D., & Balasubramanian, S. Genome-wide DNA Methylation Signatures Are Determined by DNMT3A/B Sequence Preferences. Biochemistry, 59(27), 2541–2550 (2020).

[6] Mallona, I., Ilie, I. M., Karemaker, I. D., Butz, S., Manzo, M., Caflisch, A., & Baubec, T. Flanking sequence preference modulates de novo DNA methylation in the mouse genome. Nucleic Acids Research, 49(1), 145–157 (2021).

[7] Zhang, Z. M., Lu, R., Wang, P., Yu, Y., Chen, D., Gao, L., Liu, S., Ji, D., Rothbart, S. B., Wang, Y., Wang, G. G., & Song, J. Structural basis for DNMT3A-mediated de novo DNA methylation. Nature, 554(7692), 387–391 (2018).

[8] Lin, C. C., Chen, Y. P., Yang, W. Z., Shen, J. C. K., & Yuan, H. S. Structural insights into CpG-specific DNA methylation by human DNA methyltransferase 3B. Nucleic Acids Research, 48(7), 3949–3961 (2020).

[9] Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E., & Berendsen, H. J. GROMACS: fast, flexible, and free. Journal of computational chemistry, 26(16), 1701-1718 (2005).

[10] Ivani, I., Dans, P. D., Noy, A., Pérez, A., Faustino, I., Hospital, A., ... & Portella, G. Parmbsc1: a refined force field for DNA simulations. Nature methods, 13(1), 55 (2016).

COARSE-GRAINED MOLECULAR DYNAMICS OF IONIC TRANSPORT AND DNA TRANSLOCATION THROUGH PROTEIN NANOPORES

NATHALIE BASDEVANT[a], CAGLA OKYAY[a], JÉRÔME MATHÉ[a]

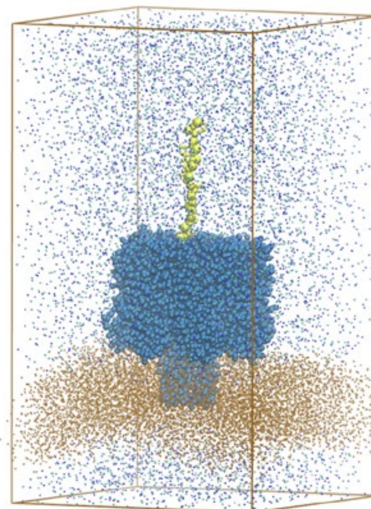
[a] Laboratoire Analyse et Modélisation pour la Biologie et l'Environnement – LAMBE – CNRS-Université d'Evry Paris-Saclay, 91025 Évry-Courcouronnes

Nanopore experiments are a popular technique which consists in applying an electric potential difference to guide a charged biopolymer through an artificial or biological nanopore. The pore is inserted in a solid or lipid membrane, surrounded by an ionic solution. When a macromolecule passes through the pore, it partially blocks it and induces a decrease in the ionic current, according to the nature of the molecule and the pore properties. Due to its commercial availability, the alpha-hemolysin toxin channel is a membrane protein widely used for these experiments. This protein nanopore enables the passage of a single-stranded DNA and has been the subject of several DNA translocation and unzipping experiments in our lab[1]. However, the molecular details are not experimentally accessible. Molecular dynamics simulations are therefore a valuable tool to understand the physical processes involved in these experiments.

Coarse-grained (CG) models are an excellent alternative to classical all-atom models because they enable longer and faster simulations for large systems, closer to the characteristic experimental times. In collaboration with experimentalists of our lab, we first performed coarse-grained molecular dynamics simulations of the ionic transport through alpha-hemolysin, inserted into a lipid bilayer surrounded by solvent and ions, using the MARTINI coarse-grained force field and polarizable water[2]. Our system is reduced to 90,000 coarse grains, instead of around 400,000 atoms, in the presence of several different electric fields to mimic the electric potential difference.

We were able to observe several specific features of this pore, current asymmetry, and anion selectivity, in agreement with previous studies and experiments, and identified the charged amino-acids responsible for these current behaviors using ionic density maps[3].

Moreover, we are now studying the translocation of single-stranded DNA molecules through alpha-hemolysin using steered CG molecular dynamics in order to elucidate the mechanisms involved in DNA translocation experiments. We performed simulations of several systems (~500,000 atoms,140,000 CG beads) for at least 1 µs with ssDNA of different lengths and direction at the nanopore entry. Our preliminary results, such as the distribution of translocation times per base are qualitatively consistent with experiments and are thus very promising.



References

[1] Mathé J., Visram H., Viasnoff V., Rabin Y. & Meller A. Nanopore unzipping of individual DNA hairpin molecules. Biophys. J., 87: 3205- 3212 (2004).

[2] Basdevant N., Dessaux D. & Ramìrez R. Ionic transport through a protein nanopore: a Coarse-Grained Molecular Dynamics Study. Scientific Reports, 9: 15740 (2019).

[3] Dessaux D., Mathé J., Ramìrez R & Basdevant N. Current rectification and ionic selectivity of alpha-hemolysin studied by Coarse-Grained Molecular Dynamics. J. Phys. Chem. B, 126, 4189-4199 (2022).

SUBSTRATE SELECTIVITY OF SULFOTRANSFERASE ISOENZYMES, RESULTS BASED ON MOLECULAR DYNAMICS AND VIRTUAL SCREENING

**DÁNIEL TOTH[a,b] , BÁLINT DUDAS[a,c] , DAVID PERAHIA[c]
ERIKA BALOG[b], MARIA A. MITEVA[a]**

[a] Inserm U1268 MCTR, CiTCoM UMR 8038 CNRS - Université Paris Cité, France
[b] Department of Biophysics and Radiation Biology, Semmelweis University, Hungary
[c] Laboratoire de biologie et pharmacologie appliquee, Ecole Normale Superieure Paris-Saclay, France

Sulfotransferase enzymes (SULTs) are a family of cytosolic globular proteins in the chain of metabolism. By catalysing a sulfate transfer from their co-factor, 3'-Phosphoadenosine 5'-Phosphosulfate (PAPS), they help to eliminate a large variety of small molecules like drugs, hormones and neurotransmitters. Even though the tertiary structure across the family is very similar, their substrates vary considerably in size and complexity.[1] The aim of our project is to better understand the reasons of selectivity between different SULTs, by comparing the broad targeting hepatic detoxifier 1A1, and the ileum located, dopamine selective 1A3. We utilize multiple in silico methods. We employ molecular dynamics (MD)[2] simulations and the recently developed approach of MD with excited Normal Modes (MDeNM)[3] to elucidate molecular mechanisms guiding the recognition of diverse ligands by SULT1A1. We used SULT structures from MD and MDeNM clustered into ensembles[4] and combined them with categorised substrates and inhibitors, performing Virtual Screening.[5]

Based on our results, we identified the most important amino acid differences, that are responsible to change the protein dynamics and binding mechanisms implicated in the selectivity of these isoenzymes. These results show the selectivity is likely to be governed by certain amino acid sidechains in 1A3 by opening the binding pocket to an unfavourable conformation for the most common ligands of 1A1, thus acting as efficient selectors. These results can be helpful in the future to develop an algorithm for machine learning, that could differentiate and even predict new substrates, thus helping in the development of ADME-Tox profiling of novel drug candidates and xenobiotics.

References

[1] Martiny V Y, Carbonell P, Lagorce D, Villoutreix B O, Moroy G, & Miteva M A. In silico mechanistic profiling to probe small molecule binding to sulfotransferases, PLoS One, 2013, 8(9) e73587.

[2] Brooks B R, Brooks C L, Mackerell A D, Nilsson L, Petrella R, J, Roux B, & Karplus M. CHARMM: The Biomolecular Simulation Program. Journal of Computational Chemistry, 2009, 30(10) 1545-1614.

[3] Costa M G S, Batista P R, Bisch P M, & Perahia D. Exploring Free Energy Landscapes of Large Conformational Changes: Molecular Dynamics with Excited Normal Modes. Journal of Chemical Theory and Computation, 2015, 11(6) 2755-2767.

[4] Humphrey W, Dalke A, & Schulten K. VMD: visual molecular dynamics. J. Mol. Graph., 1996, 14(1) 33-38 27-38.

[5] Trott O, & Olson A J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading, J. Comput. Chem., 2010 31(2) 455-461.

CHARMM FORCE FIELD PARAMETERIZATION PROTOCOL FOR THE PESTICIDE FLUPYRADIFURONE (SIVANTO™): AN IN-DEPTH DIVE IN THE OPTIMIZATION MAZE!

ZAKARIA BOUCHOUIREB[a], JEAN-YVES LE QUESTEL[a]

[a] Nantes Université, CNRS, CEISAM UMR 6230, F-44000, Nantes, France

Keywords: Pesticides • Flupyradifurone • nAChRs • Parametrization • CHARMM FF

According to the projections given by the **World Population Prospect 2022**, the world population should reach 8 billion on November 2022 [1]. This exponential growth creates an increasing necessity in food supplies. Thus, an important aspect of the modern sustainable agriculture is the design and development of potent, effective and safe pesticides to alleviate the threat of crops being damaged by pests. The latter should be highly selective of insects over mammals and pollinators. One way of tackling this selectivity issues is to target the nicotinc acetylcholine receptors (nAChRs) of insects. This strategy has shown a growing interest for many years and proved to be effective in particular with the insecticides of the neonicotinoids family that were developed as main solution to regulate insect pests not only in for crops but also in the field of animal health [2].

Flupyradifurone (FLU) also known by its trademark name **Sivanto** is the last member of nAChR competitive modulators designed by *Bayer CropScience* [3,4] and the most recent member of the neonicotinoid class of insecticides to reach market with an innovative pharmacophore including a butenolide moiety. Mainly claimed as harmless for non-targeted organisms such as pollinators (*e.g.* honey bees) and mammals, concerns about FLU resistance and toxicity for mammalian and the environment have been raised. However, these effects and the corresponding mechanisms are still debated and poorly described. Indeed, if the mechanism of action of FLU appears admitted, the information related to its interactions at the atomic level at the targeted insect nAChRs is limited.

Molecular modeling methodologies (*e.g.* molecular docking, molecular dynamics (MD)) are imperative tools in the absence of experimental information. However the question of results reliability is strongly correlated to force field parameters quality. Indeed, if force field parameters for biopolymers are readily available and widely used in MD simulations, this not the always the case for small ligands. Furthermore, knowing the bias of such fields toward pharmaceutical applications, the availability of force field parameters is even scarcer for pesticides related compounds like FLU.

In the present talk, we will discuss the workflow used for the development and optimization of **CGenFF force field** compatible parameters for FLU in order to be used in molecular dynamic and/or molecular docking. A set of initial parameters were used as a starting point and were derived from CHARMM General Force Field database (CGenFF) using CGenFF program [5, 6]. Then the non-bonded and bonded terms were optimized using Force Field Tool Kit (FFTK) software to reproduce MP2/6-31G* quantum calculations accuracy. The relevance of the obtained parameters was validated by comparing physical properties such as IR-spectra, Normal Modes Analysis (NMA) and water-octanol partition coefficient ($logK_{ow}$) to their experimental values. MD simulations of FLU/AChBP complexes were carried out to evaluate the ability of the optimized parameters in reproducing recently observed crystallographic trends.

References

[1] United Nations Department of Economic and Social Affairs, Population Division (2022). World Population Prospects 2022: Summary of Results. UN DESA/POP/2022/TR/NO.3.

[2] Jeschke Peter et al., Journal of Agricultural and Food Chemistry, 2011, 121(7):2897–2908

[3] Nauen Ralf et al., Pest Management Science, 2014, 71(6):850–862

[4] Jeschke Peter et al., Pesticide Biochemistry and Physiology, 2015, 121:31–38

[5] Vanommeslaeghe Kenno et al., Journal of Computational Chemistry, 2009, 31(4):671–690

[6] Vanommeslaeghe Kenno et al., Journal of Chemical Information and Modeling, 2012, 52(12):3144–3154

MOLECULAR DYNAMICS STUDIES OF PLATELET MEMBRANES BASED ON LIPIDOMICS

FLORENTIN ALLEMAND[a], SEMEN YESYLEVSKYY[b,c,d] JENNIFER LAGOUTTE-RENOSI[a,e], SIAMAK DAVANI[a,e] CHRISTOPHE RAMSEYER[b]

[a] EA 3920 Université de Franche-Comté, 25000 Besancon, France

[b] Laboratoire Chrono Environnement UMR CNRS 6249, Université de Bourgogne Franche-Comté, 16 route de Gray, 25000 Besançon, France.

[c] Department of Physics of Biological Systems, Institute of Physics of the National Academy of Sciences of Ukraine, Prospect Nauky 46, 03028 Kyiv, Ukraine.

[d] Institute of Organic Chemistry and Biochemistry, Czech Academy of Sciences, CZ-166 10 Prague 6, Czech Republic.

[e] Service de Pharmacologie Clinique et Toxicologie-CHU de Besançon, 25000 Besancon, France

The plasma membrane of platelets plays a major role in platelet activation, particularly through the P2Y receptors, a family of purinergic G-proteins. In addition, platelet plasma membrane is enriched in arachidonic acid compared to other cells. This unsaturated fatty acid is also necessary for platelet activation, as a precursor of thromboxane A2. But its role in the plasma membrane is not known. Our aim is therefore (i) to reveal the precise content and organization of lipids inside the plasma membrane of platelets and (ii) to see how P2Y12 receptors, the target of antiplatelet drugs, embedded accommodates such complex lipid environment.

For this purpose, coarse-grained molecular dynamics was used to study the dynamics of membrane lipids and P2Y12 receptors. This work is done according to lipidomic study conducted by our group[1]. Three models were built: a single bilayer with 10 lipid species representative of the lipidomic results (M1), a bilayer with 9 receptors P2Y12 embedded (M2) and a control model without acid arachidonic (M3). 100 µs of each model were simulated with the GROMACS 2018 suite of programs and MARTINI forcefields.

Dynamic lipid microdomains containing sphingomyelin and cholesterol enriched clusters were observed in M1, M2 and M3 models. More unexpectedly, phospholipid containing arachidonic acid-rich clusters were also identified in both M1 and M2. These enrichments are particularly found near P2Y12 receptors in the M2 model (between +10% and +125% depending on the species, within 5 Å of proteins). This is not observed for the same species where arachidonic acid has been substituted by oleic acid in the M3 model. Moreover, after 50 µs of simulation, there are two dimers and one trimer in M2 model while there is only one trimer in M3 model. All together, these results show that arachidonic acid promotes the oligomerization of P2Y12 receptors, which is apparently necessary for their functioning.[2]

References

[1] Lagoutte-Renosi J, Allemand F, Ramseyer C, et al. Influence of Antiplatelet Agents on the Lipid Composition of Platelet Plasma Membrane: A Lipidomics Approach with Ticagrelor and Its Active Metabolite, Int J Mol Sci, 22, 1432.

[2] Mundell SJ, Rabbolini D, Gabrielli S, et al. Receptor homodimerization plays a critical role in a novel dominant negative P2RY12 variant identified in a family with severe bleeding, J Thromb Haemost 2018, 16: 44–53.

COMPUTATIONAL ENZYME DESIGN OF COFACTOR SPECIFICITY 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.

Natural metabolic pathways have been promoted as a route to green synthesis of biofuels and biochemicals. However, these processes can be difficult to engineer into chassis microorganisms, and, when successful, are often hampered by the limitations imposed by complex cellular metabolism such as toxicity of products and intermediates, slow growth rates, and maintaining cell viability. One potential solution to these problems is to perform the reactions *in vitro* with partially purified enzymes or cell lysates.

Redox cofactor utilization is currently one of the major barriers to the realization of efficient and cost competitive cell-free biocatalysis, especially where multiple redox steps are concerned. The design of versatile, cofactor balanced modules for canonical metabolic pathways, such as glycolysis, is one route to overcoming such barriers.

In this regard, the non-phosphorylating glyceraldehyde-3-phosphate dehydrogenase (GapN) can be used as a shortcut into the glycolysis pathway by catalyzing the transformation of glyceraldehyde-3-phosphate into 3-phosphoglycerate. It is a crucial enzyme that can be involved in the regulation of ATP concentration in cell-free biocatalysis processes. However, GapN is strictly dependent on the NADP+ cofactor[1], which prevents its use for NAD+ cofactor-dependent pathways or those based on the use of more stable and less expensive biomimetic cofactors.

Therefore, we set up a computer-aided design framework to engineer GapN for enabling a NADH linked efficient cell-free glycolytic pathway with a net zero ATP usage. This rational design approach combines molecular dynamics simulations with Artificial Intelligence-based multistate computational design methods[2-3] that allowed us to consider different conformational states encountered along the GapN enzyme catalytic cycle. In particular, the cofactor flip, characteristic of this enzyme family and occurring before product hydrolysis[4], was taken into account to redesign the cofactor binding pocket for NAD+ utilization.

While GapN exhibits only tiny trace activity with NAD+, an enhancement of more than 10,000-fold in this activity was achieved, corresponding to a recovery of more than 72 % of the activity of the wild-type enzyme on NADP+. The development of this engineered GapN variant with comparable kinetics to the WT enzyme on its native cofactor represents an important step towards establishing an efficient, robust, and versatile cell-free glycolysis module for biochemical syntheses that could have broad applications in the field of green chemistry. Moreover, the computer-aided engineering approach we set up here may also provide a template for other researchers wishing to conduct similar engineering campaigns.

References

[1] Boyd D A, Cvitkovitch D G, Hamilton I R. J Bacteriol, 1995, 177(10), 2622–2627.

[2] Ruffini M, Vucinic J, de Givry S, Katsirelos G, Barbe S, Schiex T. Algorithms, 2021, 14(6), 168.

[3] Vucinic J, Simoncini D, Ruffini M, Barbe S, Schiex T. Bioinformatics, 2020, 36(1), 122-130.

[4] Rahuel-Clermont S, Bchini R, Barbe S, Boutserint S, André I, Talfournier F. ACS Catalysis, 2019, 9(2), 1337–1346.

CHARACTERIZATION AND FUNCTIONAL COMPREHENSION OF A PROTEIN ASSEMBLY: THE UBIQUINONE METABOLON FROM ESCHERICHIA COLI

ROMAIN LAUNAY[a], SOPHIE-CAROLE CHOBERT[b], ELIN TEPPA[a] CARLA MARTINS[a], SOPHIE ABBY[b], FABIEN PIERREL[b] ISABELLE ANDRÉ[a] *, JÉRÉMY ESQUE[a] *

[a] Toulouse Biotechnology Institute, TBI, Université de Toulouse, CNRS, INRAE, INSA, Toulouse, France. 135, avenue de Rangueil, F-31077 Toulouse Cedex 04, France

[b] TIMC, Université Grenoble Alpes, CNRS, CHU Grenoble Alpes, Grenoble INP, Grenoble, France

Protein-protein interactions and supramolecular complexes are essential for the functioning of living cells. They play a huge role in a number of biological functions. In this study, we were interested in a cytosolic complex, called Ubi metabolon[1] from *Escherichia coli.* This complex is involved in the biosynthesis of ubiquinone, a redox-active prenol localized in biological membranes[2] of eukaryotes and proteobacteria.

The Ubi metabolon of *E. coli* is composed of seven subunits including enzymes (methyl-transferases UbiG and UbiE, hydroxylases UbiI, UbiH and UbiF) and structural proteins (UbiJ and UbiK), with an unknown stoichiometry. Determination and characterization of the structure and the interactions of this supramolecular complex are highly challenging. Our goal was thus to provide alternative multi-scale computational approaches to predict the most probable structural model(s) of the different protein partners composing the metabolon and to investigate their interactions, trying to propose in fine a stable and functional molecular assembly of the Ubi metabolon. To this aim, an integrative approach has been considered, combining sequence information and structure prediction together with biochemical and biophysical data.

In view of this context, AlphaFold2[3] was central in our strategy, given its accuracy for both monomeric and multimeric predictions, by integrating evolutionary information. Our approach relied thus on the use of AlphaFold2 to predict first the molecular bricks corresponding to the different protein partners, and then to predict oligomer subunits, which were further expanded using a combination of both AlphaFold2 prediction and structural superposition. To elucidate the stoichiometry of the complex, AlphaFold2 was used to discriminate some potential protein-protein interactions, reducing the number of possibilities.

To evaluate the biological relevance of the generated molecular models, we completed this work by investigating interactions of some of the oligomer subunits with the E. coli inner membrane, and the compatibility of the system with ubiquinone precursors substrate channeling. To do so, several computational approaches were used, such as molecular docking, multiscale molecular dynamics approaches, using Coarse-grained (Martini3[4]) and all-atom simulations (CHARMM36[5]) or umbrella sampling approaches. These models are compared when possible to available experimental data, such as Cryo-EM which is under progress.

This approach has been exemplified to model the UbiJ-UbiK$_2$ complex, an essential oligomer subunit composing the Ubi metabolon[6]. This led to the prediction of heterotrimer structure of UbiJ-UbiK$_2$, as well as a possible binding mode between the heterotrimer and the membrane. Finally, this enabled to propose a pathway release of ubiquinone from Ubi the metabolon into the membrane.

This work is part of the PhD thesis of Romain Launay funded by French Ministry of Education, Research and Innovation (MESRI). This work was also partially funded by the French National Research Agency (ANR Project Deepen, ANR-19-CE45-0013-02). This work was granted access to the HPC resources of the Computing mesocenter of Région Midi-Pyrénées (CALMIP, Toulouse, France).

References

[1] M. Hajj Chehade et al. A soluble metabolon synthesizes the isoprenoid lipid ubiquinone. Cell Chemical Biology, (26): 482-492, 2019.

[2] J.A Stefely and D. J. Pagliarini. Biochemistry of Mitochondrial Coenzyme Q Biosynthesis. Trends in Biochemical Sciences, (42): 824-843, 2017

[3] J. Jumper et al. Highly accurate protein structure prediction with AlphaFold. Nature (596): 583-589, 2021.

[4] P.C.T Souza et al. Martini3: a general purpose force field for coarse-grained molecular dynamics. Nat Methods (18): 382-388, 2021.

[5] J. Huang and A.D. MacKerell Jr. CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. Journal of Computational Chemistry (34): 2135-2145, 2013.

[6] R. Launay et al. Towards Molecular Understanding of the Functional Role of UbiJ-UbiK2 Complex in Ubiquinone Biosynthesis by Multiscale Molecular Modelling Studies. Int. J. Mol. Sci. 2022, 23, 10323.

POSTER SESSION II

Tue. 16/05
18h-20h

Poster **n°1** to **n°60**

EXPLICIT MODELING OF BINDING PARTNERS TO INVESTIGATE THE INTRINSIC DYNAMICS AND MODULATION OF PROTEIN OLIGOMERIZATION

SANDHYA P. TIWARI[a]

[a] Institute of Protein Research, Osaka University, 3-2 Yamadaoka, Suita-Shi, Osaka 565-0871, Japan

Many proteins form oligomers under physiological conditions, which can confer thermal stability, greater complexity in structural and functional activity. Understanding the mechanisms of assembly provide insight into protein evolution, providing a framework for how protein structures adapt to gain new function. Previously, due to computational limits, intrinsic dynamics, or vibrational signatures, have typically been modelled implicitly by considering the conformation of a participating protein subunit in isolation. Generally, it is assumed the isolated subunit is in a conformation which captures the implicit effect of the binding partner on the intrinsic dynamics, suggesting the influence of a partnering subunit is already integrated. However, this description lacks detailed information on the influence of critical contacts at the protein-protein interface. Since the binding of many proteins to their protein partners is tightly regulated via control of their relative intrinsic dynamics, investigation of the intrinsic dynamics of proteins is necessary for the comprehensive understanding of function. In this study, we examine the case of a protein family, pyrimidine synthesis attenuator PyrR[1], to understand the effect of the binding partners in the stability of the tetramer vs. the dimer, and to uncover signals that link to their modulation via allostery. By partitioning the covariance matrices from elastic network models to obtain normal modes[2], we found that explicitly modelling the partnering subunits revealed the influence of perturbations that extend from the tetrameric interface, that is not captured by modelling the subunits in isolation.

References

[1] Perica, T., et al. Evolution of oligomeric state through allosteric pathways that mimic ligand binding. Science, 346, 6216 (2014). https://doi.org/ 10.1126/science.1254346

[2] Dasgupta, B., Tiwari, S.P. Explicit versus implicit consideration of binding partners in protein–protein complex to elucidate intrinsic dynamics. Biophys Rev 14, 1379–1392 (2022). https://doi.org/10.1007/s12551-022-01026-5

MOLECULAR MODELLING, HOMO-OLIGOMERISATION AND MEMBRANE INTERACTIONS OF HEPATITIS E VIRUS pORF1 REPLICATION POLYPROTEIN

THIBAULT TUBIANA[a], SONIA FIEULAINE[a], STÉPHANE BRESSANELLI[a]

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

Hepatitis E virus (HEV) is one of the main causative agents of acute hepatitis and jaundice, affecting both developed and developing countries. At least 20 million people are infected each year worldwide. HEV usually causes a minor and asymptomatic infection that resolves on its own but around 70,000 people die per year after HEV infection. HEV is particularly problematic in immunocompromised patients such as tranplant recipients, where infection progresses in most cases to a chronic and highly liver-damaging disease, for which antivirals are needed. However, despite its severity, its prevalence and its impact in health systems, HEV is poorly characterized and many questions remain unanswered about its functioning, especially at the structural level.

In recent years, the subject of structural biology has witnessed major revolutions. The most recent of these is the capacity of the software AlphaFold to properly predict protein 3D structures from their sequences alone. We used AlphaFold modeling to study I) the full-length pORF1 and ii) oligomers of a pORF1 subdomain (MetY)[1]. It turns out that MetY is highly similar to nsP1 of Chikungunya virus, for which the structure was published of a dodecamer putatively topping the membrane spherule harbouring the replication complex [2]. We modelled a similar MetY dodecamer and used coarse-grained[3] molecular dynamic simulations (MD) to investigate its interaction with membranes. To mimic a membrane spherule system, we used a torus (or donut-like) membrane system (generated with BUMPy[4] and TS2CG[5]) to reprensent the membrane curvature at a spherule exit.

We will discuss the conclusions we can draw from these models, specifically I) the breakdown of pORF1 into a limited number of domains and possible alternate conformations, ii) the possible organization, oligomerization, and membrane association of the pORF1 MetY.

References

[1] Fieulaine S, Tubiana T, Bressanelli S. De novo modelling of HEV replication polyprotein: Five-domain breakdown and involvement of flexibility in functional regulation. Virology. 2023;578:128-140. doi:10.1016/j.virol.2022.12.002

[2] Jones R, Bragagnolo G, Arranz R, Reguera J. Capping pores of alphavirus nsP1 gate membranous viral replication factories. Nature. 2021;589(7843):615-619. doi:10.1038/s41586-020-3036-8

[3] Monticelli L, Kandasamy SK, Periole X, Larson RG, Tieleman DP, Marrink SJ. The MARTINI Coarse-Grained Force Field: Extension to Proteins. J Chem Theory Comput. 2008;4(5):819-834.

[4] Boyd KJ, May ER. BUMPy: A Model-Independent Tool for Constructing Lipid Bilayers of Varying Curvature and Composition. J Chem Theory Comput. 2018;14(12):6642-6652. doi:10.1021/acs.jctc.8b00765

[5] Pezeshkian W, König M, Wassenaar TA, Marrink SJ. Backmapping triangulated surfaces to coarse-grained membrane models. Nat Commun. 2020;11(1):2296. doi:10.1038/s41467-020-16094-y

COMPUTATIONAL MODELLING FOR AN ANTICANCER THERAPY ENZYME METHIONINE-γ-LYASE

XINGYU CHEN[a], PIERRE BRIOZZO[b]
DAVID MACHOVER[c], THOMAS SIMONSON[a]

[a] Laboratoire de Biologie Structurale de la Cellule (CNRS UMR7654), Ecole Polytechnique, Palaiseau, France
[b] Institut Jean-Pierre Bourgin, INRAE-AgroParisTech, University Paris-Saclay, Paris, France,
[c] INSERM U935-UA09, University Paris-Saclay, Hôpital Paul-Brousse, Paris, France

Cancer cells require large amounts of methionine and cannot survive methionine deprivation, whereas normal cells are more resistant. Methionine-γ-lyase (MGL), a pyridoxal-5'-phosphate (PLP)-dependent enzyme, catalyzes the degradation of methionine to the products α-ketobutyrate, ammonia and methanethiol. Therefore, MGL has been proposed as a candidate for anticancer therapy. However, MGLs are only found in bacteria, protozoa and plants, not in mammals. For therapeutic applications, natural MGLs may not be sufficient, due to immunogenicity and limited stability, activity or specificity. Therefore, engineered variants may be needed. The catalytic reaction of MGL for methionine elimination is a multistep reaction. Limited structures of intermediates were solved. Improved understanding of the reaction mechanisms is also needed, to guide engineering efforts. Here, we use molecular modelling to study the internal aldimine intermediate of MGL with and without the presence of substrate Met at the active site. By using molecular dynamics simulations and thermodynamic perturbation theory, our models reveal the preferred protonation states of PLP in MGLs, and fine details of the structures and dynamics.

SWORD2: HIERARCHICAL ANALYSIS OF PROTEIN 3D STRUCTURES

GABRIEL CRETIN[a], TATIANA GALOCHKINA[a], YANN VANDER MEERSCHE[a], ALEXANDRE G. DE BREVERN[a], GUILLAUME POSTIC[b] JEAN-CHRISTOPHE GELLY[a]

[a] Université Paris Cité and Université des Antilles and Université de la Réunion, INSERM, BIGR, F-75014 Paris, France

[b] Université Paris-Saclay, Univ Évry, IBISC, Évry-Courcouronnes, France

Background: Structural domains are the usual unit of analysis of protein structure. Thus, understanding the functions and evolution of proteins requires dividing them into regions that are both relevant in terms of structural independence, folding, activity, or evolution. Besides, apart from structural domains, subdomains and supersecondary structures, which are shorter regions, they also provide valuable biological insights.

Objectives: Using Protein Units [1,2,3,4] (PUs), compact and evolutionarily preserved protein regions, the objective is to provide and assess multiple alternative decompositions of proteins into domains.

Methods: We present SWORD2 [1,6], a strategy for the exploration and analysis of protein structure. An input protein structure may be decomposed into PUs that can be hierarchically assembled to delimit structural domains. For each partitioning solution, the relevance of the identified substructures is estimated through different measures. This multilevel analysis is achieved by integrating our previous work on domain delineation, PU identification through 'protein peeling' and model quality assessment using the IG-score [5].

Results: The web server is freely available online: https://www.dsimb.inserm.fr/SWORD2

Conclusion: The SWORD2 partitioning algorithm produces multiple alternative domain assignments for a given protein structure. This unique approach handles ambiguous protein structure partitioning, admitting multiple relevant solutions.

References

[1] Postic, G., Ghouzam, Y., Chebrek, R., & Gelly, J. C. (2017). An ambiguity principle for assigning protein structural domains. Science advances, 3(1), e1600552.10.1126/sciadv.1600552

[2] Gelly JC, de Brevern AG. Protein Peeling 3D: new tools for analyzing protein structures. Bioinformatics. 2011 Jan 1;27(1):132-3.

[3] Gelly JC, Etchebest C, Hazout S, de Brevern AG. (2006) Protein Peeling 2: a web server to convert protein structures into series of protein units. Nucleic Acids Res. Jul 1;34 (Web Server issue):W75-8

[4] Gelly JC, de Brevern AG, Hazout S. (2006) 'Protein Peeling': an approach for splitting a 3D protein structure into compact fragments. Bioinformatics. Jan 15;22(2):129-33. Epub 2005 Nov 14.

[5] Postic G, Janel N, Tufféry P, Moroy G. An information gain-based approach for evaluating protein structure models. Comput Struct Biotechnol J. 2020;18:2228-2236.

[6] Cretin, G., Galochkina, T., Vander Meersche, Y., de Brevern, A. G., Postic, G., & Gelly, J. C. (2022). SWORD2: hierarchical analysis of protein 3D structures. Nucleic acids research, gkac370. 50(W1), W732–W738 10.1093/nar/gkac370

EVALUATION OF THE PREDICTION OF AMYLOID FIBRILLAR STRUCTURES BY ALPHAFOLD2

MEHDI MUNIM[a], VÉRONIQUE ARLUISON[b,c]
GAUTIER MOROY[a], SAMUEL MURAIL[a]

[a] Université Paris Cité, CNRS UMR 8251, INSERM U1133, Unité de Biologie Fonctionnelle et Adaptative, F-75013 Paris, France.
[b] Laboratoire Léon Brillouin LLB, CEA, CNRS UMR 12, Université Paris-Saclay, CEA Saclay, 91191 Gif-sur-Yvette, France.
[c] Université Paris Cité, UFR Sciences du vivant, 75006 Paris cedex, France.

Recent development of AlphaFold 2 algorithm allows structure prediction of single chain proteins from their sequences with an unprecedented accuracy, close to experimental methods[1]. Application of AlphaFold algorithm to protein complexes structure prediction, AlphaFold-Multimer, provides acceptable results in 63% of cases[2], making it the reference method for complex structure prediction.

In this study, we aim to assess Alphafold algorithm capability to predict the 3D structure of amyloid peptides, which can considered as protein complexes. Amyloids are not only involved in several pathologies (Alzheimer, Parkinson) but they also have functional roles (bacterial biofilm[3]). Based on AmyPro[4], we built a data-set of amyloid peptides for which structure have been determined experimentally. We selected 56 sequences ranging in size from 6 to 140 residues, and belonging to 15 different amyloid families (amyloid beta, alpha synuclein, HET-s prion...). This data-set contains 77 amyloid peptide structures from the Protein Data Bank.

To evaluate the capability of AlphaFold to predict the structure of the amyloid sequences, we systematically tested several crucial parameters of AlphaFold-Multimer/ColabFold (AlphaFold version, sampling, recycle, dropout[5], peptide number, ...) and compare prediction to available PDB structures. In total we ran around ~200.000 AlphaFold model calculation. We intended to correlate the quality of prediction with the input parameters as long as features derived from the sequences. As of now, the number of recycling as well as the dropout slightly improved AlphaFold results on amyloids. Our results showed conversely that the prediction of amyloids clearly depends on the peptide sequence. For example the structure of the HET-s prion was successfully predicted by AlphaFold whereas it was unable to predict the structure of the Tau protein.

References

[1] Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. Nature 596, 583–589 (2021).
[2] Evans, R. et al. Protein complex prediction with AlphaFold-Multimer.
http://biorxiv.org/lookup/doi/10.1101/2021.10.04.463034 (2021)
[3] Otzen, D. et al. Functional Amyloids. Cold Spring Harb. Perspect. Biol. 11, a033860 (2019).
[4] Varadi, M. et al. AmyPro: a database of proteins with validated amyloidogenic regions. Nucleic Acids Res. 46, D387–D392 (2018).
[5] Wallner, B. AFsample: Improving Multimer Prediction with AlphaFold using Aggressive Sampling.
http://biorxiv.org/lookup/doi/10.1101/2022.12.20.521205 (2022)

MDLON: COMMUNITY DETECTION IN MOLECULAR DYNAMICS SIMULATIONS VIA LOCAL OPTIMA NETWORKS ANALYSIS

JELENA VUCINIC[a], JINMEI GAO[a], DAVID SIMONCINI[a]

[a] Université Fédérale de Toulouse, ANITI, IRIT-CNRS UMR 5505, Université Toulouse I Capitole, 31000 Toulouse, France

Molecular Dynamics (MD) simulations have emerged over the last decades as the method of reference for analyzing protein motions. It allows understanding countless biological properties and key molecular level mechanisms through the analysis of quantitative information that is generated over the course of the simulations. Treating this abundant data with pertinent analysis and visualization techniques may uncover particularly important features that would remain elusive otherwise. Standard techniques such as RMSD, RMSF and PCA analysis have proved very useful but are not without drawbacks such as loss of spatial correlation or lossy data reduction. In this work we propose *MDlon*, a new local optima network-based analysis tool for MD simulations. *MDlon* identifies local optima according to potential energy in a graph in which conformations are linked by an edge using a relative neighborhood relationship. The local optima network (LON) is constructed by connecting local optima when the intersection of their basins of attraction is non-null. In such cases, edges between local minima are weighted by the probability of reaching one node from the other. The Leiden algorithm is applied to detect the community structure of the LON using modularity. To showcase our approach, we performed 100 *ns* MD simulations on two different proteins and analysed the trajectories with *MDlon*. The LON we construct allows detecting strongly connected communities corresponding to protein macrostates and computing transition probabilities between different macrostates. We show that *MDlon* encompasses all the information revealed by standard PCA analysis. Also, network statistics such as betweenness centrality allow identifying key microstates and capture the free-energy landscape topology. *MDlon* is a new MD analysis and visualization tool which aims at extracting additional knowledge from MD simulations through the lens of local optima networks. It is developed in Python using holoviews and bokeh libraries for interactive and scalable network visualization. Our software will be made available to the community as an easy-to-use open source python script which allows to fastly run network-oriented MD analysis.

MODELING RNA POLYMORPHISM

SAMUELA PASQUALI[a]

[a] Laboratoire Biologie Fonctionnelle et Adaptative, CNRS UMR 8251 Equipe Modélisation Computationnelle des Interactions Protéine-Ligand, Univesité Paris Cité

RNA molecules are characterized by the existence of a multitude of stable states that result in a frustrated energy landscape, where the observed structures depend sensibly on experimental conditions and can depend on the initial, unfolded, structure. Using both atomistic [1,2,3] and coarse-grained models for RNAs [4], combined with enhanced sampling methods, we investigate the energy landscape of these systems to understand what are the most relevant structures in the different conditions. Because of the complexity of the landscape, it is useful to guide the modeling process with experimental data and placing the molecule in the environment sampled in experiments. We have recently developed a simulation technique allowing to bias MD coarse-grained simulations with SAXS data onthe-fly [5], and a theoretical framework to perform fast constant pH simulations where we can model the system considering the exchange of charges with the solvent [6]. These developments allow us to account for the environment to obtain reasonable structures to then be studied more thoroughly with high-resolution modeling, also introducing the comparison with SHAPE data from first principles [7].

References

[1] K Röder, G Stirnemann, AC Dock-Bregeon, DJ Wales, S Pasquali, "Structural transitions in the RNA 7SK 5' hairpin and their effect on HEXIM binding", Nucleic Acids Research 48 (1), 373-389 (2020)

[2] K Röder, AM Barker, A Whitehouse, S Pasquali, "Investigating the structural changes due to adenosine methylation of the Kaposi's sarcoma-associated herpes virus ORF50 transcript", PLOS Computational Biology 2022, 18(5):e1010150, doi: 10.1371/journal.pcbi.1010150, PMID: 35617364

[3] G Lazzeri, C Micheletto, S Pasquali, P Faccioli, "RNA Folding Pathways from All-Atom Simulations with a Variationally Improved History-Dependent Bias" arXiv:2205.12603 (2022)

[4] T. Cragnolini, Y. Laurin, P. Derreumaux, S. Pasquali, "The coarse-grained HiRE-RNA model for de novo calculations of RNA free energy surfaces, folding, pathways and complex structure predictions", J. Chem. Theory Comput., 11, 3510 (2015)

[5] L Mazzanti, L Alferkh, E Frezza, S Pasquali, "Biasing RNA coarse-grained folding simulations with Small--Angle X--ray Scattering (SAXS) data", J. Chem. Theory Comput., 17, 6509-6521 (2021)

[6] S. Pasquali, E. Frezza, F.L. Barroso da Silva, Coarse-grained dynamic RNA titration simulations, Interface Focus 9: 20180066 (2019)

[7] E. Frezza, A. Courban, D. Allouche, B. Sargueil, S. Pasquali, "The interplay between molecular flexibility and RNA chemical probing reactivities analyzed at the nucleotide level via an extensive molecular dynamics study", Methods 162-163:108-127 (2019)

MODELLING AND MOLECULAR DYNAMICS OF CARBAMYLATED COLLAGEN I WITH FOCUS ON AGEING PROCESSES

ZARA MSOILI[a], PINAR HASLAK[a], MANUEL DAUCHEZ[a,b]
HUA WONG[a], STÉPHANIE BAUD[a,b]

[a] Université de Reims Champagne-Ardenne, CNRS UMR 7369, MEDyC, Reims 51687, France
[b] Université de Reims Champagne-Ardenne, P3M, Multi-Scale Molecular Modeling Platform, Reims 51687, France.

Background: Collagens, the most abundant proteins in mammals, are key components in the integrity, the structure, and the physico-chemical properties of the extracellular matrix (ECM). Type I collagen is a triple helix [1], that can be found in skin, eyes, bones, and many other tissues. One of the modifications occurring in collagens is carbamylation: a post-translational (PTM) nonenzymatic reaction2. In chronic kidney disease (CKD), the accumulation of urea can react with arginine or lysine and form citrulline or homocitrulline (HCT). Recent studies from our research unit showed experimentally that the accumulation of carbamylated derived products such as HCT in skin could alter the properties of collagen and be correlated to skin ageing[3,4].

Objectives: The aim of this project is to decipher the mechanisms responsible for this damage at the molecular and atomic level. To this end, we propose a numerical strategy mainly based on the use of molecular dynamics (MD) as a powerful computational model applied to collagen I.

Methods: An in-silico model of collagen I was built, and MD simulations were performed on specific regions with Gromacs 2020.4[5] and AMBER99SB*-ILDNP forcefield[6]. Four different systems containing 0 to 3 HCT were investigated. Since there is no parameter to describe HCT in the AMBER forcefield, this PTM was parametrized with quantum mechanics using HF/6-31G(d) level of theory. The MD trajectories were analysed with in house scripts to evaluate: (i) the geometry of proline and hydroxyproline residues[7] located near the HCT residue, and (ii) the conformation of the HCT side chain.

Results: The presence of one to three HCT in the same region has no impact on the overall structure (polyproline-II structure is conserved) but seems to influence locally the dynamics of triple helices. Indeed, the characterization of the dihedral angles of the HCT side chain highlights a change in the χ_4 behaviour compared to lysine. In the presence of a single carbamylation, no major difference in ring topology was observed for proline or hydroxyproline.

Conclusions: Our study provides insight into the impact of carbamylation at the atomic level: only local and not global effects were demonstrated in the presence of multiple carbamylations. Further research will focus on the potential collaborative effects that could occur in a system with multiple modifications on different regions and thus closer to the physiological reality.

References

[1] Ricard-Blum S. The Collagen Family. Cold Spring Harbor Perspectives in Biology. 2011;3(1):a004978-a004978. doi:10.1101/cshperspect.a004978

[2] Jaisson S, Pietrement C, Gillery P. Carbamylation-Derived Products: Bioactive Compounds and Potential Biomarkers in Chronic Renal Failure and Atherosclerosis. Clinical Chemistry. 2011;57(11):1499-1505. doi:10.1373/clinchem.2011.163188

[3] Gorisse L, Pietrement C, Vuiblet V, et al. Protein carbamylation is a hallmark of aging. Proc Natl Acad Sci USA. 2016;113(5):1191-1196. doi:10.1073/pnas.1517096113

[4] Gorisse L, Jaisson S, Piétrement C, Gillery P. Carbamylated Proteins in Renal Disease: Aggravating Factors or Just Biomarkers? Int J Mol Sci. 2022;23(1). doi:10.3390/ijms23010574

[5] Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJC. GROMACS: Fast, flexible, and free. J Comput Chem. 2005;26(16):1701-1718. doi:10.1002/jcc.20291

[6] Aliev AE, Kulke M, Khaneja HS, Chudasama V, Sheppard TD, Lanigan RM. Motional timescale predictions by molecular dynamics simulations: Case study using proline and hydroxyproline sidechain dynamics: Proline Force Field Parameters. Proteins. 2014;82(2):195-215. doi:10.1002/prot.24350

[7] Wu D. The puckering free-energy surface of proline. AIP Advances. 2013;3(3):032141. doi:10.1063/1.4799082

COARSE-GRAINED SIMULATION MODEL OF PROTEIN TDP43 LIQUID-LIQUID PHASE SEPARATION BEHAVIOR

XIAOFEI PING[1,2,3,4], LUKAS STELZL [1,2,3,4]

[1] Max Planck Graduate Center mit der Johannes Gutenberg-Universität Mainz, 55128 Mainz, Germany
[2] Institute of Molecular Biology (IMB), 55128 Mainz, Germany
[3] Faculty of Biology, Johannes Gutenberg University, 55128 Mainz, Germany
[4] KOMET 1, Institute of Physics, Johannes Gutenberg University, 55099 Mainz, Germany

TDP-43 (TAR DNA-binding protein 43) is a protein that involves in the pathogenesis of several neurodegenerative diseases, including Alzheimer's disease (AD), Amyotrophic lateral sclerosis (ALS), and Frontotemporal lobar degeneration (FTLD). The formation of protein aggregates in the brain and spinal cord of patients is underpinned by liquid-liquid phase separation (LLPS)[1].

In the context of TDP-43 and neurodegenerative diseases, molecular dynamics simulations can provide insight into the phase separation behavior of TDP-43 condensates and offer a powerful computational tool for investigating the behavior of molecules at the atomic or molecular scale, providing a detailed understanding of the driving force behind phase separation and aggregate formation[2].

Two main types of simulations, atomistic and coarse-grained, are utilized in the study of molecular systems. Atomistic simulations provide a high level of detail and accuracy by representing each atom in the molecule. In contrast, coarse-grained models, such as the near-atomic resolution MARTINI and the residue-level Hydropathy scale (HPS) models, enable the investigation of larger systems and longer timescales. Simulations of TDP-43 phase behavior with a chemically detailed coarse-grained model (Martini3 Go-like model), which accurately captures the dynamics of both folded and unfolded domains in molecular dynamics simulations. Increasing protein-water interaction strength can improve the model to capture the LLPS behavior of TDP-43 as determined by experiments[3]. The molecular dynamics simulations highlight parts of the protein which may be driving phase separation in health and disease.

References

[1] Austin, James A., et al. Disease Causing Mutants of TDP-43 Nucleic Acid Binding Domains Are Resistant to Aggregation and Have Increased Stability and Half-Life, PNAS, 2014, 4309–4314.

[2] Thomasen, F. Emil, et al. Improving Martini 3 for Disordered and Multidomain Proteins, J. Chem. Theory Comput., 2022, 2033–2041.

STRUCTURAL DYNAMICS OF SUPEROXIDE DISMUTASE (SOD) 1 A4V VARIANT ON A 10 MICROSECOND TIMESCALE

MUNISHIKHA KALIA[a,b], AMMAR AL-CHALABI[b,c]
ALFREDO IACOANGELI[a,b,c]

[a] Department of Biostatistics and Health Informatics, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK
[b] Department of Basic and Clinical Neuroscience, Maurice Wohl Clinical Neuroscience Institute, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, SE5 9NU, UK
[c] National Institute for Health Research Biomedical Research Centre and Dementia Unit at South London and Maudsley NHS Foundation Trust and King's College London, London, UK
*Email : munishikha.kalia@kcl.ac.uk

Amyotrophic Lateral Sclerosis (ALS) is a fatal neurodegenerative disease, primarily affecting upper and lower motor neurons, that results in progressive weakness and culminates in death from neuromuscular respiratory failure, typically 2–5 years after diagnosis. Superoxide Dismutase 1 (SOD1) was the first gene to be associated with ALS. SOD1 is an antioxidant enzyme which destroys superoxide O2- radical in mammalian cell cytoplasm. SOD1 mutations are the second commonest known cause of ALS, responsible for approx. 20% of familial cases and 2% of sporadic cases. More than 200 SOD1 variants have been reported in ALS patients and new variants are identified in ALS patients on a regular basis. A4V variant is responsible for 50% of SOD1 mutations in North American familial ALS patients[1]. One of the most striking features of A4V variant is the rapid disease progression, with a mean survival time of less than 2 years from clinical onset[2].

The aim of our study is to investigate the structural and conformational differences between the A4V variant of the SOD1. For this, we performed 10 microsecond length molecular dynamics (MD) simulations of the A4V SOD1. We observed that the A4V variant adopted an open and a closed conformation. In addition, we observed that the electrostatic loop region was highly flexible in the A4V SOD1. These extensive MD simulations shed light into the role of the ESL loop of SOD1 in the pathogenesis of SOD1 induced ALS.

References

[1] Andersen, P.M., Current neurology and neuroscience reports, 2006. 6(1): p. 37-46.
[2] Saeed, M, Y Yang, H-X Deng, W-Y Hung, N Siddique, L Dellefave, C Gellera, P M. Andersen, and T Siddique, Neurology, 2009. 72(19): p. 1634-1639.

INVESTIGATION OF A NOVEL PRESSURE-SENSITIZED TACAN STRUCTURE UNDER HIGH-THRESHOLD SYSTEM PRESSURE.

HUIXIA LU[a,b,c], JORDI FARAUDO[a], JORDI MARTI[b], BUYONG MA[c]

[a] Department of Materials Simulation and Theory, Institut de Ciencia de Materials de Barcelona (ICMAB-CSIC), Campus de la UAB, Bellaterra, Barcelona E-08193, Spain

[b] Department of Physics, Technical University of Catalonia-Barcelona Tech (UPC), B4-B5 Campus Nord, Jordi Girona 1-3, 08034 Barcelona, Catalonia, Spain

[c] Department of Pharmacy, Shanghai Jiao Tong University, Dongchuan Road No. 800, Shanghai, China

TACAN has been shown to contribute to the mechanical hyperalgesia with the mechanism unknown. In 2020, Lou Beaulieu-Laroche et al. proposed that TACAN may function as a high threshold mechanically activated ion channel involved in sensing mechanical pain[1]. However, other research groups deny the potential role that TACAN may play in the mechanical signaling path cross the cell membrane. Taken all together, despite the previous report that suggested TACAN might function as a high-threshold mechanically activated cation channel responsible for sensing mechanical pain in mice, the mechanism of mechanosensation and ion permeation mediated by TACAN channel remains elusive and awaits further investigation. In this study, we show a novel human TACAN dimer structure. It is a homodimer with each monomer consisting of a body, a spring and a blade domains. More importantly, we found that all the helices of the body and the spring domains are specifically associated with 11 membrane lipid molecules. Particularly, a lipid core, residing within a cavity formed by the two body and spring domains, contacts with the helices from the body and spring domains and extends to reach two symmetrically arranged lipid clusters. Actually, a previous study showed that the membrane lipids have the intrinsic feature leading to the possibility of soliton propagation in nerves[2]. Therefore, our results shall suggest that the cryo-EM structure of TACAN associated with lipids is novel and the membrane lipids may play an important biological role in tunning the function of TACAN upon mechanical (eg. pressure) stimuli. TACAN may use membrane lipids to directly sense and transduce the mechanic stimulus, whereas the protein part of TACAN may function to modulate the action of lipids. Molecular dynamics simulations suggest that TACAN isn't an ion channel, but signal transduction along TACAN tends to be gated by the membrane cardiolipin lipids upon high-threshold pressure stimuli upon upstream mechanical force.

References

[1] Beaulieu-Laroche, Lou, et al. «TACAN is an ion channel involved in sensing mechanical pain.» Cell 180.5 (2020): 956-967.

[2] Heimburg T, Jackson AD. On soliton propagation in biomembranes and nerves. Proceedings of the National Academy of Sciences of the United States of America 102, 9790-9795 (2005).

STRUCTURAL ANALYSIS OF THE EVOLUTION OF PROTEIN-RNA INTERACTIONS

IKRAM MAHMOUDI[a], JESSICA ANDREANI[a]

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

Protein-RNA interactions are crucial for the proper functioning of the cells. They are also involved in various pathologies particularly in cancer[1]. Even if recent advances greatly expanded the large-scale binding data of protein-RNA interactions, the scarcity of available structural data remains problematic to understand the molecular details of the interactions. In our study, we aim at developing reliable computational methods to predict protein-RNA 3D structures that may enrich the detailed molecular data. A robust computational prediction method is homology-based protein-RNA interface modeling[2]. This approach requires detailed understanding of protein-RNA interfaces and how they evolve. In this context, we performed a detailed evolutionary analysis of protein-RNA interface structures.

Our starting point is collecting the available structural data for protein-RNA complexes, 4,307 structures of protein-RNA complexes available in the Protein Data Bank as of May 2022. The 4307 complexes were then splitted into 86000 protein-RNA interfaces. We filtered out NMR structures, low-resolution structures and small interfaces. We merged pairwise interfaces containing RNA chains forming a double helix. Conservative clustering to remove protein and RNA chains identical in sequence yielded 976 representative interface clusters.

We then performed all-against-all structural comparisons between the 976 representative interfaces. We computed structural similarity scores for interface, protein and RNA, as well as sequence identity and coverage for protein and RNA based on either sequence or structural comparisons, interface RMSD and proportion of conserved interface contacts. Combining multiple similarity score thresholds, we identified 1956 pairs of structural homologs (called interologs) which were splitted into ribosome/non-ribosome interologs. Overall, 69% of all distance-based interface contacts are conserved between interologs in our dataset. However, interestingly, only 21% of hydrogen bonds and 25% of stacking contacts are conserved. Within the ribosome interologs the proportions of conserved contacts were 67%, 11% and 9% respectively for distance-based contacts, H bonds and stacking, while for the non-ribosome interologs those percentages were 76%, 46% and 44% respectively. Similar trends were observed whether the hydrogen bonds were formed through the phosphate, sugar or base.

In the near future, we aim to understand which key interface features are relevant when performing homology modeling of protein-RNA structures and what information we can transfer within a pair of structural homologs. In parallel areas of the project, we are integrating omics data with structural information to boost protein-RNA interactions datasets. We also plan to derive a statistical potential for the scoring of protein-RNA interface models by including evolutionary information.

References

[1] Neelamraju, Y., Gonzalez-Perez, A., Bhat-Nakshatri, P., Nakshatri, H., & Janga, S. C. (2018). Mutational landscape of RNA-binding proteins in human cancers. RNA biology, 15(1), 115-129.

[2] Zheng, J., Kundrotas, P. J., Vakser, I. A., & Liu, S. (2016). Template-based modeling of protein-RNA interactions. PLoS computational biology, 12(9), e1005120.

A NEW HAC ALGORITHM TO CREATE THE SMALLEST REPRESENTATIVE SET FROM AN ENSEMBLE OF 3D CONFORMATIONS

ISAURE CHAUVOT DE BEAUCHENE[a] , ANTOINE MONIOT[a], YANN GUERMEUR[a]

[a] LORIA (CNRS - INRIA - Université de Lorraine), Nancy, 54000, France

Background: Many molecular modeling methods use 3D structural libraries that discretize the conformational space of (fragments of) molecules. A classical way to create such libraries is to cluster experimental structures extracted from databases, then to select few representatives or "prototypes". For most applications, the set of prototypes must satisfy two contradictory constraints: be representative of all existing conformations (to maximize the precision of the final models) and be of cardinality as small as possible (to avoid combinatorial explosions).

Objectives: We want to create the smallest set of prototypes from an ensemble of 3D conformations such that each conformation is well represented, i.e. is within a threshold deviation from at least one prototype.

Methods: Our problem corresponds to inferring an epsilon-net of minimal cardinality. This combinatorial optimization problem is NP-hard and still requires the specification of dedicated methods. However, feasible solutions can be derived at negligible cost from dendrograms produced by hierarchical agglomerative clustering (HAC) algorithms. But the corresponding linkage functions are unevenly suited to the task. We thus developed an HAC algorithm with a new linkage called radius, based on the computation of minimum enclosing balls, i.e the smallest ball that comprise all points of a cluster[1]. This approach necessitates an Euclidean space or a kernel. To tackle the issue of RMSD after superimposition not being a distance (no triangular inequality), we iteratively aligned all conformations of the to-be-merged clusters on one and computed an approximate (upper bound) pairwise RMSD. We stop agglomerating when the next ball radius is above our deviation criteria, then keep the center of each ball as a prototype.

Results: As a first theoretical test, on a classical benchmark of images, we produced smaller epsilon-nets than all common HAC algorithms. Then, in the context of RNA fragment-based docking, we applied it on the 3D structures of trinucleotides (~20.000 per sequence) extracted from all protein-RNA complexes in the PDB. We obtained 566-1012 prototypes per RNA sequence with a 1 A threshold, which is a reduction by a factor of 4-5 compared to the conventional star-shape clustering that keeps the center point of each cluster as a prototype. This resulted in correspondingly faster docking calculations without decreasing the accuracy of the docking poses.

Conclusion: This approach is applicable to any set of 3D conformations of an object, and we foresee many applications within and beyond the field of molecular modeling.

References

[1] Moniot A, Chauvot de Beauchêne I, Guermeur Y. Inferring Epsilon-nets of Finite Sets in a RKHS, Advances in Self-Organizing Maps, Learning Vector Quantization, Clustering and Data Visualization (WSOM+), Jul 2022, Prague, 53-62.

DENSITY FITTING METHODS FOR QM/MM MODELS

IOANNA-MARIA LYGATSIKA[a], YVON MADAY[b]
JEAN-PHILIP PIQUEMAL[c], LOUIS LAGARDÈRE[d]
[a] ioanna-maria.lygatsika@sorbonne-universite.fr
[b] yvon.maday@sorbonne-universite.fr
[c] jean-philip.piquemal@sorbonne-universite.fr
[d] louis.lagardere@sorbonne-universite.fr

The background of this talk concerns the intermolecular (non-covalent) interactions in large molecular clusters, consisting of identical molecules, called fragments. The frozen-core approximation is assumed, in which atomic orbitals corresponding to core electrons of fragments remain fixed throughout electronic structure calculations. The electronic density of each fragment is assumed to be independent of other fragment positions and is approximated by applying Hartree-Fock or Kohn-Sham methods applied to the isolated fragment. In the context of energy decomposition methods, our objective is to perform fast evaluations of intermolecular interaction energies (Coulomb and exchange repulsion) of polymers. Such energies are four-center molecular integrals of frozen electronic densities of monomers and their cost scales as N^4, where N is the number of atomic orbital basis functions of a monomer. In force field calculations for biomolecular systems, the Gaussian electrostatic model[1] reduces this cost using the density fitting technique. In this talk, the existing method is demonstrated to be a special type of solution to a more general summation problem, that admits further optimization using dynamic programming techniques[2]. We present a new fast summation algorithm to accelerate intermolecular energies. The numerical results of our method are demonstrated on small TIP3P rigid water clusters, for geometries optimised using molecular dynamics. As a conclusion, results show that our method achieves N^2 scaling and lower approximation error than standard density fitting.

References

[1] J. Chem. Phys. 125, 184101 (2006).
[2] European Journal of Operational Research Volume 261, Issue 2, 1 September 2017, Pages 530-539.

EXPLORING THE RELATIONSHIP BETWEEN MICROTUBULE ARCHITECTURE AND MECHANICS THROUGH NETWORK-BASED MULTISCALE ANALYSIS OF TUBULIN DYNAMICS

MARCO CANNARIATO[a], ERIC A. ZIZZI[a], MARCO A. DERIU[a]

[a] Polito[BIO]Med Lab, Department of Mechanical and Aerospace Engineering, Politecnico di Torino, Turin, Italy

The study of microtubule (MT) mechanics and the pathways involved in the transfer of vibrations between tubulins is crucial to understand how MTs are stabilized within the cell. Indeed, the hierarchical organization of MTs is the basis of their stability, mechanics, and function[1]. MTs mechanics may depend on tubulin interactions, protofilament (PF) number, or the presence of stabilizers such as the anti-cancer drug Taxol. Therefore, this work aims at shedding light on the relationship between MT architecture and mechanics, considering the effect of Taxol and the number of PFs. We employed a multiscale approach integrating all-atom molecular dynamics (MD) simulations, Protein Structure Networks (PSN), and Normal Mode Analysis (NMA) on Elastic Network Models (ENM). All-atom structures of the dimers, at different PF numbers and in presence of Taxol, were simulated in systems representative of the MT wall. Through the PSN, we studied the propagation of vibrations between tubulins within the MT. MTs of lengths between 250 nm and 400 nm were built by fitting tubulin conformations from MD simulations onto an Electron Microscopy Density (EMD) map and replicating them axially with an experimental step[2]. Mechanical properties were then derived by MT vibrational frequencies from NMA of ENM, where residues were connected if closer than 1.2 nm[3]. The results highlight that β-tubulin drives the transfer of vibrations between PFs, information relatable to previous evidence1, and that such communication is altered in presence of Taxol. Moreover, Taxol stabilizes the MT by reducing the fluctuations of the tubulin dimer bending angle. At a higher scale, our results revealed that Taxol influences mainly the shear modulus, reflecting the atomistic result that Taxol alters the inter-PFs interaction. Moreover, MTs with different PF numbers were characterized by changes in mechanical properties, with the persistence length being linearly correlated with the number of PFs. Finally, we observed remarkable differences comparing the mechanics of MTs built with tubulins derived from simulations both at the proper wall curvature and at the 13PF curvature, highlighting the importance of the multiscale approach coupling MD with ENM. Taken together, we provide additional information to understand the nanomechanics of MT and how it can be altered to achieve its stabilization, which could be exploited for the development of biomimetic materials and to guide the design of new MT-stabilizing drugs.

References

[1] Janke et al, The tubulin code and its role in controlling microtubule properties and functions, Nat Rev Mol Cell Biol, 2020, 21:307-326.

[2] Sui et al, Structural Basis of Interprotofilament Interaction and Lateral Deformation of Microtubules, Structure, 2010, 18:1022-1031.

[3] Deriu et al, Anisotropic Elastic Network Modelling of entire Microtubules, Biophysical Journal, 2010, 99:2190-2199

PREDICTING NON-COVALENT INTERACTIONS BETWEEN ANTIOXIDANTS IN BIOLOGICAL MEMBRANES THROUGH MOLECULAR DYNAMICS

MARVING MARTIN*[a,b], PATRICK TROUILLAS[a, c]

[a] INSERM U1248 Pharmacology & Transplantation, Univ. Limoges, CBRS, 2 rue du prof. Descottes, F-87000 Limoges, France
[b] InSiliBio, 1 avenue d'Ester, Ester Technopôle, F-87000 Limoges, France
[c] RCPTM CATRIN Palacky University, Slechtitelu 27, 783 71, Olomouc, Czech Republic

Background. There have been many evidences about non-covalent interactions between polyphenols. In water, the non-covalent association between anthocyanins and other polyphenols is called copigmentation. This association is mainly ensured by π-π stacking interactions. The copigmentation process participates in stabilization of the anthocyanin color in plants and derived beverages like red wine. This non-covalent association has also been described in lipid bilayer membranes in a joint fluorescence and in-silico study. This association is key to rationalize synergism between π-conjugated antioxidants including polyphenols (e.g. quercetin) and vitamin E (α-tocopherol). Such synergism process has gained much interest in food industry as it allows developing efficient antioxidant cocktails, reducing concentration of active agents and subsequently potential toxicity.

Objectives. This study aims at benchmarking the performance of molecular dynamics (MD) to predict the formation of non-covalent complexes between π-conjugated antioxidants, including two prototypes (quercetin and vitamin E), in a pure 1,2-dipalmitoylphosphatidylcholine (DOPC) lipid bilayer.

Methods. The association between quercetin and vitamin E was studied through free MD simulations of µs-time scale with the lipid17 forcefield. Experimentally, the quenching of vitamin E fluorescence by quercetin was found to be mainly ruled by a sphere-of-action quenching model and to a lower extent (20%) by the formation of transient π-π stacking complexes. The equation fitting picturing this model exhibited a correlation coefficient R^2 of 0.99.

Results. MD simulations revealed high capacity to reproduce quenching intensity, with a 0.39 mean absolute error, whilst the mean ratio of π-π stacking complex formation was evaluated at 18%. The predictive capability of MD simulations at capturing this non-covalent association was evaluated with five other π-conjugated potential partners of vitamin E, namely the five antioxidants catechin, caffeic acid, myricetin, kaempferol and galangin. The observed trend agreed with experimental studies, validating again the use of MD simulations as a tool to measure the potential synergism between active ingredients.

References

[1] Fabre G., Defining a subject in a concise format, Chem. Commun., 2015, 7713-7716

UNDERSTANDING PASSIVE MEMBRANE PERMEATION OF PEPTIDES: PHYSICAL MODELS AND SAMPLING METHODS COMPARED

LIUBA MAZZANTI[a], TÂP HA-DUONG[a]

[a] Université Paris-Saclay, CNRS, BioCis, 17 Avenue des Sciences, 91400 Orsay, France

Characterizing the permeation process of drug candidate molecules across biomembranes is a crucial aspect in the machinery of drug design. While active transport occurs, namely for endogenous substances, passive diffusion across a lipid bilayer represents the dominant mechanism for xenobiotics. Among them, peptides are considered suitable compounds for modulating protein-protein interactions thanks to their relatively large size, making them promising drug candidates. Since their size may represent an obstacle for membrane permeation, studying this process is especially relevant to anticipate their pharmacokinetic properties. With the aim to select a method that could profitably be applied to peptide-based therapeutics, we compared two protocols by assessing their solidity and computational costs[1].

On the one hand, we considered the wide-spread inhomogeneous solubility-diffusion model (ISDM)[2]. In this framework, the permeability coefficient is related to the free energy profile (FEP) and diffusion coefficient along the transport direction, both evaluated from umbrella sampling (US) simulations. On the other hand, in the kinetic model characterized by four phases (outer water, outer lipid, inner lipid, inner water compartments), the permeability coefficient is proportional to the rate constant associated to the fastest transition[3]. We built a Markov state model (MSM)[4] from unconstrained simulations in order to estimate kinetic quantities. This requires a prior knowledge of the FEP, which is provided by the weighted histogram analysis method[5] for the ISDM US trajectories. Hence, higher computational costs are inevitably entailed in the MSM framework, with respect to the ISDM protocol.

We compare the two computational approaches for a benchmark peptide to experimental results from planar artificial membrane assays. While ISDM underestimates the permeability coefficient, possibly because of the difference between the experimental and computational membrane setup, MSM largely overestimates it. Although this might be due to insufficient sampling, overcoming this issue would require an extremely large amount of computational resources as compared to the ISDM approach. We conclude that efficiency and reliability favor ISDM for peptide membrane permeability predictions.

References

[1] Mazzanti L., Ha-Duong T. Understanding passive membrane permeation of peptides: physical models and sampling methods compared, Int. J. Mol. Sci., 2023, 24, 5021.

[2] Marrink, S.J., Berendsen, H.J.C. Simulation of water transport through a lipid membrane, J. Phys. Chem., 1994, 98, 4155-4168.

[3] Prinz, J.H., Wu, H., Sarich, M., Keller, B., Senne, M., Held, M., Chodera, J.D., Schutte, C., Noé, F. Markov models of molecular kinetics: Generation and validation, J. Chem. Phys., 2011, 134, 174105.

[4] Dickson, C.J., Hornak, V., Pearlstein, R.A., Duca, J.S. Structure–Kinetic Relationships of Passive Membrane Permeation from Multiscale Modeling, J. Am. Chem. Soc., 2017, 139, 442-452.

[5] Kumar, S., Rosenberg, J.M., Bouzida, D., Swendsen, R.H., Kollman, P.A. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method, J. Comput. Chem., 1992, 13, 1011-1021.

MDVERSE: SHEDDING A FIRST LIGHT ONTO THE DARK MATTER OF MOLECULAR DYNAMICS

JOHANNA KS TIEMANN[a], LISA BOUARROUDJ[b]
MOHAMED OUSSAREN[b], MAGDALENA SZCZUKA[c]
MATTHIEU CHAVENT[c], PIERRE POULAIN[b]

[a] Linderstrøm-Lang Centre for Protein Science, Section for Computational and RNA Biology, Department of Biology, University of Copenhagen, DK-2200 Copenhagen N, Denmark

[b] Université Paris Cité, CNRS, Institut Jacques Monod, F-75013 Paris, France

[c] Institut de Pharmacologie et Biologie Structurale, CNRS, Université de Toulouse, 205 route de Narbonne, 31400, Toulouse, France

The rise of open science and the absence of a dedicated data repository for molecular dynamics (MD) simulations has led to the accumulation of MD files in generalist data repositories[1], constituting *the dark matter of MD*. Leveraging an original search strategy based on file types and keywords, we found and indexed about 250,000 files and 2,000 datasets from the data repositories Zenodo, Figshare and Open Science Framework. We have shown that the number of MD files deposited each year in these data repositories is constantly increasing, especially in Zenodo. With a focus on files produced by the Gromacs MD software, we illustrated the high potential offered by the mining of structures (.gro files), parameters (.mdp files) and trajectories (.xtc files) from publicly available MD datasets. We identified for instance systems with specific molecular composition (i.e. protein, lipid, nucleic acid, saccharide...). We were also able to characterize essential parameters of MD simulations, like temperature and simulation length, and to identify model resolution, such as all-atom and coarse-grain[2]. Based on this analysis, we proposed a search engine prototype for researchers to explore collected MD data.

https://mdverse.streamlit.app/

References

[1] Hospital A, Battistini F, Soliva R, Gelpí JL, Orozco M. Surviving the deluge of biosimulation data. WIREs Comput Mol Sci. 2020;10(3). doi:10.1002/wcms.1449

[2] Ingólfsson HI, Lopez CA, Uusitalo JJ, et al. The power of coarse graining in biomolecular simulations. WIREs Comput Mol Sci. 2014;4(3):225-248. doi:10.1002/wcms.1169

BINDING POSE PREDICTION OF PROTEASE INHIBITOR USING SUPERVISED MOLECULAR DYNAMICS (SUMD)

PETER E.G.F. IBRAHIM[a], XIAO HU[a], FABIO ZUCCOTTO ULRICH ZACHARIAE[a], IAN GILBERT[a]

[a] Drug Discovery Unit, Division of Biological Chemistry and Drug Discovery, University of Dundee, Dow St, Dundee DD1 5EH, UK

The correct determination of the ligand mode of binding in a biological target binding site is essential to gain insight into the molecular recognition process, understand the nature of the interactions, and drive the molecular design in the drug discovery process. Current in-silico approaches rely mainly on molecular docking that is known to be fast but treats the protein as a rigid body. This creates a significant limitation as several side chains of binding site residues and loop conformation rearrangements are associated with induced-fit binding. To address this limitation, we have been exploring the implementation of supervised molecular dynamics (SuMD) a novel molecular dynamic approach recently developed by Moro et al[1].

Methods: SuMD allows exploration of the entire ligand-receptor binding event, in a nanosecond timescale, reducing the computational efforts in comparison to classical molecular dynamics protocols. The ligand is placed at a distance of at least 30 Å away from the protein catalytic cleft to avoid premature intramolecular interactions. A collection of short unbiased MD simulations is generated and concatenated together while monitoring the distance between the centre of mass of the ligand and the residues composing the binding site of the receptor (ΔdL-R) over a fixed time window (Δt, e.g., 300 ps).

Papain-like protease protein (PL-pro) for SARS-CoV-2 coronavirus[2] was used as a case study, where its catalytic binding site is composed of a P-loop that was observed with high flexibility in both crystallography and classical MD simulations (CMD)[3]. This P-loop is also one gate-keeping structure of a PLpro inhibitor. Three major conformations observed of P-loop: Closed in ligand-protein holo-crystal complex, semi-open in apo-crystal structure, and open after CMD simulation. This creates a critical challenge in the discovery and optimization of new hit compounds against such highly flexible binding site. Starting from the three different P-loop conformations; closed, semi-open, and open, we apply SuMD simulations to investigate the mode of binding of the PLpro inhibitor in solvent.

Results: The induced-fit binding event of PLpro inhibitor was successfully simulated using SuMD without interference to the potential energy landscape, all succeed in reproducing the crystalized reference. The preliminary results shows that SuMD emphasizes a noticeable capability to reproduce the crystallographic structure of ligand-protein complex independent from the starting position of the ligand and receptor binding site conformation. This can be a promising approach to overcome the challenges of exploring small molecules binding modes in a dynamic environment.

References

[1] Salmaso, V., & Moro, S. (2018). Bridging molecular docking to molecular dynamics in exploring ligand-protein recognition process: An overview. Frontiers in Pharmacology, 9(AUG), 1–16.

[2] Amin, S. A., & Jha, T. (2020). European Journal of Medicinal Chemistry Fight against novel coronavirus: A perspective of medicinal chemists. European Journal of Medicinal Chemistry, 201, 112559

[3] Ismail, M. I., Ragab, H. M., Bekhit, A. A., & Ibrahim, T. M. (2021). Targeting multiple conformations of SARS-CoV2 Papain-Like Protease for drug repositioning: An in-silico study. Computers in Biology and Medicine, 131(December 2020), 104295.

BINDING PREFERENCES OF SMALL BIOMOLECULES FOR SPHERICAL ZnS NANOPARTICLE AND ZnS (110) PLANE: EFFECT OF SURFACE COATING AND SHAPE OF ZnS NANOSTRUCTURES

ROJA RAHMANI[a], ALEXANDER P. LYUBARTSEV[a]

[a] Department of Materials and Environmental Chemistry, Stockholm University, Stockholm, Sweden

Exposure to nanoparticles may result in major health issues, demanding research of any potential hazardous consequences as well as the discovery of their interactions in biological environments, which are connected to the physiochemical features of nanomaterials. Hence the modifications in these properties such as their shape, and surface coatings might affect their interactions in biological environments. ZnS nanomaterials have several applications in optoelectronic devices and biotechnology, used as biosensors. As well as their lower toxicity when compared to other materials in the II-VI group, the underlying mechanisms of interaction between ZnS NPs and macromolecules, such as proteins, lipids, and sugars, are still unknown [1].

The free energy of biomolecule adsorption on nanoparticles provides insight into the biological activities of these particles in the body. Here, we investigated the interactions of 29 small biomolecules (amino acid side chains, lipid fragments, glucose) as building blocks of proteins, lipid bilayers, and sugars with ZnS spherical nanoparticles (NPs) and ZnS (110) plane surface in aqueous solution using enhanced sampling, metadynamics, and atomistic molecular dynamics simulations [2]. We discovered that negatively charged amino acid side chains (aspartic acid (ASP), cysteine ion (CYM), and glutamic acid (GLU) bind to both NP and ZnS (110) and that NP is preferred over ZnS (110). In order to investigate the effect of coating on the binding affinity of biomolecules, we carried out simulations of ZnS nanostructures with polymethyl methacrylate (PMMA). The adsorption tendency of previously adsorbed biomolecules (ASP, CYM, and GLU) decreased significantly, but some new biomolecules were adsorbed on coated ZnS (110) because of their hydrophobic structure. Our research can be utilized to rationalize the design of peptides that bind selectively to ZnS nanostructures regarding their shape and character of coating molecules. In addition, our findings suggest that ZnS could be a safe choice in the biological environment because of its poor interactions with most of the building blocks of biomolecules, particularly lipid fragments.

References

[1] Power, D., Rouse, I., Poggio, S., Brandt, E., Lopez, H., Lyubartsev, A., & Lobaskin, V. (2019). A multiscale model of protein adsorption on a nanoparticle surface. Modelling and Simulation in Materials Science and Engineering, 27(8), 084003.

[2] Brandt, E. G., & Lyubartsev, A. P. (2015). Molecular dynamics simulations of adsorption of amino acid side chain analogues and a titanium binding peptide on the TiO2 (100) surface. The Journal of Physical Chemistry C, 119(32), 18126-18139.

PREDICTING PEPTIDES PERMEABILITY OF COMPLEX PEPTIDES

SELMA BENGAOUER[a], ELEONORA SERRA[b]
MASSINA ABDERRAHMANE[a], BRICE HOFFMANN[a]
NICOLAS DEVAUX[a], MAUD JUSOT[a]

[a] Iktos, 65 rue de Prony, 75017 Paris, France
[b] Istituto Italiano di Tecnologia, Genoa, Italia

Peptides represent a promising class of therapeutic molecules. They are of great interest for undruggable targets like protein-protein interactions (PPI) as they can occupy a larger surface of interaction compared to classical small molecules and reach higher specificities[1,2]. Despite these remarkable properties, peptides have suffered for many years from several drawbacks in the drug discovery field, with major barriers such as lack of oral bioavailability and poor membrane permeability. However, advances in chemistry to produce more stable peptides have been made in the last few decades, triggering a new interest for peptides as therapeutic drugs.

Predicting peptide properties using machine learning methods has gained interest in recent years, with a particular focus on anticancer[3] and antimicrobial[4] peptides or improving permeability properties[5]. Various methods have been proposed for prediction like support vector machines or random forest[3]. These predictive approaches have shown interesting performance, but also limitations, especially on the type of amino acids considered, most often restricted to natural amino acids, whereas the importance of modified amino acids has been demonstrated for peptide drug design.

To overcome those limitations, new representations of peptides, including different graph representations notably at the amino-acid level have been developed. These representations are suitable for designed peptides, taking into account modified amino acids and cross-links. We applied a circular algorithm to transform these graphs into a vectorial representation and evaluated them for predictive tasks, including permeability. Our models have shown improved performances compared to the baseline.

Finally, as the flexibility of peptides is one of the main bottlenecks in their drug likeness property prediction, the incorporation of 3D descriptors related to flexibility could be a real game changer in the strength of our models. Work has been initiated to include those 3D descriptors with an automated pipeline for molecular dynamics simulation including the parametrization of unnatural amino acids. Yet, the sampling of the peptide conformational landscape is challenging because of its multiple energetic minima and high energy barriers[6]. We will discuss how advanced molecular dynamics (replica exchange molecular dynamics and meta-dynamics) and the incorporation of local 3D descriptors could improve the prediction.

References

[1] Muttenthaler, Markus, et al. “Trends in peptide drug discovery.” Nature Reviews Drug Discovery 20.4 (2021): 309-325.

[2] Wang, Lei, et al. “Therapeutic peptides: current applications and future directions.” Signal Transduction and Targeted Therapy 7.1 (2022): 1-27

[3] Basith, Shaherin, et al. “Evolution of machine learning algorithms in the prediction and design of anticancer peptides.” Current Protein and Peptide Science 21.12 (2020): 1242-1250.

[4] Rondon-Villarreal, Paola, Daniel A Sierra, and Rodrigo Torres. “Machine learning in the rational design of antimicrobial peptides.” Current Computer-Aided Drug Design 10.3 (2014): 183-190.

[5] Wei, Huan-Huan, et al. “The development of machine learning methods in cell-penetrating peptides identification: a brief review.” Current Drug Metabolism 20.3 (2019): 217-223.

[6] Damjanovic, Jovan, et al. «Elucidating solution structures of cyclic peptides using molecular dynamics simulations.» Chemical reviews 121.4 (2021): 2292-2324.

IMPACT OF CONFORMATIONAL SHIFTS ON THE HAIRPIN RIBOZYME REACTIVITY

**SÉLÈNE FORGET[a] , MARIE JUILLE[a]
ELISE DUBOUE-DIJON[a], GUILLAUME STIRNEMANN[a]**

[a] Laboratoire de Biochimie Théorique, Institut de Biologie Physico-Chimique, PSL University, Université de Paris, 13 rue Pierre et Marie Curie, 75005 Paris, France

A key question for the RNA world hypothesis[1] is the emergence of autocatalytic networks in abiotic conditions without protein-based enzymes. Ribozymes (RNA-enzymes) are likely components[2].

However, large ribozymes, which are known to form self-replicating networks, cannot have selfassembled from the short RNA fragments available in abiotic conditions, and typical short ribozymes tend to favor the cleavage reaction over the required ligation for the formation of larger molecules. A promising direction is to adapt specific environmental conditions (temperature, ions), strand sequence and length, which were shown by experiments to favor ligation e.g. in the small hairpin ribozyme[3], but an understanding of these factors' impact on the catalytic steps is still missing.

Our aim is to provide a molecular understanding of the tertiary structure effect on the ligation/ cleavage equilibrium using all-atom molecular dynamics with enhanced sampling approaches and to clarify the reaction thermodynamics and mechanism for this system using alchemical transformations and mixed quantum-classical approaches.

We will present our first results regarding the identification of relevant conformations in the reactive state, which strongly depends on the force field parametrization and on the specific characteristics of the catalytic site in the initial structures. Specific emphasis was put on estimating the free energy landscape along several key variables, such as the inline attack angle of the cleavage reaction and the hydrogen bonds network. Beyond shedding light on the behavior of the hairpin ribozyme, our findings highlight the crucial importance of using specific enhanced sampling techniques to provide a reliable conformational sampling of the reactant and product states, which is typically not achieved even with microsecond brute force simulations.

References

[1] Leslie E., O. Prebiotic Chemistry and the Origin of the RNA World. Critical Reviews in Biochemistry and Molecular Biology 2004, 39 (2), 99–123.
https://doi.org/10.1080/10409230490460765.

[2] Hayden, E. J.; von Kiedrowski, G.; Lehman, N. Systems Chemistry on Ribozyme Self-Construction: Evidence for Anabolic Autocatalysis in a Recombination Network. Angew. Chem. Int. Ed. 2008, 47 (44), 8424–8428.
https://doi.org/10.1002/anie.200802177.

[3] Fedor, M. J. Structure and Function of the Hairpin Ribozyme. Journal of Molecular Biology 2000, 297 (2), 269–291.
https://doi.org/10.1006/jmbi.2000.3560

THE CONTRIBUTION OF SHORT LINEAR MOTIFS (SLIMS) TO THE MECHANOSTABILITY OF MECHANOSENSITIVE PROTEINS

ISMAHENE MESBAH[a], BIANCA HABERMANN[b], FELIX RICO[c]

[a] Turing Center for Living Systems (CENTURI), Aix Marseille Univ, Parc Scientifique de Luminy, Marseille, France.

[b] Aix Marseille Univ, Developmental Biology Institute of Marseille (IBDM), Turing Center for Living Systems, Parc Scientifique de Luminy, Marseille, France.

[c] Aix Marseille Univ, LAI, U1067 INSERM, Parc Scientifique de Luminy, Marseille France.

Mechanical forces are involved in a variety of biological processes, including protein degradation, cell adhesion, tissue organization, and muscle function. Proteins are the main participants in these events. These nanomachines fold into specific structures that allow them to use mechanical force at cellular, subcellular and molecular level. Therefore, studying the relationship between the structure and mechanics of proteins is important to understanding protein function. While it is known that β-sheet proteins are more mechanically stable than α-helix proteins, we are still not able to predict the unfolding forces from specific protein folds.

Atomic force microscopy (AFM) is an excellent tool to measure forces and determine the strength of protein structural folds. Computational biology, on the other hand, allows predicting structural folds, their associated functions and their evolutionary conservation.

The main objective of the project is to find correlations between protein folds, sequence and structural motifs with their mechanical stability by combining AFM experiments to computational tools. This will allow to establish a link between sequence, structure and mechanical stability of proteins.

Gathering and clustering data from experiments and literature was an important step to start our project. We created MechanoProtein database (MPDB) which stores relevant information about mechanical proteins such as their unfolding pattern, unfolding forces and loading rates. In MPDB, proteins are classified into different categories according to their structural folds.

The 3D comparison of proteins having similar folds using distance matrix (DALI) permits their spatial superimposition which allows extracting sequence alignments based on structural features. We used DALI to compare protein domains including ImmunoGlobulin domain of titin (I27), 4th domain of InterCellular Adhesion Molecule (ICAM1), 2nd domain of Vascular Cellular Adhesion Molecule (VCAM1) and ExtraCellular domain of Cadherin-23 (Cdh23EC1). From the sequence alignment of these domains, sequence patterns can be observed. These are called Short Linear Motifs (SLiMs). For example, we observed the SLiM [NE][LI][KQR]V located in the last beta sheet (G) involved in the unfolding process of I27 of titin1,2. This SLiM is also found in Cdh23EC1 , ICAM1 and VCAM1, all mechanosensitive proteins.

To understand which role each amino acid plays within the motif, we use Steered Molecular Dynamics (SMD) simulations. By fixing the protein from one end (N-ter) and pulling from the other end (C-ter), SMD simulations mimic AFM pulling experiments with the advantage of knowing the position of each atom at each time step which enables us to have an atomistic description of the unfolding process. Preliminary results coughs amino acids in the found SLiM as important for the mechanical stability of the studied domains.

References

[1] Rico, F. et al. (2013) "High-speed force spectroscopy unfolds titin at the velocity of molecular dynamics simulations," Science, 342(6159), pp. 741–743. Available at: https://doi.org/10.1126/science.1239764.

[2] Rief M, Gautel M, Oesterhelt F, Fernandez JM, Gaub HE. Reversible unfolding of individual titin immunoglobulin domains by AFM. Science. 1997 May 16;276(5315):1109-12. doi: 10.1126/science.276.5315.1109. PMID: 9148804

[3] Habermann B. HH-MOTiF: de novo detection of short linear motifs in proteins by Hidden Markov Model comparisons. Nucleic Acids Research, 2017, 45 (W1), pp.W470-W477. 10.1093/nar/gkx341 SMASH. hal-02000485

IDENTIFICATION OF THE TWENTY PROTEINOGENIC AMINO ACIDS USING SOLID-STATE NANOPORES

ANDREINA URQUIOLA HERNÁNDEZ[a], PATRICE DELARUE[a] CHRISTOPHE GUYEUX[b], ADRIEN NICOLAÏ[a*], PATRICK SENET[a]

[a] Laboratoire Interdisciplinaire Carnot de Bourgogne, UMR 6303 CNRS Université de Bourgogne. 9, Av. Savary - B.P. 47 870 21078. Dijon Cedex - France

[b] Institut FEMTO-ST, UMR 6174 CNRS Université de Franche-Comté, Besançon – France

* Corresponding author

Two-dimensional solid-state nanopores (SSNs), including single-layer molybdenum disulfide () nanopores, show immense potential to achieve single-residue resolution on protein detection technology [1], which is one of the field with major impact on medical diagnosis. The detection principle behind SSNs is based on measuring the small fluctuations in ionic current as charged biomolecules, in reaction to an external voltage applied across the membrane, move through the nanopore while immersed in an electrolyte. The fundamental premise of this technology is that the single biomolecule passage through a nanopore will alter the ionic current flowing through the pore in a particular way, enabling the current record to reproduce the protein properties such as its amino acid sequence. In this work, we produce translocation data of biological peptides across single-layer nanopores through extensive unbiased all-atom classical Molecular Dynamics simulations. Each peptide is made of one of the twenty proteinogenic amino acids linked to a charged carrier. Based on ionic current blockade signal processing and clustering techniques, we define relevant information that can lead to discrimination of the twenty proteinogenic amino acids. From MD, we extract the peptide-induced blockade events following the two threshold-method which revealed different levels of blockade current. First, we characterize the sub-populations of blockade ionic current 1D distributions from the parameters of Gaussian Mixture Model (GMM). Second, blockade levels of current have been detected by using structural break detection algorithm and characterized by their duration and depth (2D). These data are then used as input in GMM to estimate the mean values of duration and depth clusters for each amino acid, which are extracted by a repeated process. This analysis reveals visually characteristic regions for each amino acid and showing that charged amino acids present the most well-distinguishable regions. This result is an important step forward to the recognition of protein simplified sequences partitioning amino acids into three groups: positive, negative and neutral amino acids, significantly decreasing the size of protein sequence space. These promising findings may offer a route toward protein sequencing using solid-state nanopores.

References

[1] Barati Farimani, A.; Heiranian, M.; Aluru, N. R. Identification of amino acids with sensitive nanoporous : towards machine learning-based prediction. npj 2D Materials and Applications 2018, 2, 1–9, Number: 1 Publisher: Nature Publishing Group.

[2] A. Nicolaï and P. Senet. Challenges in Protein Sequencing using Nanopores, 2022. In Single Molecule Sensing Beyond Fluorescence. Nanostructure Science and Technology. Springer, Cham.

IN SILICO STUDIES OF THE LRRK2/PP1 INTERACTION IN THE CONTEXT OF PARKINSON'S DISEASE

JOSE LUIS DOMINGUEZ MEIJDE[a], GAUTIER MOROY[a] SAMUEL MURAIL[a], PIERRE TUFFERY[a]

[a] BFA, Université Paris Cite, CNRS UMR 8251, Inserm U1133, Paris, France

Mutations in Human leucine-rich repeat serine/protein phosphatase Kinase 2 (LRRK2) are associated with familial and sporadic Parkinson disease (PD);[1] more specifically, the mutations that cause dephosphorilation of certain key sites could be implied in the pathological mechanism of PD. Serine/Threonine Phosphatase 1 (PP1) is a phosphoprotein phosphatase involved in the regulation of multiple cellular functions, including the desphosphorilation and subsequent inhibition of LRRK2.[2] The LRRK2/PP1 interaction has become a target of interest and we plan to develop peptides interfering with this interaction in order to understand its role in the etiology of PD. Previous research in our group[3] has identified the specific region of LRRK2 that binds to PP1. We employed the LRRK2 sequence identified in the previous experiment (PMGFWSRLINRLLEISPY) to model a peptide that we will use for the next step of our experiments, which is identify the specific region of PP1 that interacts with LRRK2, and identify the key modes of interaction between these two regions.

In order to identify the modes of interaction, we took the LRRK2-based peptide and used it in combination with the catalytic unit of PP1 in docking and Simulated Tempering simulations. These in-silico experiments explored multiple angles of approach between between the two partners. The results we obtained from these simulations showed two main forms of interaction between the LRRK2-based peptide and PP1, both based on salt bridges, but each revolving around one respective key residue of the LRRK2-based peptide: the first complex configuration showed the LRRK2-peptide R7 as the main interactor (with a 37% time fraction of contact in the trajectory), whereas the second complex revolved around E14 (with a 29% time fraction of contact in the trajectory).

With this information in hand, we tested the stability of the two different interaction configurations. We subjected these two main interaction configurations to mutations in the key interaction amino acids of the LRRK2-based peptide. We then proceeded with new Simulated Tempering simulations to test the stability of this poses when the key amino acids are replaced, in order to determine which of these two amino acids has a higher impact on the integrity of the complex.

Pending experimental confirmation, our preliminary results indicate that substituting R7 causes the complex to break faster, and is therefore is the most influential residue in the contact infrastructure of the LRRK2-based peptide and PP1.

References

[1] Li, JQ., Tan, L. & Yu, JT. The role of the LRRK2 gene in Parkinsonism. Mol Neurodegeneration 9, 47 (2014). https://doi.org/10.1186/1750-1326-9-47

[2] Lobbestael E, Zhao J, Rudenko IN, Beylina A, Gao F, Wetter J, Beullens M, Bollen M, Cookson MR, Baekelandt V, Nichols RJ, Taymans JM. Identification of protein phosphatase 1 as a regulator of the LRRK2 phosphorylation cycle. Biochem J. 2013 Nov 15;456(1):119-28. doi: 10.1042/BJ20121772. PMID: 23937259; PMCID: PMC5141581.

[3] Dong CZ, Bruzzoni-Giovanelli H, Yu Y, Dorgham K, Parizot C, Zini JM, Brossas JY, Tuffery P^, Rebollo A. (2020) Identification of peptides interfering with the LRRK2/PP1 interaction. PLOS ONE 15(8): e0237110. https://doi.org/10.1371/journal.pone.0237110

TRACKING LIGAND-BINDING EFFECT ON PROTEIN STABILITY BY CD SPECTROSCOPY AND MOLECULAR DYNAMICS SIMULATIONS

JEREMIE TOPIN[a], MAXENCE LALIS[a], NYKOLA C. JONES[b]
SØREN V. HOFFMANN[b], LOIC BRIAND[c], CORNELIA MEINERT[a]

[a] Institut de Chimie de Nice, CNRS UMR 7272, Université Côte d'Azur, Nice, France

[b] ISA, Department of Physics and Astronomy, Aarhus University, 8000 Aarhus C, Denmark

[c] INRAE, CNRS, Université de Bourgogne-Franche Comté, AgroSup Dijon, Centre des Sciences du Goût et de l'Alimentation, 21000 Dijon, France

The influence of protein-ligand interactions on protein stability is usually assessed by measurements in the liquid phase. CD spectroscopy appears to be a tool of choice to i) measure the conformation of the protein in different phases and to ii) follow the conformational changes of the protein upon binding.

We therefore studied the stability of the rat odorant binding protein 3 (OBP3), its ability to remain functional and a preliminary test of its ligand binding specificity in the dry state. Solid-state spectra were performed on dry thin films prepared by drop casting of initial buffered aqueous solutions of the rat OBP3 onto optically transparent CaF2 windows and subsequently dried under low vacuum.

We successfully recorded CD spectra of solid-state apo OBP3 from 280 to 130 nm. A previously unknown positive dichroic band became measurable in the solvent-free state at 175 nm. The reproducibility of the solid-state CD spectrum of apo OBP3 was confirmed by measuring several individually prepared films. We then assessed the time-dependent alteration of the protein in this dry environment. No change in the spectra was observed (storage at constant humidity with binary saturated salt solution), highlighting the stability of the OBP films on a monthly basis. Our results revealed that protein folding is not affected during film formation and remains stable over long-time scales.

Finally, molecular dynamics simulations revealed that the flexibility of the protein is affected when the surrounding water molecule network decreases. The analysis of unconstraint molecular dynamics simulations revealed a conserved pathway for the ligand entry. Mutations of the identified residues abolish the binding of several odorants, confirming the *in silico* results.

References

[1] Li, JQ., Tan, L. & Yu, JT. The role of the LRRK2 gene in Parkinsonism. Mol Neurodegeneration 9, 47 (2014). https://doi.org/10.1186/1750-1326-9-47

[2] Lobbestael E, Zhao J, Rudenko IN, Beylina A, Gao F, Wetter J, Beullens M, Bollen M, Cookson MR, Baekelandt V, Nichols RJ, Taymans JM. Identification of protein phosphatase 1 as a regulator of the LRRK2 phosphorylation cycle. Biochem J. 2013 Nov 15;456(1):119-28. doi: 10.1042/BJ20121772. PMID: 23937259; PMCID: PMC5141581.

[3] Dong CZ, Bruzzoni-Giovanelli H, Yu Y, Dorgham K, Parizot C, Zini JM, Brossas JY, Tuffery P^, Rebollo A. (2020) Identification of peptides interfering with the LRRK2/PP1 interaction. PLOS ONE 15(8): e0237110. https://doi.org/10.1371/journal.pone.0237110

UNDERSTANDING THE INTERACTION MECHANISM BETWEEN CASPASE3 AND LINKER-BIR2 OF XIAP IN PRESENCE AND ABSENCE OF A SMAC MIMETIC

MARC RAGUI[a*], CHARLINE KIEFFER[a], ANNE-SOPHIE VOISIN-CHIRET[a] JANA SOPKOVA-DE OLIVIERA SANTOS[a]

[a] Normandie Univ, UNICAEN, CERMN, EA 4258, Boulevard Becquerel 14032 Caen Cedex, France.

XIAP (X-linked chromosome) is one of the human inhibitory apoptosis protein family considered as key regulators of cell death (apoptosis).[1] XIAP is known for its ability to bind caspases enzymes (cell apoptosis initiators and effectors) released from mitochondria, through its baculoviral IAP repeats domains (BIR). Also, it contains an ubiquitin-associated domain (UBA) and a really interesting new gene with a ligase activity (RING) domain. XIAP/caspases binding mechanism results in promoting cell survival. This mechanism is balanced by the action of an endogenous antagonist, the second mitochondrial activator of caspases (SMAC/DIABLO). SMAC binds to XIAP BIR domains through its N-*terminal* tetrapeptide 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]

SMAC binding to XIAP-BIR2 domain inhibits XIAP interaction with caspases 3/7. Caspases 3/7 bind more strongly to the linker between BIR1 and BIR2 domains than XIAP-BIR2 domain.[3] The monomeric form of caspase 3/7 (two protein subunits) complexed with linker-BIR2 is unable by itself to elucidate this cross interaction. A dimeric complex is indeed required to well understand the whole process as described in some studies.[4] They suggest that the small subunits of caspase 3/7 of each

monomer are interacting in parallel with different parts of XIAP. The first small subunit interacts with the linker, while the other one interacts with the BIR2 domain. This binding mechanism is not very well illustrated. Nevertheless, the comprehension of this mechanism is a key element to design small non-peptide molecules that could reactivate apoptosis effector caspases 3/7.

Replicas of MD simulations of the dimeric system (linker-BIR2/Caspase 3)$_2$ were carried out in absence and presence of a SMAC mimetic ligand in the literature[5]. Throughout these simulations, we highlighted the conformational change induced in the linker/caspase 3 binding following the interaction of the SMAC mimetic ligand with the XIAP-BIR2 domain, and hence leading to liberation of the caspase.

In conclusion, our results have revealed the conformational change in the hydrophobic cluster formed by the caspase small subunit interacting with the linker. Fortunately, this goes along with biological findings, considering this change as a major determinant for the interaction.[6]

References

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

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

[3] Scott, F. L. et al. XIAP inhibits caspase-3 and -7 using two binding sites: evolutionarily conserved mechanism of IAPs. EMBO J. 24, 645–655 (2005).

[4] Riedl, S. J. et al. Structural Basis for the Inhibition of Caspase-3 by XIAP. Cell 104, 791–800 (2001).

[5] Kester, R. F. et al. Optimization of Benzodiazepinones as Selective Inhibitors of the X-Linked Inhibitor of Apoptosis Protein (XIAP) Second Baculovirus IAP Repeat (BIR2) Domain. J. Med. Chem. 56, 7788–7803 (2013).

[6] Huang, Y. et al. Structural basis of caspase inhibition by XIAP: differential roles of the linker versus the BIR domain. Cell 104, 781–790 (2001).

EVOLUTIONAL AND STRUCTURAL ANALYSIS OF THE INTERLEUKIN-3 INTERACTION WITH ITS ALPHA SUBUNIT RECEPTOR IN THE CONTEXT OF ATOPIC DISEASES

JADE FOGHA[a], ALEXANDRE G. DE BREVERN[a]
JULIEN DIHARCE*[a]

[a] Biologie intégrée du Globule rouge, UMR S_1134 Inserm Université Paris Cité, Hôpital Broussais, 8 Rue Maria Helena Vieira Da Silva, 75014 Paris, France

Protein-protein interactions (PPIs) are attractive targets for drugs development as they are critical in a variety of biological processes and pathologies. As an illustration, the interleukin 3 (IL-3) and its α subunit receptor (IL-3Rα) are two proteins belonging to the cytokine (or receptor βc) family, which comprises also IL-5 and GM-CSF. Those proteins are characterized by a two-subunit receptor (subunits α and β) which binds sequentially to the protein (first the α subunit and second the βc). Moreover, they are involved in several disorders like inflammatory diseases or hematological malignancies but IL-3 is specifically involved in atopic diseases, such as allergy. The IL-3/IL-3Rα exhibits a low binding affinity and a complex formed by a mutant form of IL-3 (superkine) and IL-3Rα have emerged from the literature, with an increase of the affinity. The BASIN ANR project is therefore centered around the study of this complex, using experimental and theoretical approaches, in order to decipher in the most precise way its behavior and to eventually propose small compounds of interest with therapeutic potential in the case of atopic diseases. Here, we performed a structural, evolutional and dynamical analyses on this system, more precisely on the interface of interaction. Multiple sequence alignments with several species and molecular dynamics simulations on the human complex have been realized in this context. The evolutionary analysis made on the alignment reveals that the interaction interface is less conserved than the rests of the two proteins. Especially, it appears that the IL-3/IL-3Rα mouse interface is quite different than the human interface.[1] Regarding the dynamical behavior, we observe during long timescale simulations that three different reference conformations are sampled including two already described in the literature and one entirely new which could be associated to a new free energy basin. Finally, since these two proteins are glycosylated, impact of the glycosylation was investigated through MD simulations, but the behavior between the glycosylated and non-glycosylated systems is almost identical, revealing a modest impact of the glycosylation of this system. All those findings will be quite important for the drug design phase of our research project.

References

[1] Fogha, J., Bayry, J., Diharce, J and de Brevern A.G. Structural and evolutionary exploration of the IL-3 family and its alpha subunit receptors. Amino Acids, 2021, 53, 1211–1227.

[2] Fogha J., de Brevern A.G. and Diharce J. Conformational and binding mode assessment of the human IL-3 recognition by its alpha receptor, Proteins: structure, function and Bioinformatics, 2023, 1-16.

MOLECULAR DYNAMICS BASED PREDICTION OF FZO1 TRANSMEMBRANE DOMAINS STRUCTURE

RAPHAËLLE VERSINI [a,b] , ANTOINE TALY [a], PATRICK FUCHS [b]

[a] Laboratoire de Biochimie Théorique, Institue de Biologie Physico-Chimique, CNRS, 13 rue Pierre et Marie Curie, 75005, Paris
[b] Laboratoire des Biomolécules, Sorbonne Université, 4 place Jussieu, 75005, Paris

Outer mitochondrial membrane (OMM) fusion is an important process for the cell and organism survival, as its dysfonction is linked to neurodegenerative diseases [1, 2] and cancer [3]. The OMM fusion is mediated by members of the dynamin-related protein (DRP) family, named mitofusins [4]. The exact mechanism by which the mitofusins contributes to these deseases, as well as the exact molecular fusion mechanism mediated by mitofusin, remains elusive. Fzo1, the only mitofusin homologue of the yeast Saccharomyces cerevisiae and embedded in the OMM, was modeled by homology [5] with the mitofusin related bacterial dynamin-like protein (BDLP) as template [6]. 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. The previous prediction [5] used the webserver PREDIMMER [7] to predict the structure of the transmembrane domain. Furthermore, TM1 has a lysine (Lys716) located inside the membrane which could either be protonated or unprotonated. The first construction assumed this lysine to be neutral.

The project consist of solving the structure of the two transmembrane domains of the protein Fzo1, using a multiscale molecular modeling approach, in order to improve the prediction of the structure of TM1 and TM2 with physics based methods. Coarsegrained representations, with the force field MARTINI3, are used in order to sample a massive amount of conformations. We found that the neutral state of the LYS is preferable in the membrane. The most frequent contacts were also studied, and we found that the GX3G motif was often involved. A model was extracted, backmapped into an all-atom representation and it's robustness was tested using an all-atom Temperature-Replica Exchange Molecular Dynamics (T-REMD) method simulations. The latter simulations allowed us to produce a refined version of the model of Fzo1.

References

[1] Hsiuchen Chen and David C. Chan. Mitochondrial dynamics–fusion, fission, movement, and mitophagy–in neurodegenerative diseases. Hum Mol Genet, 18:R169–R176, 2009.

[2] Andrew B. Knott, Guy Perkins, Robert Schwarzenbacher, and Ella Bossy-Wetzel. Mitochondrial fragmentation in neurodegeneration. Nature Reviews Neuroscience, 9(7):505–518, 2008.

[3] Keshav K Singh and Josephine S. Modica-Napolitano. Special issue: Mitochondria in cancer. Seminars in Cancer Biology, 47:iv–vi, 2017. Mitochondria in Cancer.

[4] Rajesh Ramachandran. Mitochondrial dynamics: The dynamin superfamily and execution by collusion. Semin. Cell Dev. Biol., 76:201–212, 2018.

[5] Dario De Vecchis, Laetitia Cavellini, Marc Baaden, Jérôme Hénin, Mickaël M. Cohen, and Antoine Taly. A membrane-inserted structural model of the yeast mitofusin fzo1. Scientific Reports, 7(1):10217, 2017.

[6] Harry H. Low and Jan Löwe. A bacterial dynamin-like protein. Nature, 444:766–769, 2006.

[7] Anton A. Polyansky, Anton O. Chugunov, Pavel E. Volynsky, Nikolay A. Krylov, Dmitry E. Nolde, and Roman G. Efremov. Preddimer: a web server for prediction of transmemrane helical dimers. Bioinformatics, 30:889–890, 2014.

STRUCTURAL AND DYNAMIC DIFFERENCES BETWEEN CALRETICULIN MUTANTS ASSOCIATED WITH ESSENTIAL THROMBOCYTHEMIA

RAGOUSANDIRANE RADJASANDIRANE[a]
ALEXANDRE G. DE BREVERN[b]

[a] Université Paris Cité, INSERM, BIGR, DSIMB Bioinformatics team, Paris, France
[b] Université Paris Cité, INSERM, BIGR, DSIMB Bioinformatics team, Paris, France

Essential thrombocythemia (ET) is a blood cancer. It is characterized by an overproduction of blood platelets and can induce deadly thrombosis. ET was already associated with mutations of JAK2[1] and MPL[2] proteins in megakaryocytes. MPL is a receptor that triggers the JAK2 signaling pathway when activated.

A third protein was characterized in 2013, namely Calreticulin wild-type (CALRwt)[3]. CALR is found in the endoplasmic reticulum (ER) and contains a KDEL motif which is an ER retention motif. Naturally, CALR can't activate the MPL receptor that is outside the megakaryocyte to induce platelet proliferation. However, in 25-30% of patients with ET, the CALR protein has mutations that cause a frameshift; it changes half of the sequence of its last domain (named C-domain) and loses the KDEL motif, i.e. CALR cannot stay in the ER. Moreover, these new variant sequences contain cysteines that give CALR mutants (CALRm) the ability to dimerize. Therefore, CALRm exits the megakaryocytes and forms a homodimer that activates platelet production in an uncontrolled manner by binding the MPL receptor.

We previously proposed a novel classification in five classes (A, B, C, D and E) of these variants that is segmental and allows a better vision of the CALRm sequences[4]. However, among these new classes, some seem unable to induce the disease according to their sequences. Hence, class E and D are devoid of cysteines, contrary to classes A, B and C. Indeed, C-domains are supposed to dimerize through interchain disulphide bonds[5].

The aim of this study is to analyze the oncogenicity of each class in an ET context based on the analysis of the last domain of CALR that is affected by the frameshift. We analyzed the dynamics of each CALRm class (and CARLwt) in both monomeric and dimeric forms to correctly classify these variants in terms of dynamic properties. Our results show that the dynamic properties of classes A, B and C are different but share characteristics such as the ability to form a dimer. Class E and D can't form dimers since they don't have cysteines, and even by forcing two monomers by distance constraint to form a dimer, the dimer chains instantly repulse each other.

The conclusion is that classes A, B and C could induce ET, but only as they share new Cysteines. Classes E and D are similar to CALRwt, and cannot provide a dimerization. Therefore we propose the hypothesis that they are probably not involved in the disease, and class E represents only simple human polymorphism[6].

References

[1] Chuzi, S.; Stein, B.L. Essential thrombocythemia: A review of the clinical features, diagnostic challenges, and treatment modalities in the era of molecular discovery. Leuk. Lymphoma 2017, 58, 2786–2798.

[2] James, C.; Ugo, V.; Le Couédic, J.-P.; Staerk, J.; Delhommeau, F.; Lacout, C.; Garçon, L.; Raslova, H.; Berger, R.; Bennaceur-Griscelli, A.; et al. A unique clonal JAK2 mutation leading to constitutive signalling causes polycythaemia vera. Nature 2005, 434, 1144–1148.

[3] Klampfl, T.; Gisslinger, H.; Harutyunyan, A.S.; Nivarthi, H.; Rumi, E.; Milosevic, J.D.; Them, N.C.C.; Berg, T.; Gisslinger, B.; Pietra, D.; et al. Somatic Mutations of Calreticulin in Myeloproliferative Neoplasms. N. Engl. J. Med. 2013, 369, 2379–2390

[4] El Jahrani, N.; Cretin, G.; de Brevern, A.G. CALR-ETdb, the database of calreticulin variants diversity in essential thrombocythemia. Platelets 2022, 33, 157–167.

[5] Venkatesan, A.; Geng, J.; Kandarpa, M.; Wijeyesakere, S.; Bhide, A.; Talpaz, M.; Pogozheva, I.; Raghavan, M. MPN-274: Mpl is Activated by Dimers of MPN-Linked Calreticulin Mutants Stabilized by Disulfide Bonds and Ionic Interactions. Clin. Lymphoma Myeloma Leuk. 2020, 20, S335.

[6] Radjasandirane, R.; de Brevern, A.G. Structural and Dynamic Differences between Calreticulin Mutants Associated with Essential Thrombocythemia. Biomolecules 2023, 13, 509.

DECIPHERING THE STRUCTURAL DEFECTS FROM SNP VARIANTS OF ABCB4 TRANSPORTER BY MEANS OF MOLECULAR DYNAMICS SIMULATIONS

VERONICA CRESPI[a], ÁGOTA TÓTH[a], ANGELIKA JANASZKIEWICZ[a] THOMAS FALGUIÈRES[b], FLORENT DI MEO[a]

[a] InsermU1248 Pharmacology & Transplantation, University of Limoges, Limoges, France
[b] Inserm U1193 Physiopathogénèse et traitement des maladies du foie, Université Paris-Saclay, Orsay, France
veronica.crespi@inserm.fr, florent.di-meo@inserm.fr

Bile secretion is an essential function of the liver for the elimination of xenobiotics and endogenous metabolites. This function mostly relies on membrane transporters located at the canalicular membrane, such as the ATP-binding cassette (ABC) transporter ABCB4. Variations in the ABCB4 gene were reported with rare liver diseases. Most of them are missense variations that were associated with the modulation of protein expression, intracellular trafficking, or ABCB4 secretion activity[1]. In this context, the present study aims to provide insights into structural defects of selected ABCB4 SNP mutants by means of molecular dynamics (MD) simulations, supported by cellular and molecular biology experiments.

Taking advantage of the recent cryo-EM resolutions[2,3] of ABCB4 transporter, ms-scaled MD simulations were performed considering (i) different bound states (i.e., ATP- and/or PC substrate ABCB4), and (ii) two lipid bilayer models (namely, symmetric and asymmetric). The overall ABCB4 dynamics was monitored to decipher key structural features along the transport cycle. Furthermore, local structural impact arising from SNP mutations were assessed by mutating WT-ABCB4 using alchemical transformation. Particular attention was paid to the local structural impact of I490T mutant in order to provide structural hints associated with functional impairments.

References

[1] Jean�Louis Delaunay et al., "A Functional Classification of ABCB4 Variations Causing Progressive Familial Intrahepatic Cholestasis Type 3," Hepatology 63, no. 5 (May 2016): 1620–31, https://doi.org/10.1002/hep.28300.

[2] Kamil Nosol et al., "Structures of ABCB4 Provide Insight into Phosphatidylcholine Translocation," Proceedings of the National Academy of Sciences 118, no. 33 (August 17, 2021): e2106702118, https://doi.org/10.1073/pnas.2106702118.

[3] Jeppe A. Olsen et al., "Structure of the Human Lipid Exporter ABCB4 in a Lipid Environment," Nature Structural & Molecular Biology 27, no. 1 (January 2020): 62–70, https://doi.org/10.1038/s41594-019-0354-3.

INVESTIGATION OF A NOVEL PRESSURE-SENSITIZED TACAN STRUCTURE UNDER HIGH-THRESHOLD SYSTEM PRESSURE

HUIXIA LU[a,b,c], JORDI FARAUDO[a], JORDI MARTI[b], BUYONG MA[c]

[a] Institut de Ciencia de Materials de Barcelona (ICMAB-CSIC), Campus de la UAB, Bellaterra, Barcelona E-08193, Spain

[b] Department of Physics, Technical University of Catalonia-Barcelona Tech (UPC), B4-B5 Campus Nord, Jordi Girona 1-3, 08034 Barcelona, Catalonia, Spain

[c] Department of Pharmacy, Shanghai Jiao Tong University, Dongchuan Road No. 800, Shanghai, China

TACAN has been shown to contribute to the mechanical hyperalgesia with the mechanism unknown. In 2020, Lou Beaulieu-Laroche et al. proposed that TACAN may function as a high threshold mechanically activated ion channel involved in sensing mechanical pain[1]. However, other research groups deny the potential role that TACAN may play in the mechanical signaling path cross the cell membrane. Taken all together, despite the previous report that suggested TACAN might function as a high-threshold mechanically activated cation channel responsible for sensing mechanical pain in mice, the mechanism of mechanosensation and ion permeation mediated by TACAN channel remains elusive and awaits further investigation. In this study, we show a novel human TACAN dimer structure. It is a homodimer with each monomer consisting of a body, a spring and a blade domains. More importantly, we found that all the helices of the body and the spring domains are specifically associated with 11 membrane lipid molecules. Particularly, a lipid core, residing within a cavity formed by the two body and spring domains, contacts with the helices from the body and spring domains and extends to reach two symmetrically arranged lipid clusters. Actually, a previous study showed that the membrane lipids have the intrinsic feature leading to the possibility of soliton propagation in nerves[2]. Therefore, our results shall suggest that the cryo-EM structure of TACAN associated with lipids is novel and the membrane lipids may play an important biological role in tunning the function of TACAN upon mechanical (eg. pressure) stimuli. TACAN may use membrane lipids to directly sense and transduce the mechanic stimulus, whereas the protein part of TACAN may function to modulate the action of lipids. Molecular dynamics simulations suggest that TACAN isn't an ion channel, but signal transduction along TACAN tends to be gated by the membrane cardiolipin lipids upon high-threshold pressure stimuli upon upstream mechanical force.

References

[1] Beaulieu-Laroche, Lou, et al. «TACAN is an ion channel involved in sensing mechanical pain.» Cell 180.5 (2020): 956-967.

[2] Heimburg T, Jackson AD. On soliton propagation in biomembranes and nerves. Proceedings of the National Academy of Sciences of the United States of America 102, 9790-9795 (2005).

UNRAVELLING WATER TRANSPORT PATHS IN HALOALKANE DEHALOGENASE AND EVALUATING THE ROLE OF WATER MODELS IN TRANSPORT EFFICIENCY

ARAVIND SELVARAM THIRUNAVUKARASU[a,b]
CARLOS EDUARDO SEQUEIROS BORJA[a,b]
CEDRIX J DONGMO FOUMTHIUM[a,b], JAN BREZOVSKY[a,b]

[a] Laboratory of Biomolecular Interactions and Transport, Institute of Molecular Biology and Biotechnology, Faculty of Biology, Adam Mickiewicz University, Poland

[b] International Institute of Molecular and Cell Biology in Warsaw, 02-109 Warsaw, Poland

Background: Enzyme active sites can be buried deep within the core of the protein[1]. The buried active sites are accessible through tunnels, which allow access of ligands, solvents, and ions [2]. Molecular dynamics simulations are the preferred method for studying tunnel dynamics and molecular transport within proteins, as their viability often change notably with a protein conformation [3]. DhaA, an environmentally relevant enzyme with buried active site that uses water as a major catalytic component, was selected as model system due to its well-studied nature [4].

Objectives: i) to establish a method for routine analysis of transport through tunnels, ii) to characterize the tunnels and their use by water molecules in DhaA, and iii) to evaluate the effects of different water models on probability of their transport via tunnels

Methods: The starting structure was obtained from the RCSB PDB[5] database, and the simulations were performed using Amber software[6]. A total simulation time of 5 µs was achieved with adaptive sampling (HTMD[7]). The tunnels were then calculated using CAVER 3.0[8] and the movement of the waters was tracked using the AQUA-DUCT tool[9]. The migration of water molecules through the tunnels was analyzed using TransportTools[10]. The effects of different water models on the tunnels and their transport were evaluated through Comparative analysis module of TransportTools, on simulations performed using different tunnel geometries extracted from the adaptive sampling simulation with protein backbone restrained.

Results: The developed method enabled us to efficiently identify several ensembles of tunnels used by waters molecules to reach the active site of DhaA. Among those, all well-established tunnels were present but also novel transient tunnels capable of water transport were detected. Curiously, about 20 % of the water transport occurred through the tunnels with radii below 1.4 Å but enriched in hydrogen bonding facilities. We have established that the water model used in the simulation markedly affects the volume of waters entering the protein.

Conclusions: A methodology was developed to routinely analyze the transport paths used in proteins. We have observed that DhaA has multiple tunnels through which water is transported, including those with unexpectedly low radii. Importantly, the water model used in simulations affects the rates of transport via those paths making the selection of an appropriate model critical for accurate inference into transport kinetics.

This research was supported by POWR.03.02.00-00-I006/17 project. The calculations were performed at the Poznan Supercomputing and Networking Center.

References

[1] S. M. Marques, L. Daniel, T. Buryska, Z. Prokop, J. Brezovsky, and J. Damborsky, "Enzyme Tunnels and Gates As Relevant Targets in Drug Design," Med Res Rev, vol. 37, no. 5, pp. 1095–1139, Sep. 2017, doi: 10.1002/MED.21430.
[2] M. Krone et al., "Visual Analysis of Biomolecular Cavities: State of the Art," Computer Graphics Forum, vol. 35, no. 3, pp. 527–551, Jun. 2016, doi: 10.1111/CGF.12928.
[3] V. Vonásek, A. Jurčík, K. Furmanová, and B. Kozlíková, "Sampling-Based Motion Planning for Tracking Evolution of Dynamic Tunnels in Molecular Dynamics Simulations," Journal of Intelligent and Robotic Systems: Theory and Applications, vol. 93, no. 3–4, pp. 763–785, Mar. 2019, doi: 10.1007/S10846-018-0877-6/METRICS.
[4] M. Klvana et al., "Pathways and mechanisms for product release in the engineered haloalkane dehalogenases explored using classical and random acceleration molecular dynamics simulations," J Mol Biol, vol. 392, no. 5, pp. 1339–1356, Oct. 2009, doi: 10.1016/J.JMB.2009.06.076.
[5] H. M. Berman et al., "The Protein Data Bank," Nucleic Acids Res, vol. 28, no. 1, pp. 235–242, Jan. 2000, doi: 10.1093/NAR/28.1.235.
[6] R. Salomon-Ferrer, A. W. Götz, D. Poole, S. le Grand, and R. C. Walker, "Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald," J Chem Theory Comput, vol. 9, no. 9, pp. 3878–3888, Sep. 2013, doi: 10.1021/CT400314Y.
[7] S. Doerr, M. J. Harvey, F. Noé, and G. de Fabritiis, "HTMD: High-Throughput Molecular Dynamics for Molecular Discovery," J Chem Theory Comput, vol. 12, no. 4, pp. 1845–
1852, Apr. 2016, doi: 10.1021/ACS.JCTC.6B00049/ASSET/IMAGES/LARGE/CT-2016-00049E_0005.JPEG.
[8] E. Chovancova et al., "CAVER 3.0: a tool for the analysis of transport pathways in dynamic protein structures," PLoS Comput Biol, vol. 8, no. 10, 2012, doi: 10.1371/JOURNAL.PCBI.1002708.
[9] T. Magdziarz et al., "AQUA-DUCT 1.0: structural and functional analysis of macromolecules from an intramolecular voids perspective," Bioinformatics, vol. 36, no. 8, pp. 2599–2601, Apr. 2020, doi: 10.1093/BIOINFORMATICS/BTZ946.
[10] J. Brezovsky et al., "TransportTools: a library for high-throughput analyses of internal voids in biomolecules and ligand transport through them," Bioinformatics, vol. 38, no. 6, pp. 1752–1753, Mar. 2022, doi: 10.1093/BIOINFORMATICS/BTAB872.

MULTISCALE SIMULATIONS OF MOLECULAR RECOGNITION BY PHASE SEPARATED MUT-16: A SCAFFOLDING PROTEIN OF MUTATOR FOCI

KUMAR GAURAV[a,b,c], RENÉ KETTING[b], LUKAS STELZL[a,b,c]

[a] Faculty of Biology, Johannes Gutenberg University Mainz, Germany
[b] Insitute of Molecular Biology (IMB), Mainz
[c] KOMET1, Institute of Physics, Johannes Gutenberg University Mainz, Germany

The RNA silencing pathway is a crucial biological process that regulates gene expression and protects the genome from foreign nucleic acids. In the nematode \textit{C. elegans}, this pathway involves small RNA amplification mediated by RNAdependent RNA polymerases (RdRPs) and the formation of perinuclear germline foci called Mutator foci. The Mutator foci is a biomolecular condensate that arises from liquid-liquid phase separation and plays a critical role in the efficient processing and silencing of target RNAs. Several Mutator complex proteins, including Mut-16 and MUT-7, localize to these foci and facilitate small RNA amplification.[1] Mut-16 acts as a scaffolding protein to recruit other Mutator complex components. Recent studies have identified the protein Rde-2, which is recruited by Mut-16 and has a prion-like N-terminal domain that may interact with the Mut-16 condensate. The C-terminal domain of Rde-2 is globular and recruits the exonuclease (Mut-7).[2] In this study, we performed the multi-scale molecular dynamic simulations of the Rde-2/ Mut-16 interaction that revealed the particular residues and type of interaction responsible for the recognition of Rde-2 to Mutator foci. The simulations were performed in three resolutions: residue-level coarse-grained simulation (HPS/CALVADOS2 model), near-atomic coarse-grained simulations (Martini3), and atomistic simulation. The results showed that the recognition of Rde-2 to Mutator foci was facilitated by cationpi interaction and are possibly sensitive to the protein's post-translational modifications. The elucidation of the molecular details of protein recruitment in mutator foci will deepen our understanding of the RNA silencing pathway and its regulatory role in gene expression.

References

[1] Phillips CM, Montgomery TA, Breen PC, Ruvkun G. MUT-16 promotes formation of perinuclear mutator foci required for RNA silencing in the C. elegans germline. Genes Dev. 2012 Jul 1;26(13):1433-44.

[2] Uebel CJ, Anderson DC, Mandarino LM, Manage KI, Aynaszyan S, Phillips CM. Distinct regions of the intrinsically disordered protein MUT-16 mediate assembly of a small RNA amplification complex and promote phase separation of Mutator foci. PLoS Genet. 2018 Jul 23;14(7):e1007542.

USING MODELLING AND SIMULATION TO ACCOUNT FOR MAGNETIC-DRIVEN SELF-ASSEMBLY OF NANOPARTICLES

NATHALIE SAOULI[a], HICHEM DAMMAK[a,b], ANDREA COSOLA[a] MARC HAYOUN[a], GIANCARLO RIZZA[a]

[a] FLaboratoire des Solides Irradiés (LSI), Institut Polytechnique de Paris, CEA/DRF/IRAMIS, CNRS, Ecole polytechnique, Route de Saclay, 91128 Palaiseau, France

[b] Laboratoire Structures Propriétés et Modélisation des Solides, CentraleSupélec, CNRS, Université Paris-Saclay, F 91190 Gif-sur-Yvette, France

Due to their widespread applications, ranging from diagnosis and drug delivery to actuators, and sensors, there is a growing interest in mastering the behaviour of magnetic nanoparticle assemblies under the influence of external magnetic fields [1-4].

Here, we present the first observations of in situ X-ray microscopy showing the dynamical behaviour in liquid phase of magnetic aggregates upon application of a magnetic field. The varieties of, sometimes unexpected, observations have been modelled and simulated using a Molecular Dynamics approach. Our interpretation allowed to validate the experimental results and it pushes forward the understanding of the magnetic-driven self-assembly processes under dynamical conditions.

From an operative point of view, a relaxed state of clusters with different shapes and sizes is first generated, wherein each cluster is composed of an assembly of nanoparticles, modelled as spheres with a diameter of 100 nm. Particles are considered as ferrimagnetic (Fe3O4) and characterised by a magnetic dipole to enable dipole-dipole interactions. Van der Waals interactions between first neighbours are incorporated in the model to favour the aggregation of particles into clusters and repulsive Lennard Jones potential to prevent their overlapping. Newton's equations of motion are implemented in a Verlet algorithm to compute both position and velocity of the ensemble of particles for each simulation step. Additionally, magnetic moments are evaluated inside a loop of convergence where dipoles are allowed to rotate to align along the local field, i.e. to reach their lowest magnetic energy. Finally, a damping method is used to quickly drive the system towards a stable configuration.

The evolution of several clusters of nanoparticles have been followed by varying both the intensity and the direction of the applied magnetic field. Simulations indicate that the clusters elongate up to a critical field value, above which a decrease of their length is observed. Besides, the separation of large clusters into sub-structures and their rotation is observed in regions presenting a lower density of nanoparticles, i.e. cluster's fragile zones. We demonstrated that these fragile zones play a key role in the dynamical evolution and stability of the clusters upon the applied field.

Finally, numerical simulations are in excellent agreement with the experimental observations and additional simulations are currently being performed in order to deeply dive into the mechanism driving this behaviour of the nanoparticles.

References

[1] R. A. Revia and M. Zhang, "Magnetite nanoparticles for cancer diagnosis, treatment, and treatment monitoring: recent advances," Materials Today, vol. 19, no. 3, pp. 157–168, Apr. 2016, doi: 10.1016/j.mattod.2015.08.022.

[2] L. Shen, B. Li, and Y. Qiao, "Fe3O4 Nanoparticles in Targeted Drug/Gene Delivery Systems," Materials, vol. 11, no. 2, p. 324, Feb. 2018, doi: 10.3390/ma11020324.

[3] J. Yuan, Y. Xu, and A. H. E. Müller, "One-dimensional magnetic inorganic–organic hybrid nanomaterials," Chemical Society Reviews, vol. 40, no. 2, p. 640, 2011, doi: 10.1039/c0cs00087f.

[4] M. V. Rastei et al., "Defect-Driven Magnetization Configuration of Isolated Linear Assemblies of Iron Oxide Nanoparticles," Advanced Functional Materials, vol. 29, no. 45, p. 1903927, Sep. 2019, doi: 10.1002/adfm.201903927

ATLAS: LARGE-SCALE ANALYSIS OF PROTEIN FLEXIBILITY THROUGH MOLECULAR DYNAMICS SIMULATIONS

VANDER MEERSCHE YANN[a,b], CRETIN GABRIEL[a,b], GELLY JEAN-CHRISTOPHE[a,b], GALOCHKINA TATIANA[a,b]

[a] Université Paris Cité and Université des Antilles and Université de la Réunion, INSERM, BIGR, F-75014 Paris, France

[b] Laboratoire d'Excellence GR-Ex, 75014 Paris, France

Despite the disruption caused by Alphafold2 release in the field of structural bioinformatics, the problem of the analysis and prediction of protein dynamic properties still remains one of the most important challenges. In our recent work [1], we have developed a Deep Learning based prediction tool of protein flexibility in terms of B-factor classes. However, the precision of the predictions is limited by the experimental variability of the protein flexibility measurements, mainly B-factors in this case [2]. To overcome this issue we decided to use molecular dynamics simulations following a standardized protocol to obtain diverse and precise flexibility descriptors well suited for machine learning. We therefore created ATLAS, a database of molecular dynamics simulation on a large and representative set of protein structures, covering more than 84% of the proteins with known ECOD [3] class X (possible homology). ATLAS contains the dynamics of 1,000 proteins organized in overlaping datasets, each non-redundant in terms of ECOD domains Xclass: mono-domain proteins filtered at 0.5 and at 0.6 of TM-score to obtain dynamical properties not biased by potential partner proteins, as well as chains coming from multi-chains complexes to sample even more protein folds. All these dynamics have been carefully analyzed, manually reviewed and integrated in the form of a website containing interactive diagrams of protein analysis, trajectory visualization, and the associated downloadable data.

Every protein chain of the ATLAS database was simulated in 3 all-atom molecular dynamics replicates, using GROMACS [4] with the CHARMM36m force field [5]. Each replicate was performed in a water solution (TIP3P) under room temperature (300K) for 100 ns, using a triclinic box and Na+/Cl- ions at a concentration of 150 mM.

The obtained data allow us to propose novel flexibility prediction tools as well as important observations on the role of protein flexibility in molecular function. We report detailed analysis of the molecular dynamics trajectories in terms of the global protein behavior (RMSD, gyration radius, contact map dynamics), as well as its local flexibility of the protein backbone (RMSF, dihedral angle deviation, Neq, DSSP fluctuations). We compare our simulation results to experimental data, functional residue annotation as well as to the confidence of prediction of AlphaFold2 [6] models (pLDDT). We demonstrate that the obtained data provides evidence of protein dynamic properties, which was not available neither in the original structural data, nor from the AF2 predictions.

References

[1] Vander Meersche, Y., Cretin, G., de Brevern, A. G., Gelly, J. C., & Galochkina, T. 2021. MEDUSA: Prediction of protein flexibility from sequence. Journal of Molecular Biology, 166882.

[2] Carugo O. B-factor accuracy in protein crystal structures. Acta Crystallogr D Struct Biol. 2022 Jan 1;78(Pt 1):69-74.

[3] Cheng H, Schaeffer RD, Liao Y, Kinch LN, Pei J, Shi S, Kim BH, Grishin NV. ECOD: an evolutionary classification of protein domains. PLoS Comput Biol. 2014 Dec 4;10(12):e1003926.

MEMCROSS, A NEW TOOL TO EVALUATE MEMBRANE PERMEATION COEFFICIENTS OF DRUG-LIKE MOLECULES BY MD SIMULATIONS

MEHDI BENMAMERI[a], BENJAMIN CHANTEMARGUE[b], ANTOINE HUMEAU[a], PATRICK TROUILLAS[a,c], GABIN FABRE[a]

[a] Pharmacology & Transplantation, INSERM UMR-1248, Faculté de Pharmacie, Université de Limoges, France

[b] InSiliBio, Limoges, France

[c] CATRIN RCPTM, Olomouc, Czech Republic

Membrane crossing events are key in the pharmacokinetics of xenobiotics. Among them, passive permeation is the most ubiquitous. Therefore, an accurate and cost-effective prediction of permeation coefficient (logPerm) is highly valuable. Among the theoretical methods to predict logPerm, all-atom molecular dynamics simulations are versatile and can provide an accurate description of all intermolecular interactions. However, the cost associated with the required sampling often limits to the study of a few small permeants.

Here, we present MemCross, a tool based on the inhomogeneous solubility-diffusion model and the Accelerated Weight Histogram (AWH) method, and which takes subdiffusion[1] into account. We report the first use of AWH on all-atom membrane simulations.[2] The ease of use of MemCross and its relatively fast convergence allowed to benchmark it on more than 350 xenobiotics (mostly drugs) for which experimental logPerm are available. We believe this is the largest study performed with drug-like database, while performing all-atom MD simulations, totaling more than 1.2 millisecond of simulation time. A very good correlation was obtained with experimental logPerm of PC-based liposomes (R^2 = 0.83). Thus, MemCross is a flexible and affordable tool to evaluate logPerm of xenobiotics, while providing an atomistic description of the permeation process.

References

[1] Tse, C. H., Comer, J., Sang Chu, S. K., Wang, Y. & Chipot, C. Affordable Membrane Permeability Calculations: Permeation of Short-Chain Alcohols through Pure-Lipid Bilayers and a Mammalian Cell Membrane. J. Chem. Theory Comput. 15, 2913–2924 (2019).

[2] Benmameri, M., Chantemargue, B., Humeau, A., Trouillas, P. & Fabre, G. MemCross: Accelerated Weight Histogram method to assess membrane permeability. Biochimica et Biophysica Acta (BBA) - Biomembranes 1865, 184120 (2023).

MODELING THE EFFECTS OF PHOTO-INDUCED DRUGS IN A BIOLOGICAL MEMBRANE MODEL BY MOLECULAR DYNAMICS SIMULATIONS

ANASTASIIA DELOVA[a], RAÚL LOSANTOS[a,b]
ANTONIO MONARI[a]

[a] Université Paris Cité and CNRS, ITODYS, F-75006 Paris, France.

[b] Universidad de La Rioja, Departamento de Química, Centro de Investigación en Síntesis Química, 26006 Logroño, Spain.

In this work we study, by molecular modeling and simulation, the effects of a photo-induced molecular switch based on a cyclocurcumin derivative on model lipid bilayers of different composition. Two systems consisting of 1,2-dipalmitoyl-sn-glycero-3-phosphatidylcholine (DPPC) as well as a mixture of three different lipids (DPPC, DOPC, cholesterol) were used to mimic cell membranes. The cyclocurcumin derivative can be used in light-activated chemotherapy to selectively induce cell death by perturbing cellular membranes due to its structural perturbation brought by the E/Z photoisomerization[1]. This approach is an appealing alternative to conventional photodynamic therapy (PDT) because it is operative also in absence of oxygen, and hence could be efficient for hypoxic tumors.

Classical molecular dynamics simulations were used to investigate the interaction of the chromophore with the lipid bilayers. We also used enhanced sampling simulations via the coupling of ABF and Metadynamics (meta-eABF) to determining free energy profiles for the penetration of the switch in the membranes.

Additionally, the effects of different concentrations of chromophore, as well as the photoswitching, i.e. E/Z photoisomerization, on the membranes structural parameters were investigated and compared. We show that the interaction with the DPPC-only membrane is highly dependent on the concentration, furthermore we also evidenced a transition in the arrangement of the photoswitches from ordered to disordered state[2]. In the more complex membrane, we showed that the cyclocurcumin derivative interacted differently, causing less profound damages and changes in the lipid bilayer. However, for both membrane models we confirmed that the structural parameters of the bilayer are differently affected by the two isomers, and hence can be modulated through photoswitching, offering interesting perspectives for future applications.

References

[1] Pecourneau J. et al., “Biomimetic Photo-Switches Softening Model Lipid Membranes”, Langmuir, 2022

[2] A. Delova, R. Losantos, J. Pecourneau, “Perturbation of Lipid Bilayers by Biomimetic Photoswitches Based on Cyclocurcumin”, JCIM, 2022

FINDING HITS IN LARGE CHEMICAL SPACES BY COMBINING DOCKING WITH DEEP LEARNING

DAVID RINALDO[a], KUN YAO[b], MATTHEW P. REPASKY[c]
KARL LESWING[b], ROBERT ABEL[b], STEVEN V. JEROME[d]

[a] Schrödinger GmbH, Glücksteinallee 25, 68163 Mannheim, Germany
[b] Schrödinger, Inc., New York, New York 10036, United States
[c] Schrödinger, Inc., Portland, Oregon 97239, United States
[d] Schrödinger, Inc., San Diego, California 92121, United States

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

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

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

References

[1] Y. Yang et al., Efficient Exploration of Chemical Space with Docking and Deep Learning, J. Chem. Theory Comput., 2021 17, 7106-7119

[2] K. Konze et al., Reaction-Based Enumeration, Active Learning, and Free Energy Calculations To Rapidly Explore Synthetically Tractable Chemical Space and Optimize Potency of Cyclin-Dependent Kinase 2 Inhibitors, J. Chem. Inf. Model. 2019, 59, 3782-3793.

COARSE-GRAINED MOLECULAR DYNAMIC METHODS IN EXPLORATION OF LIGAND TRANSPORT PATHWAYS INSIDE LARGER PROTEINS WITH BURIED ACTIVE SITES

NISHITA MANDAL[a,b], BARTŁOMIEJ SURPETA[a,b]
JAN BREZOVSKY[a,b]

[a] Laboratory of Biomolecular Interactions and Transport, Institute of Molecular Biology and Biotechnology, Faculty of Biology, Adam Mickiewicz University, Poznan, Poland
[b] International Institute of Molecular and Cell Biology, Warsaw, Poland

Background: Biochemical reactions occur in active sites of enzymes, which are often buried deep inside the protein core and connected to the bulk solvent through molecular transport paths, called tunnels. These tunnels were proven to be important targets for protein engineering and drug discovery [1]. As proteins are dynamic in nature and most tunnels are transient, molecular dynamics (MD) simulations are themethod of choice for their study. However, given the rare opening of tunnels, their identification and analysis often require extensive and time-demanding simulations. Coarse-grained (CG-MD) simulations are, in general, capable of overcoming sampling limitations of classical MD (cMD). Therefore, we investigated to what extent CG models are suitable for tunnel analysis given their compromised resolution.

Methods: Replicated cMD & CG-MD simulations (5x 5 μs for each model) were carried out for a model system haloalkane dehalogenase enzyme(DEHAL), using AMBER18 package with SIRAH CG [2] and MARTINI CG models [3]. Tunnels were analyzed with the CAVER3 and TransportTools package [4].

Results: Since we have initially observed the compromised secondary structure of DHEAL in SIRAH CG simulations, we have developed a protocol for restraining α-helices and β-sheets. This approach reduced the secondary structure loss to around 5% in α-helices and effectively prevented disruption of β-sheets, while maintaining the protein RMSD (< 5 Å), comparable to other extensively used CG models. Exploration using MARTINI CG method with elastic network generated RMSD (< 4 Å) and relatively stable CG simulations. Regarding their application for tunnel analysis, both CG methods managed to identify all experimentally confirmed tunnels in agreement with cMD. The geometry of CG tunnels was similar to cMD, but there was a tendency for tunnels being somewhat narrower and longer. Additionally, we have systematically identified several tunnels not observed in cMD previously, suggesting improved sampling of tunnels in significantly reduced run-times.

Conclusions: CG models constitute a promising approach to investigate tunnels in proteins with significantly reduced computational demands compared to traditional methods, opening new possibilities for the identification of tunnels in large protein systems and massive datasets.

Acknowledgments: This research is supported by National Science Centre, Poland (2017/26/E/NZ1/00548), and the calculations were performed at the Poznan Supercomputing and Networking Center.

References

[1] S. M. Marques, L. Daniel, T. Buryska, Z. Prokop, J. Brezovsky, and J. Damborsky, "Enzyme Tunnels and Gates As Relevant Targets in Drug Design: TUNNELS AND GATES IN DRUG DESIGN," Med. Res. Rev., vol. 37, no. 5, pp. 1095–1139, Sep. 2017, doi: 10.1002/med.21430.

[2] M. R. Machado, E. E. Barrera, F. Klein, M. Sóñora, S. Silva, and S. Pantano, "The SIRAH 2.0 Force Field: Altius, Fortius, Citius," J. Chem. Theory Comput., vol. 15, no. 4, pp. 2719–2733, Apr. 2019, doi: 10.1021/acs.jctc.9b00006.

[3] S. J. Marrink, H. J. Risselada, S. Yefimov, D. P. Tieleman, and A. H. de Vries, "The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations," J. Phys. Chem. B, vol. 111, no. 27, pp. 7812–7824, Jul. 2007, doi: 10.1021/jp071097f.

[4] J. Brezovsky et al., "TransportTools: a library for high-throughput analyses of internal voids in biomolecules and ligand transport through them," Bioinformatics, vol. 38, no. 6, pp. 1752–1753, Mar. 2022, doi: 10.1093/bioinformatics/btab872.

POLARIZABLE FORCE FIELD AND DFT CALCULATIONS IN ANALYSIS OF THE INTERACTIONS BETWEEN CYCLODEXTRINS AND ENDOCRINE DISRUPTING CHEMICALS

ANNA HELENA MAZUREK[a], AS SUPERVISORS: PROF. THOMAS SIMONSON[b] AND PROF. ŁUKASZ SZELESZCZUK[c]

[a] Department of Organic and Physical Chemistry, Faculty of Pharmacy, Doctoral School, Medical University of Warsaw, Banacha 1 Str., 02-093 Warsaw, Poland

[b] Laboratoire de Biochimie (CNRS UMR7654), Ecole Polytechnique, 91-120 Palaiseau, France

[c] Department of Organic and Physical Chemistry, Faculty of Pharmacy, Medical University of Warsaw, Banacha 1 Str., 02-093 Warsaw, Poland

Background: Endocrine Disrupting Chemicals (EDCs)[1] are chemical substances external to a human body which mimic the endogenous hormones and cause a hormonal imbalance evoking for ex. fertility alterations or sex organ cancers. Also hormones delivered as pharmaceutics are described as EDCs. Those substances are usually steroid hormones characterized by a low solubility in water hence by a poor bioavailability. Therefore, in terms of EDCs we stand in front of two problems: EDCs removal from the environment and bioavailability enhancement of EDC-like medication. Cyclodextrins (CDs) are cyclic oligosaccharides. A chemical substance can be placed in their void so they are used both as toxin removing agents and as drug-carriers[2]. Inner structure and interaction energies of the specific EDCs-CDs complexes is a crucial information about the complexation possibility. Performance of experimental analyses can be supported by the molecular modelling. According to our literature reviews[3-5], the main computational approaches gaining the importance in this topic are: Density Functional Theory (DFT) calculations, Free Energy Perturbation[6,7] calculations, application of polarizable force fields (FFs)[8-11]. However, in terms of DFT, there is no consequence in the applied computation parameters what strongly influences the obtained results. In terms of FEP: this method has been so far used only for the macromolecules but is promising also for smaller systems. In turn, the polarizable FFs which can be also used for FEP, describe the ionic and polar interactions more accurately when compared to additive force fields but this method requires parametrization process for small molecules (ligands).

Objectives: Analysis of the chosen CD-EDC (estradiol, progesterone, bisphenol A) complexes by using various molecular modelling methods, choice of computational parameters and methods which most accurately describe such systems.

Methods: In the DFT approach, the following calculation parameters were compared: functionals (B3LYP, M06-2X), water models (PCM, SMD), presence or absence of dispersion correction. Thermodynamical parameters were calculated and compared to the experimental ones. Calculations were performed in Gaussian16 software. Semi-empirical approach (PM6 and PM7 with and without water models and dispersion corrections) and MD-MMGBSA calculations were made as a comparison. As a polarizable force field, AMOEBA FF8-11 was used. For the FF parametrization the Tinker software was applied.

Results and conclusions: DFT and semi-empirical approaches show the best agreement with the experimental data when a water model and dispersion correction are applied. Polarizable FF parametrization for estradiol, progesterone, bisphenol A and cyclodextrin unit has been successfully completed. The obtained parameters can be now used to perform FEP in classical and polarizable FF.

References

[1] Sellami, A.; Montes, M.; Lagarde, N. Predicting Potential Endocrine Disrupting Chemicals Binding to Estrogen Receptor α (ERα) Using a Pipeline Combining Structure-Based and Ligand-Based in Silico Methods. Int. J. Mol. Sci. 2021, 22, 2846 10.3390/ijms22062846

[2] Saokham P, Muankaew C, Jansook P, Loftsson T. Solubility of Cyclodextrins and Drug/Cyclodextrin Complexes. Molecules 2018, 23(5), 1161 10.3390/molecules23051161

[3] Mazurek, A.H.; Szeleszczuk, Ł. Current Status of Quantum Chemical Studies of Cyclodextrin Host–Guest Complexes. Molecules 2022, 27, 3874 10.3390/molecules27123874

[4] Mazurek, A.H.; Szeleszczuk, Ł.; Gubica, T. Application of Molecular Dynamics Simulations in the Analysis of Cyclodextrin Complexes. Int. J. Mol. Sci. 2021, 22, 9422 10.3390/ijms22179422

[5] Mazurek, A.H.; Szeleszczuk, Ł.; Simonson, T.; Pisklak, D.M. Application of Various Molecular Modelling Methods in the Study of Estrogens and Xenoestrogens. Int. J. Mol. Sci. 2020, 21, 6411 10.3390/ijms21176411

[6] Fratev, F., Sirimulla, S. An Improved Free Energy Perturbation FEP+ Sampling Protocol for Flexible Ligand-Binding Domains. Sci Rep 9, 16829 (2019) 10.1038/s41598-019-53133-1

[7] Simonson T., Archontis G., Karplus M. Free energy simulations come of age. Accts Chem Res. 2002, 35, 6, 430-437 10.1021/ar010030m

[8] Ponder, J. W., Wu, C., Ren, P., Pande, V. S., Chodera, J. D., Schnieders, M. J. et. al. Current Status of the AMOEBA Polarizable Force Field. The Journal of Physical Chemistry B, 2010, 114(8), 2549–2564 10.1021/jp910674d

[9] Shi, Y., Xia, Z., Zhang, J., Best, R., Wu, C., Ponder, J. W., Ren, P. Polarizable Atomic Multipole-Based AMOEBA Force Field for Proteins. Journal of Chemical Theory and Computation, 2013, 9(9), 4046–4063. 10.1021/ct4003702

[10] Wu JC, Chattree G, Ren P. Automation of AMOEBA polarizable force field parameterization for small molecules. Theor Chem Acc. 2012, 131(3):1138. 10.1007/s00214-012-1138-6

[11] Zheng C., Lu C., Jing, Z. Wu, C. Piquemal, JP., Ponder, JW., Ren P. AMOEBA Polarizable Atomic Multipole Force Field for Nucleic Acids. J Chem Theory Comput. 2018,14(4):2084-2108 10.1021/acs.jctc.7b01169

BLOCKS SELFIES, A NEW METHOD TO EXPLORE THE CHEMICAL SPACE OF PROTEIN LIGAND INTERACTIONS

ETIENNE REBOUL[a], ZOE WEFERS[b], JÉRÔME WALDISPÜHL[b] ANTOINE TALY[a]

[a] Laboratoire de biochimie théorique, CNRS, Institut de biologie physico-chimique, Université Paris Cité
[b] Waldispühl group, McGill University

The exploration of the chemical space is a critical task in De novo drug Design. Using genetic algorithm (GA) and derivative methods, it is theoretically possible to progressively change the chemical structure of a molecule. Therefore, it allows exploring the neighborhood of the original molecule in the chemical space.

Our method was developed from the STONED SELFIES algorithm[1]. It is a pseudo GA that is based upon a molecular string representation called SELFIES[2]. A Molecular string representation can be understood as a language where a molecule is a "sentence" and its bounds and atoms are "words". The "words" of SELFIES langage are string tokens. They are particularly suitable for GAs because they can be easily deleted, inserted and replaced. However, SELFIES have a complex syntax that is relatively fragile. The random mutations used in GAs may introduce errors that result in the truncation of the chemical structure.

In our study, we have quantified the truncation rate of mutated SELFIES obtained with STONED SELFIES. We found out that even limited truncation can lead to malformed chemical structure. Therefore, we propose a new method where the mutations are not made tokens by tokens but by blocks of tokens. A block is defined as an ensemble of tokens that are completely independent from the rest of the tokens in a SELFIES. We present a case study of optimizing the chemical structure of Celecoxib[3], a well-known Non-steroidal anti-inflammatory drug, to improve its affinity to cyclo-oxygenase 2.

References

[1] Nigam, A., Pollice, R., Krenn, M., dos Passos Gomes, G. & Aspuru-Guzik, A. Beyond Generative Models: Superfast Traversal, Optimization, Novelty, Exploration and Discovery (STONED) Algorithm for Molecules using SELFIES. https://chemrxiv.org/engage/chemrxiv/article-details/60c753f00f50db6830397c37 (2021) doi:10.26434/chemrxiv.13383266.v2.

[2] Krenn, M., Häse, F., Nigam, A., Friederich, P. & Aspuru-Guzik, A. Self-referencing embedded strings (SELFIES): A 100% robust molecular string representation. Mach. Learn. Sci. Technol. 1, 045024 (2020).

[3] Clemett, D. & Goa, K. L. Celecoxib: A Review of its Use in Osteoarthritis, Rheumatoid Arthritis and Acute Pain. Drugs 59, 957–980 (2000).

EFFICIENT GPU COMPUTATION OF LARGE PROTEIN SES FOR MOLECULAR ILLUSTRATION

CYPRIEN PLATEAU-HOLLEVILLE[a], MAXIME MARIA[a] STÉPHANE MÉRILLOU[a], MATTHIEU MONTES[b]

[a] Laboratoire XLIM, UMR CNRS 7252, Université de Limoges, France

[b] Laboratoire GBCM, EA7528, Conservatoire National des Arts et Metiers, Hésam Université, France

The Solvent-Excluded Surface (SES) is a critical representation of molecular systems. It is widely used to support the exploration and understanding of molecular interactions for basic research in structural and computational biochemistry and drug discovery applications. The calculation of the SES is computationally intensive which renders its use on large molecular systems unpracticable due to the limited scalability of its current implementations. Its fast computation has remained a challenge originating from the difficulty to process and store its structure. We show that, based on our work leveraging advances made in its depiction, we can compute and render the SES of large molecular systems on GPUs. Thanks to its ability to reduce memory consumption while keeping competitive processing performance, our method will support the illustration of molecular dynamics of small- and large-scale systems as well as making available high-quality rendering of this surface on lower-end hardware.

EXPLORING SELF-HEALING PEPTIDE HYDROGELS THROUGH EXPERIMENTAL AND COMPUTATIONAL DIALOGUE

F. BARBAULT[a], M. CRIADO-GONZALEZ[b], M.I. PEÑAS[b,c] A.J. MÜLLER[c,d], R. HERNANDEZ[b] AND F. BOULMEDAIS[e]

[a] ITODYS, Université Paris Cité, France
[b] ICTP-CSIC, Madrid, Spain
[c] POLYMAT, University of the Basque Country, San-Sebastián, Spain
[d] Ikerbasque, Bilbao, Spain
[e] Institut Charles Sadron, Strasbourg, France

The exploration of the chemical space is a critical task in De novo drug Design. Using genetic algorithm Hydrogels are 3D networks either composed of crosslinked hydrophilic polymers, block-copolymer micelles, colloids, or peptides[1]. These networks are capable of swelling in water resulting in the formation of soft materials that bear similarities to biological tissues due to their high-water content, porous structure, and mechanical properties[2].

We aim to develop supramolecular hydrogels with stimuli responsive properties using sequence-defined low molecular weight hydrogelators (LMWH). Specifically, we observed the successful gelation of a tripeptide Fmoc-FFpY with a tyrosine phosphate group (pY) that imparts water solubility and self-assembly properties in the presence of NaCl.

Interestingly, these hydrogels bear different structural organizations according to the temperature or ionic strengths, leading to tunable hydrogels with various physical properties. To understand these effect at the atomic scale, numerous computational studies were conducted, mainly through molecular dynamics, for various environments (concentration, temperature, packing) and compared to experimental data such as infra-red spectroscopy (IR), circular dichroism (CD) or transmission electron microscopy (TEM).

The present study reveals a high degree of agreement between the experimental observations and the theoretical models, thereby highlighting the efficacy of a computational methodology in generating meaningful outcomes. Collectively, these results underscore the prospective applicability of computational investigations in the design and advancement of supramolecular hydrogels exhibiting desirable properties.

References

[1] C. A. Dreiss, « Hydrogel design strategies for drug delivery », Current Opinion in Colloid & Interface Science, vol. 48, p. 1-17, août 2020, doi: 10.1016/j.cocis.2020.02.001.

[2] X. J. Loh, V. P. Nam Nguyen, N. Kuo, et J. Li, « Encapsulation of basic fibroblast growth factor in thermogelling copolymers preserves its bioactivity », J. Mater. Chem., vol. 21, no 7, p. 2246, 2011, doi: 10.1039/c0jm03051a

LIBPROTEIN: A RAPID AND VERSATILE ANNOTATION LIBRARY FOR PROTEIN POST-TRANSLATIONAL ANNOTATIONS

HAMADY BA[a], AND STÉPHANE TÉLETCHÉA[a]

[a] Nantes Université, CNRS, US2B, UMR6286, 2 chemin de la houssinière, F-44000, Nantes, France
Corresponding Author: stephane.teletchea@univ-nantes.fr

1 Introduction

Protein are complex biological molecular entities defined primarily by their amino acids sequence. Upon or after ribosomal translation, the messenger RNA is transformed into a polypeptide chain, often these amino acids may further be processed into more complex entities with the addition of functional groups such as carbohydrates, pyrophosphates or other small chemical entities. [1], [2]

Since these co- or post-translational modifications are diverse and spread into multiple databases it is difficult to assemble these data, and therefore even more complex to reuse them for bioinformatics studies. The aim of this project is to ease protein PTM retrieval and annotation, and to provide usage examples.

2 Implementation and user cases

LibProtein is a C++ library allowing the retrieval and annotation of any given protein with modifications available first in uniprot. It contains an internal representation of PTM as a 3-letter code. This library also contains annotations from SCOP, CATH and PFAM to enrich sequences with functional domains descriptions.

The library implementation is flexible enough to envision the incorporation of experimental or predictions of PTMs.[3], [4] These additional annotations will provide a more complete overview of a given protein life cycle.

The library allows to enrich structure analysis in PyMol by the addition of annotations tags into the protein sequence (PTM, families and categories), to provide data for machine learning development, and to score protein structure models often lacking PTM predictions. [5], [6]

3 Acknowledgements

Hamady Ba's internship is funded by the TROPIC project which received financial support from the ANR grant "Programme d'investissements d'Avenir" (ANR-16-IDEX-0007), from Région Pays de La Loire and from Nantes Métropole.

References

[1] D. Tsikas, « Post-translational modifications (PTM): analytical approaches, signaling, physiology and pathophysiology—part I », Amino Acids, vol. 53, no 4, p. 485487, avr. 2021, doi: 10.1007/s00726-021-02984-y.

[2] A. G. de Brevern et J. Rebehmed, « Current status of PTMs structural databases: applications, limitations and prospects », Amino Acids, janv. 2022, doi: 10.1007/s00726-021-03119-z.

[3] M. Audagnotto et M. Dal Peraro, « Protein post-translational modifications: In silico prediction tools and molecular modeling », Comput. Struct. Biotechnol. J., vol. 15, p. 307319, janv. 2017, doi: 10.1016/j.csbj.2017.03.004.

[4] Y. Deng, Y. Fu, H. Zhang, X. Liu, et Z. Liu, « Protein Post-translational Modification Site Prediction using Deep Learning », Procedia Comput. Sci., vol. 198, p. 480485, 2022, doi: 10.1016/j.procs.2021.12.273.

[5] J. Jumper et D. Hassabis, « Protein structure predictions to atomic accuracy with AlphaFold », Nat. Methods, vol. 19, n o 1, p. 1112, janv. 2022, doi: 10.1038/s41592-021-01362-6.

[6] M. Baek et al., « Accurate prediction of protein structures and interactions using a three-track neural network », Science, vol. 373, no 6557, p. 871876, août 2021, doi: 10.1126/science.abj8754.

A TOOLKIT FOR COVALENT DOCKING WITH GOLD: FROM AUTOMATED LIGAND PREPARATION WITH KNIME TO BOUND PROTEIN-LIGAND COMPLEXES

LAURIANNE DAVID[a], ANISSA MDAHOMA[a], NATESH SINGH[a,*] SÉBASTIEN BUCHOUX[b], EMILIE PIHAN[c], CONSTANTINO DIAZ[a] OBDULIA RABAL[a,*]

[a] Evotec SE, Molecular Architects, Integrated Drug Discovery, Campus Curie, 195 Route d'Espagne, 31036 Toulouse, France
[b] Evotec SE, Scientific Data Management, Campus Curie, 195 Route d'Espagne, 31036 Toulouse, France
[c] Discngine, 79 Av. Ledru Rollin, 75012 Paris, France

Covalent inhibitors are typically characterized by the formation of a chemical bond between an electrophilic moiety (warhead) on the ligand part and a nucleophilic residue, most commonly cysteine, of the protein target. Though many existing covalent docking tools are successful, they have several limitations[1]. They require time-consuming tasks such as pre-processing of ligands (standardization, filtering based on warheads type), setting up of reactions, protein preparation, complex formation, etc... They can only be used to dock a small library of compounds due to the high CPU time needed for performing covalent computations and scoring, rendering them unsuitable for performing large-scale virtual screening. We developed a toolkit for performing automated covalent docking in a fast and effective manner by combining GOLD, MOE, KNIME, and Python programs. In a first step, a KNIME workflow was developed to perform ligand preparation in a manner compatible with the covalent docking protocol of GOLD. The protein structures were prepared using the MOE QuickPrep module. Subsequently, a Python program was written to perform the covalent docking of ligands, by invoking the GOLD docking engine, in a parallelized fashion. Finally, the

protein-ligand complexes were generated for the docking poses by calling a submodule of the Python program. We applied this toolkit retrospectively on six targets (NUDT7, OTUB2, EGFR, Cathepsin K, XPO1, and HCV NS3 protease) with known crystal structures and sets of active and inactive covalent inhibitors and assessed the potential of four GOLD docking scoring functions[2] (GoldScore, ASP, ChemScore, and PLP). For most of the targets, the virtual screening metrics were satisfactory for the different scoring functions, demonstrating the program's usefulness in performing prospective virtual screening.

The KNIME workflow "Evotec_Covalent_Processing_forGOLD.knwf" for the preparation of the ligands is available in the KNIME Hub https://hub.knime.com/emilie_pihan/spaces.

The Python program is available at https://gitlab.com/seb-buch/covalent_docking_helper.

References

[1] Oyedele, AQ.K., Ogunlana, A.T., Boyenle, I.D. et al. Docking covalent targets for drug discovery: stimulating the computer-aided drug design community of possible pitfalls and erroneous practices. Mol Divers (2022). https://doi.org/10.1007/s11030-022-10523-4

[2] Verdonk, M.L., Cole, J.C., Hartshorn, M.J., Murray, C.W. and Taylor, R.D. (2003), Improved protein–ligand docking using GOLD. Proteins, 52: 609-623. https://doi.org/10.1002/prot.10465

ANALYSIS OF ALPHAFOLD2 HUMAN PROTEOME

ALEXANDRE G. DE BREVERN[a]

[a] INSERM UMR_S 1134, DSIMB Bioinformatics team, Université Paris Cité and Université de la Réunion, 75014 Paris - France

Knowledge of the 3D structure of proteins is a valuable asset for understanding their precise biological mechanisms. However, the cost of production of 3D structures and experimental difficulties limit their obtaining. The proposal of 3D structural models is consequently an appealing alternative. The release of the AlphaFold Deep Learning approach has revolutionized the field[1]. The near-complete human proteome proposal[2] makes it possible to analyse large amounts of data and evaluate the results of the approach in greater depth. The 3D human proteome was analysed in light of the classic secondary structures (α-, 310- and π-helices, bends, turns, β-sheets and coil), and many less-used protein local conformations (PolyProline II helices, type of γ-turns, of β-turns and of β-bulges, curvature of the helices, and a structural alphabet).

Some conformations are over-represented in high confidence index (pLDDT) regions, i.e. α-helices. At the opposite, PolyProline II helices are surprisingly encountered with a low confidence index. It is more striking for γ-turns that are never associated to high pLDDT while β-turns had not this issue. Another unexpected result is cis ω angles (associated or not to Proline), they are never correctly predicted (low pLDDT), a problematic case as they are often associated to biological function. The rate of β-sheet is lower than expected in the human proteome. It is well know that it is the most complex classical secondary structure to predict. However, these analyses underlined that the corresponding β-strand Protein Block d have correct occurrences. Perhaps some local conformations with Protein Block d correspond to non-finished β-sheets, i.e. there could be a potential under-representation of β-sheets by AlphaFold 2.

Without questioning the global quality of the approach, this analysis highlights certain local conformations, which maybe poorly predicted and they could therefore be better addressed.[3]

References

[1] Jumper J. et al. Highly accurate protein structure prediction with AlphaFold. Nature, 2021. 596, 583-589.

[2] Tunyasuvunakool K. et al. Highly accurate protein structure prediction for the human proteome. Nature. 596, 596-589.

[3] de Brevern AG. An agnostic analysis of the human AlphaFold2 proteome using local protein conformations, Biochimie, 2023, 207, 11-19.

COMPUTATIONAL INSIGHTS INTO THE MECHANISMS OF TRIGLYCERIDE FORMATION: A STUDY OF DGAT1 USING MOLECULAR DYNAMICS

JENNIFER SAPIA[a], PABLO CAMPOMANES[a] AND STEFANO VANNI[a]

[a] Department of Biology, University of Fribourg, Chemin du Musée 10, 1700 Fribourg, Switzerland

Lipid droplets (LDs) are cytosolic organelles that play a vital role in maintaining cellular lipid homeostasis by acting as an energy reservoir and protecting cells from lipotixicity. They have a unique composition comprising of a central hydrophobic core of neutral lipids, mainly triglycerides (TGs) and sterol esters (SEs), surrounded by a phospholipid monolayer decorated with a great diversity of LD-specific proteins. LD biosynthesis takes place in the endoplasmic reticulum (ER), whereby the neutral lipids are synthetized as well. In mammalian cells, the synthesis of TGs is catalyzed by acyl-CoA: diacylglycerol O-acyltransferase 1 (DGAT1), an enzyme that belongs to the membrane-bound O-acyltransferase (MBOAT) enzyme family [1]. Despite its physiological importance, a detailed understanding of how TG molecules are synthesized into the ER remains elusive. However, the recent determination of the 3D structure of DGAT1 using cryo-electron microscopy [2,3] has opened new avenues for understanding its mechanism. To investigate the dynamics of DGAT1, we used molecular dynamics (MD), a physics-based computational approach that allows to study the behavior of molecules over time. We carried out MD simulations at both atomistic and coarse-grain levels, using the 3D structure of DGAT1 dimer in a model lipid bilayer enriched with its natural substrate, di-oleoyl-glycerol (DOG).

Our simulations reveal that DOG can enter the catalytic pocket of DGAT1 through a huge transmembrane, hydrophobic cavity. Despite showing diverse entry pathways, likely related to the different conformations adopted by the acyl-donor substrate oleoyl-CoA, DOG mostly penetrate the catalytic pocket from the luminal leaflet of the ER. Moreover, thanks to the conical shape of DGAT1, our simulations show a non-negligible deformation in the surrounding environment, where the membrane assumes a positive curvature. Hence, our simulations suggest a preferential localization of DGAT1 on positive and highly curved membrane regions of the ER.

Taken together, our findings offer new insights into the role of DGAT1 in LD formation in the ER. We speculate that the local accumulation of DOG, the precursor of TG, in curved regions of the ER [4] matches the preferential localization of DGAT1 in the ER membrane, thereby positively promoting the tendency of LDs to form in those specific areas [5].

References

[1] Sui et al., Structure and catalytic mechanism of a human triacylglycerol-synthesis enzyme, Nature, 2020, 323-328.

[2] Wang et al., Structure and mechanism of human diacylglycerol O-acyltransferase 1, Nature, 2020, 15-27.

[3] Hofmann, K. A superfamily of membrane-bound O-acyltransferases with implications for Wnt signaling, Trends Biochem Sci, 2000, 111–112.

[4] McMahon HT, Boucrot E. Membrane curvature at a glance. J Cell Sci., 2015, 1065-1070.

[5] Chen et al, FIT2 organizes lipid droplet biogenesis with ER tubule-forming proteins and septins, J Cell Biol, 2021.

SIMULATIONS OF THE NUCLEOSOMAL DNA : MAPPING THE RADICAL CATION GUANINE

MAXIME KERMARREC[a,*], ELISE DUMONT[b,c], NATACHA GILLET[a]

[a] Université de Lyon, ENS de Lyon, CNRS UMR 5182, Université Claude Bernard Lyon 1, Laboratoire de Chimie, F69342, Lyon, France

[b] Université Côte d'Azur, CNRS, Institut de Chimie de Nice, UMR 7272 – 06108 Nice, France

[c] Institut Universitaire de France, 5 rue Descartes, 75005 Paris, France

[*]Presenting author e-mail address: maxime.kermarrec@ens-lyon.fr

An overexposure of DNA to an oxidative stress may results to various diseases related to mutations, especially cancers[1]. Therefore, understanding the behavior of the DNA oxidative damages is of medical and pharmaceutical interest. Guanine has the lowest ionization potential among the 4 DNA nucleobases[2] thus it is the most likely to be oxidized. However, the guanine redox properties can be impacted by the biomolecular environment around the nucleobases, which modulates the charge transfer ability of the DNA double strand. Then, the specific environment of the nucleosome, the fundamental unit of the chromatin, must be considered to understand DNA oxidative damages in cell. The nucleosome is composed of double-stranded DNA wrapped around four histone dimers rich in positively charged residues, which leads to a wide landscape of DNA-protein interaction.

Here, we investigate the factors and effects that can modulate the ionization potential of guanines in the nucleosome with a focus on their interactions with histone flexible N-terminal tails by means of a multi-scale approach. We combine classical molecular dynamics simulation at the µs scale and QM/MM calculations based on the FO-DFTB/MM[3,4] approach. This method allows us to determine the ionization potential of a large number of nucleobases and the electronic coupling between them.

According to our results, the proximity of positive charges from histone tails or sodium cations seems to be the most predominant cause of variation of the ionization potential. Sequence and orientation towards the nucleosome core have little impact on this redox property. The electronic coupling values mostly depend on the geometry of the considered guanine pair, but the strongest ones are compatible with sub-nanosecond charge transfers (tens of meV). Our understanding of the combinatorial impacts of the structure and dynamics of nucleosome on the DNA charge transfer parameter is currently limited by our sampling but also by our use of conventional approach. Consequently, we want develop machine learning algorithms to improve our analysis and predictive power of nucleosomal guanine redox properties.

References

[1] Greenberg, M. M. In Vitro and in Vivo Effects of Oxidative Damage to Deoxyguanosine. Biochem. Soc. Trans. 2004, 32 (1), 46–50

[2] R. Chakraborty and D. Ghosh, The effect of sequence on the ionization of guanine in DNA, Physical Chemistry Chemical Physics, 2016, 18, 6526–6533

[3] Kubar, T., Woiczikowski, P. B., Cuniberti, G., & Elstner, M. (2008). Efficient calculation of charge transfer matrix elements for hole transfer in DNA. The Journal of Physical Chemistry B, 112(26), 7937-7947

[4] Kubar, T., & Elstner, M. (2008). What governs the charge transfer in DNA? The role of DNA conformation and environment. The Journal of Physical Chemistry B, 112(29), 8788-8798

VTX: HIGH-PERFORMANCE MOLECULAR STRUCTURE AND DYNAMICS VISUALIZATION SOFTWARE

MAXIME MARIA[a,b], SIMON GUIONNIERE[b], NICOLAS DACQUAY[a], CYPRIEN PLATEAU-HOLLEVILLE[b], YASSINE NAIMI[c] JEAN-PHILIP PIQUEMAL[d], GUILLAUME LEVIEUX[e], MATTHIEU MONTES[a]

[a] Laboratoire GBCM, EA758, Hesam Universite, CNAM
[b] Laboratoire XLIM, Universite de Limoges
[c] Qubit Pharmaceuticals SAS
[d] LCT, UMR7616, Sorbonne Universite
[e] 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].

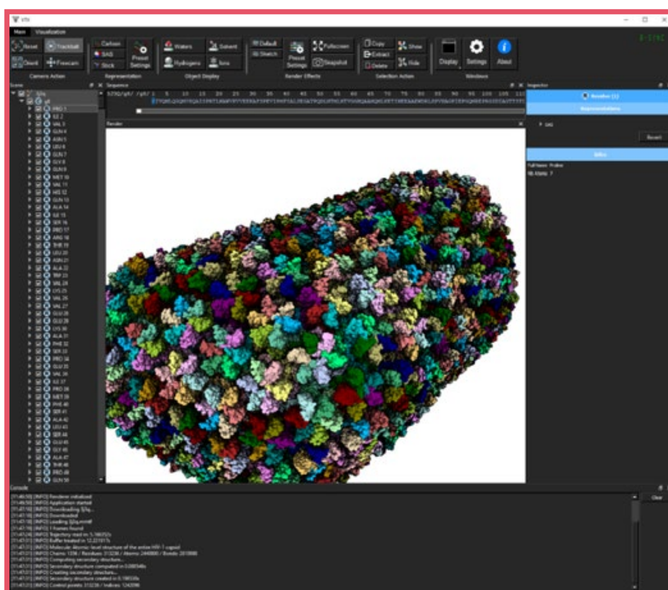


Figure 1. Illustration of VTX Graphical user interface on the structure of the human HIV Capsid (2.5 million atoms, PDB ID: 3J3Q).

Here we present VTX, a new molecular visualization software that includes a real-time high-performance molecular graphics engine dedicated to the visualization of large structure and dynamics of molecular systems. It is capable to process most molecular structures and trajectories file formats. VTX disposes of an interactive camera system controllable via the keyboard and/or mouse that includes different modes: 1. a classical trackball mode where the cam-era revolves around a fixed focus point and 2. a first-person free-fly navigation mode where the user fully controls the movement of the camera. VTX includes an intuitive and highly usable graphical user interface and tools designed for expert and non-expert users. It is free for non-commercial use at http://vtx.drugdesign.fr

References

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

IT TAKES TAU TO TANGO : INVESTIGATING THE FUZZY INTERACTION BETWEEN THE TAU-R2 REPEAT AND TUBULIN C-TERMINAL TAILS[1]

JULES MARIEN[a] , CHANTAL PRÉVOST[a] , SOPHIE SACQUIN-MORA[a]

[a] CNRS, Université de Paris, UPR 9080, Laboratoire de Biochimie Théorique, 13 rue Pierre et Marie Curie, 75005 Paris, France

Fuzzy complexes are a relatively new type of protein interaction involving one or more intrinsically disordered elements which proved more and more biologically important in the past two decades [2]. Intrinsically disordered protein Tau is known to stabilize microtubules (MT), but little is known regarding its interaction with the intrinsically disordered C-terminal tails (CTTs) of tubulins [3]. We used all-atom molecular dynamics simulations to decipher the behavior of the R2 repeat domain of the Tau protein in complex with 3 tubulin monomers β-α-β. We built the C-terminal tails corresponding to isotypes βI-αI-βI and βIII-αI-βIII by homology. Our analysis confirmed the existence of a highly stable interface area involving Ser289, a serine notably phosphorylated in Alzheimer's disease [4], and revealed a modification of the dynamics of CTTs in presence of R2. The latter result gives more weight to Lessard and Berger's idea that interactions between kinesins and MT could be indirectly mediated by the presence of Tau [5]. We also propose a "wrapping mechanism" of the CTTs around R2 which might provide more insight regarding the stabilizing impact of CTTs in Tau/MT complexes observed by Hinrichs et al [3].

References

[1] Marien J, Prévost C, and Sacquin-Mora S (2023) It takes Tau to tango : Investigating the fuzzy interaction between the Tau-R2 repeat domain and the C-terminal tails of tubulins, bioRxiv 2023.02.09.527845; doi: https://doi.org/10.1101/2023.02.09.527845

[2] Sacquin-Mora S, Prévost C. When Order Meets Disorder: Modeling and Function of the Protein Interface in Fuzzy Complexes. Biomolecules. 2021 Oct 16;11(10):1529. doi: 10.3390/biom11101529. PMID: 34680162; PMCID: PMC8533853.

[3] Hinrichs, M. H., Jalal, A., Brenner, B., Mandelkow, E., Kumar, S., and Scholz, T. (2012) Tau protein diffuses along the microtubule lattice, The Journal of biological chemistry 287, 38559-38568, doi: 10.1074/jbc.M112.369785

[4] Drummond E, Pires G, MacMurray C, Askenazi M, Nayak S, Bourdon M, Safar J, Ueberheide B, and Wisniewski T. (2020) Phosphorylated tau interactome in the human Alzheimer's disease brain. Brain. 2020 Sep 1;143(9):2803-2817. doi: 10.1093/brain/awaa223. PMID: 32812023; PMCID: PMC7526722.

[5] Lessard, D., and Berger, C. (2021) The Microtubule Associated Protein Tau Regulates KIF1A Pausing Behavior and Motility, bioRxiv, 2021.2008.2011.455914 ,doi: https://doi.org/10.1101/2021.08.11.455914

MODELLING THE CRYSTALLIZATION OF CARBAMAZEPINE

RADOST HERBOTH[a], ALEXANDER LYUBARTSEV[a]

[a] Department of Materials and Environmental Chemistry, Stockholm University, Svante Arrhenius väg 16C, 106 91 Stockholm, Sweden

Polymorphism, the ability of a material to crystallize in different solid forms, is highly common and mostly unpredictable based on molecular structure[1]. This poses a challenge especially in the manufacturing of effective drugs, where active pharmaceutical ingredients (APIs) may crystallize in polymorphic forms that differ greatly in their physical and chemical properties, like solubility and thus bioavailability. The polymorphism of many APIs is known and the various structures have been characterized and studied extensively. However, there is still a lack of feasible methods for directed crystallization of the desired forms.

This work employs computational tools to study the crystallization of APIs using an example with well-documented polymorphism, carbamazepine (CBZ). CBZ is an established drug in the treatment of epilepsy and trigeminal neuralgia[2]. Its polymorphism has been thoroughly investigated and four anhydrous forms of the molecule are known: A P-monoclinic form known as Form III, a trigonal form known as Form II, a triclinic form known as Form I and a C-monoclinic form known as Form IV[3]. Moreover, a monoclinic dihydrate form has been reported to crystallize in water[4]. The polymorphs are seen to have different thermochemical properties[5] and show differences in plasma concentration after administration as a drug[6].

The free energy landscape of the crystallization process is first explored in terms of adsorption/dissolution of molecules at a crystal surface. We computed the potential of mean force (PMF) of a carbamazepine molecule at ten different surfaces, among them surfaces of three anhydrous polymorphs (Forms I, II and IV), four crystal surfaces for the most stable one (Form III with surfaces 010, 001, 100, 011 and 102) and two surfaces of the dihydrate (010 and 001). The resulting PMFs are then used to compute the number of states that would correspond to crystal growth. We found that only in Forms I, III (100 surface), IV and the 001 surface of the dihydrate, the attachment of the molecule occurs in this way, while other surfaces show attachment with parts of the molecule rotated by 180° (Forms II and III in 010, 011 and 102 surfaces) or in completely random orientations (Form III 001 surface and dihydrate 010 surface).

We hope to relate this to experimental findings of the morphology of the crystals, so one can pinpoint the dominant direction of crystal growth and offer an explanation for the stability of certain polymorphs over others.

References

[1] Cruz-Cabeza AJ, Reutzel-Edens SM, Bernstein J. Facts and fictions about polymorphism, Chem. Soc. Rev., 2015, 8619–8635.

[2] Killian JM, Fromm GH. Carbamazepine in the Treatment of Neuralgia: Use and Side Effects, Arch. Neurol., 1968, 129–136.

[3] Grzesiak AL et al. Comparison of the Four Anhydrous Polymorphs of Carbamazepine and the Crystal Structure of Form I, J. Pharm. Sci., 2003, 2260–2271.

[4] Harris RK et al. Structural Studies of the Polymorphs of Carbamazepine, Its Dihydrate, and Two Solvates, Org. Process Res. Dev., 2005, 902-910.

[5] Clout AE et al. Polymorphic Phase Transitions in Carbamazepine and 10,11-Dihydrocarbamazepine, Chem. Eur. J., 2018, 13573–13581.

[6] Kobayashi Y et al. Physicochemical properties and bioavailability of carbamazepine polymorphs and dihydrate, Int. J. Pharm., 2000, 137–146.

UNVEILING MUTATION EFFECTS ON THE STRUCTURAL DYNAMICS OF THE MAIN PROTEASE FROM SARS-COV-2 WITH HYBRID SIMULATION METHODS AND MACHINE LEARNING

ANA L. SCOTT[a,b], PATRICIA GASPARINI[a], ERIC A. PHILOT[a], DAVID PERAHIA[c], RAFAEL SIMÕES[d], ANGELO MAGRO[e]

[a] Laboratory of Computational Biology and Biophysics- Federal University of ABC, Avenida dos Estados 5001, Santo André, SP, Brazil

[b] IGBMC, University of Strasbourg, 1 Rue Laurent Fries, 67400 Illkirch-Graffenstaden, France.

[c] LBPA, ENS Paris-Saclay, 4 Av. des Sciences, 91190 Gif-sur-Yvett, France

[d] Department of Bioprocesses and Biotechnology, School of Agriculture (FCA), Unesp, Botucatu, São Paulo, Brazil.

[e] Institute of Biotechnology (IBTEC), Unesp, Botucatu, São Paulo, Brazil.

Integrated methods combining NMA (using ENM or full atomic models), molecular dynamics, and experimental data from biophysical techniques such as X-ray crystallography, NMR, SAXS, EM, and FRET have been found useful for studying the correlation between the complex structural dynamics of macromolecules and their mechanisms of function[1,2]. These hybrid approaches allow efficient sampling of the conformational space, taking into account both slow and fast movements[2]. The main protease of SARS-CoV-2 (called Mpro or 3CLpro) is essential for processing polyproteins encoded by viral RNA. Several Mpro mutations were found in SARS-CoV-2 variants, which are related to higher transmissibility, pathogenicity, and resistance to neutralization antibodies. Macromolecules adopt several favored conformations in solution depending on their structure and shape, determining their dynamics and function. In this study, we used a hybrid simulation method to generate intermediate structures along the six lowest frequency normal modes and sample the conformational space and characterize the structural dynamics and global motions of WT SARS-CoV-2 Mpro and 48 mutations, including mutations found in P.1, B.1.1.7, B.1.351, B.1.525 and B.1.429+B.1.427 variants[3]. We tried to contribute to the elucidation of the effects of mutation in the structural dynamics of SARS-CoV-2 Mpro. A machine learning analysis was performed following the investigation regarding the influence of the K90R, P99L, P108S, and N151D mutations on the dimeric interface assembling of the SARS-CoV-2 Mpro. The parameters allowed the selection of potential structurally stable dimers, which demonstrated that some single surface aa substitutions not located at the dimeric interface (K90R, P99L, P108S, and N151D) are able to induce significant quaternary changes. Furthermore, our results demonstrated, by a Quantum Mechanics method, the influence of SARS-CoV-2 Mpro mutations on the catalytic mechanism, confirming that only one of the chains of the WT and mutant SARS-CoV-2 Mpros are prone to cleave substrates. Finally, it was also possible to identify the aa residue F140 as an important factor related to the increasing enzymatic reactivity of a significant number of SARS-CoV-2 Mpro conformations generated by the normal modes-based simulations.

References

[1] Zhang, B., Liu, D., Zhang, Y., Shen, H.-B., & Zhang, G.-J. (2022). Accurate flexible refinement for atomic-level protein structure using cryo-EM density maps and deep learning. Briefings in Bioinformatics, bbac026. https://doi.org/10.1093/bib/bbac026

[2] Krieger, J. M., Doruker, P., Scott, A. L., Perahia, D., & Bahar, I. (2020). Towards gaining sight of multiscale events: Utilizing network models and normal modes in hybrid methods. Current Opinion in Structural Biology, 64, 34–41. https://doi.org/10.1016/j.sbi.2020.05.013

[3] Gasparini P, Philot EA, Pantaleão SQ, Torres-Bonfim NESM, Kliousoff A, Quiroz RCN, Perahia D Simões RP, Magro AJ, Scott AL, Unveiling Mutation Effects on the Structural Dynamics of the Main Protease from SARS-CoV-2 with Hybrid Simulation Methods accepted to the Journal of Molecular Modeling and Graphics, 2023.

Financial Suport: Fapesp and CNPq

SITE-SPECIFIC CYSTEINE TARGETING: THE NEXT FRONTIER IN STRUCTURE-GUIDED DESIGN OF NOVEL TUBULIN BINDERS

M. SHEVELEV[a], H. PEREZ-PEÑA[a,b], A-C. ABEL[b,c], Z. BOIARSKA[b,c] F. BONATO[b,d], A. SOLIMAN[d], M. Á. OLIVA[d], J. F. DÍAZ[d], A. VARNEK[a] M. O. STEINMETZ[c], D. HORVATH[a], S. PIERACCINI[b], D. PASSARELLA[b] A. E. PROTA[c]

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

Tubulin is a protein involved in a range of cellular processes, including cell division and axonal transport, due to its ability to dynamically polymerize into microtubules, which can be modulated by small molecules, making it an attractive target for research into cancer and neurodegeneration.

In addition to seven conventionally known small molecule binding sites[1,2], a recent fragment screening campaign has highlighted eleven additional fragment binding sites[3]. Further growing of one of the fragments produced a potent tubulin polymerization modulator named *Todalam*[4].

In this study, our first objective was to identify alternative chemotypes targeting the *Todalam* binding site. We conducted a virtual screening campaign that included substructure search, pharmacophore modeling, and docking. The resulting virtual hits were validated using X-ray crystallography studies and in vitro microtubule polymerization assays, leading to the identification of potent hits.

Recognizing the proximity of a cysteine residue to the native ligand's binding moiety, our second objective was to design a covalent binder targeting the *Todalam* site as a molecular probe to study microtubule dynamics and breakdown associated with neurodegeneration. We performed virtual screening of commercial fragment libraries, looking for fragments with reactive functional groups that could produce cysteine-targeting derivatives of potent hits found by scaffold hopping. The top-scoring virtual hits were then thoroughly filtered to ensure they could form the same protein-ligand contacts as *Todalam*. Their stability and orientation within the binding pocket was assessed by molecular dynamics simulations. Virtual hits were validated using X-ray crystallography studies and in vitro microtubule polymerization assays.

Our work has resulted in the identification of several novel scaffolds that bind to the *Todalam* binding site and modulate tubulin's polymerization. Several Cys-targeting molecules have been shown to enter the site, although further work is required to promote a covalent reaction between the ligands and the cysteine residue. This research was carried out as part of collaboration within a European ITN "TubInTrain" supported by the MSCA Horizon 2020 programme.

References

[1] Steinmetz M. O., Prota A. E., Trends Cell Biol (2018), 28, 776-792;

[2] Matthew S., Chen Q. Y., Ratnayake R., Fermaintt C. S., Lucena-Agell D., Bonato F., Prota A. E., Lim S. T., Wang X., Diaz J. F., Risinger A. L., Paul V. J., Oliva M. A., Luesch H., Proc Natl Acad Sci USA (2021), 118.

[3] Mühlethaler T., Gioia D., Prota A. E., Sharpe M. E., Cavalli A., Steinmetz M. O., Angew Chem Int Ed Engl (2021), 60, 13331-13342.

[4] Mühlethaler T., Milanos L., Ortega J.A., Blum T.B., Gioia D., Roy B., Prota A.E., Cavalli A., Steinmetz M.O. Angew. Chemie Int. Ed. 61, (2022).

ACCESSING TIMESCALES FAR BEYOND MD: LANGEVIN EQUATION MODELING OF PROTEIN-LIGAND (UN)BINDING KINETICS

STEFFEN WOLF[a]

[a] Biomolecular Dynamics, Institute of Physics, University of Freiburg, Freiburg, Germany

Predicting protein-ligand (un)binding kinetics is of high interest for pharmaceutical research, as slow unbinding kinetics have been linked to improved drug efficacy[1]. However, the respective rates on the order of seconds to hours are significantly beyond the capabilities of molecular dynamics simulations. To nevertheless computationally access such timescales of biomolecular complexes, we use a combination of coarse-graining system dynamics via dissipation-corrected targeted molecular dynamics simulations[2] (dcTMD) to derive Langevin models along a suitable (un) binding reaction coordinate, and temperature-boosted Langevin equation simulations[3] to then access the desired process timescales. Usually, the Langevin models have to be parameterized along each possible diffusion path, from which a global kinetic constant has to be calculated. We have derived a robust theoretical framework[4] that provides the basis for these calculations, and have developed approaches for path detection based on dimensionality reduction as well as machine learning approaches[5]. Free energies ΔF and friction factors Γ derived via dcTMD offer complementary information on the molecular discriminants that define (un)binding rates, and while the transition state in ΔF is mostly defined by enthalpic contributions such as electrostatic interactions, Γ surprisingly correlates with well with ligand and binding site hydration events[3,6], offering approaches to design drugs with specific kinetic profiles. Lastly, I will show how we are able to predict process kinetics on the order of several hours not only for simple model systems such as the trypsin-benzamidine complex, but for Hsp90, the streptavidin-biotin complex[6], as well as for GPCRs such as the A2 adenosine receptor, and give an outlook on the applicability of our approach to monitoring protein conformational transitions and obtain the accompanying rates.

References

[1] Copeland R A, Pompliano D L, Meek T D. Drug-target residence time and its implications for lead optimization, Nat. Rev. Drug Discov., 2006, 5-730.

[2] Wolf S, Stock G. Targeted Molecular Dynamics Calculations of Free Energy Profiles Using a Nonequilibrium Friction Correction, J. Chem. Theory Comput., 2018, 14-6175.

[3] Wolf S, Lickert B, Bray S, Stock G. Multisecond ligand dissociation dynamics from atomistic simulations, Nat. Commun., 2020, 11-2918.

[4] Wolf S, Post M, Stock G. Path separation of dissipation-corrected targeted molecular dynamics simulations of protein–ligand unbinding, J. Chem. Phys., 2023, 158-124106.

[5] Bray S, Tänzel V, Wolf S. Ligand Unbinding Pathway and Mechanism Analysis Assisted by Machine Learning and Graph Methods. J. Chem. Inf. Model., 2022, 62-4591.

[6] Cai W, Jäger M, Bullerjahn J T, Hugel T, Wolf S, Balzer B N. Anisotropic Friction in a Ligand-Protein Complex, Nano Lett., 2023, http://doi.org/10.1021/acs.nanolett.2c04632.

INVESTIGATING EARLY EVENTS OF HIRE1α ACTIVATION: INSIGHTS FROM MOLECULAR DYNAMICS SIMULATIONS

ELENA SPINETTI[a], JAN STUKE[b], G. ELIF KARAGÖZ[c], ROBERTO COVINO[d]

[a] Frankfurt Insitute for Advanced Studies, 60438 Frankfurt am Main, Germany (spinetti@fias.uni-frankfurt.de)
[b] Max Planck Institute of Biophysics, 60438 Frankfurt am Main, Germany
[c] Max Perutz Labs, Vienna Biocenter (VBC), 1030 Vienna, Austria
[d] Frankfurt Insitute for Advanced Studies & International Max Planck Research
(covino@fias.uni-frankfurt.de)

The Unfolded Protein Response (UPR) is a cellular process that helps eukaryotic cells respond to stress conditions in the endoplasmic reticulum (ER)[1]. The UPR is essential for maintaining health and preventing disease[2]. When stress sensors on the ER membrane are activated, the UPR is initiated[3]. The most evolutionarily conserved sensor is the membrane protein IRE1, which triggers the UPR signaling cascade by forming dimers and larger assemblies[1]. Although IRE1 has been shown to exist as a dimer in non-stress conditions[4] and interact with unfolded polypeptides via its luminal domain (LD)[5], the mechanism by which the LD dimer directly binds to unfolded proteins in humans remains unknown.

Our study aimed to shed light on the early events of human IRE1α (hIRE1α) activation, specifically how the LD dimer detects unfolded proteins and propagates the signal. We used molecular dynamic (MD) simulations to investigate the stability of the LD dimer and its interactions with unfolded polypeptides at the atomic level of resolution. We used the deposited PDB structure of the human dimer[6] and an AlphaFold prediction[7,8] to explore the dynamics of the hIRE1α LD dimer.

Our findings indicated that the dimeric interface of the hIRE1α LD remained stable in solution, primarily due to the presence of four hydrogen bonds in the antiparallel beta-sheet interface and an adjacent disordered region that was captured in a folded conformation by AlphaFold. Additionally, we discovered that specific unfolded polypeptides could bind to the LD dimer surface through multiple interaction sites. This finding is in line with a previous model which postulated that peptides should bind in the central part of the dimer as in the yeast Ire1[9], but our simulations provide a mechanistic model for direct binding with a higher level of detail.

In conclusion, our MD simulations of the dimeric structure of the hIRE1α LD dimer demonstrated that it could constitutively form a dimer, even in the absence of ER stress. Our results also showed that specific unfolded polypeptides could stably bind to the surface of the LD dimer, even if the central groove was unable to accommodate them inside. These new findings support a model in which the direct interaction between IRE1 and unfolded proteins is a critical initial step towards the formation of clusters. Overall, our study provides new insights into the molecular mechanisms underlying IRE1 activation and may have implications for the development of therapies targeting the UPR pathway.

References

[1] Walter P, Ron D. The Unfolded Protein Response: From Stress Pathway to Homeostatic Regulation. Science. 2011;334(6059):1081-1086. doi:10.1126/science.1209038

[2] Oakes SA, Papa FR. The role of endoplasmic reticulum stress in human pathology. Annu Rev Pathol Mech Dis. 2015;10:173-194. doi:10.1146/annurev-pathol-012513-104649

[3] Radanović T, Ernst R. The Unfolded Protein Response as a Guardian of the Secretory Pathway. Cells. 2021;10(11):2965. doi:10.3390/cells10112965

[4] Belyy V, Zuazo-Gaztelu I, Alamban A, Ashkenazi A, Walter P. Endoplasmic reticulum stress activates human IRE1α through reversible assembly of inactive dimers into small oligomers. eLife. 2022;11. doi:10.7554/elife.74342

[5] Karagöz GE, Acosta-Alvear D, Walter P. The unfolded protein response: Detecting and responding to fluctuations in the protein-folding capacity of the endoplasmic reticulum. Cold Spring Harb Perspect Biol. 2019;11(9). doi:10.1101/cshperspect.a033886

[6] Jiahai Zhou, Chuan Yin Liu, Sung Hoon Back, et al. The crystal structure of human IRE1 luminal domain reveals a conserved dimerization interface required for activation of the unfolded protein response. PNAS 103.39 (2006): 14343-14348. doi: 10.1073/pnas.0606480103

[7] Evans R, O'Neill M, Pritzel A, et al. Protein complex prediction with AlphaFold-Multimer. BioRxiv. 2021: 2021-10. doi: 10.1101/2021.10.04.463034

[8] Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. ColabFold: Making Protein folding accessible to all. Nat Methods. 2022, 19.6: 679-682.. doi:10.1038/s41592-022-01488-1

[9] Elif Karagöz G, Acosta-Alvear D, Nguyen HT, Lee CP, Chu F, Walter P. An unfolded protein-induced conformational switch activates mammalian IRE1. Elife. 2017: e30700. doi:10.7554/eLife.30700.001

NEURAL NETWORK LEARNING THE RECOGNITION DYNAMICS BY CONDENSATES

YEHOR TUCHKOV[a]

[a] University of Mainz

Here we apply the variational approach for Markov processes (VAMP) to develop a deep learning framework for molecular kinetics using neural networks, dubbed VAMPnets. A VAMPnet encodes the entire mapping from molecular coordinates to Markov states and is implemented for condensate-peptide binding simulation, which is of particular interest for analysing drugs binding abilities. The method is tested for both short and long chains and the limitations of relevance of VAMP framework are deeply investigated for such trajectories.

References

[1] VAMPnets for deep learning of molecular kinetics Andreas Mardt1 , Luca Pasquali1 , Hao Wu1 & Frank Noé 1

[2] GraphVAMPNet, using graph neural networks and variational approach to Markov processes for dynamical modeling of biomolecules Cite as: J. Chem. Phys. 156, 184103 (2022)

VISCOSITY OF FLEXIBLE AND SEMIFLEXIBLE RING MELTS – FROM MOLECULAR ORIGINS TO FLOW INDUCED SEGREGATION

RANAJAY DATTA[a] , FABIAN BERRESSEM[a], FRIEDERIKE SCHMID[a]
ARASH NIKOUBASHMAN[a], PETER VIRNAU[a]

[a] Institute of Physics, Johannes Gutenberg University, Mainz, Germany

We investigate with numerical simulations the molecular origin of viscosity in melts of flexible and semiflexible oligomer rings in comparison to corresponding systems with linear chains. The strong increase of viscosity with ring stiffness is linked to the formation of entangled clusters which dissolve under shear. This together with an orientation of rings in shear direction lead to pronounced shear-thinning and non-Newtonian behavior. Viscosity in linear chains on the other hand is associated with entanglements between chains which also dissolve under shear. While mixtures of semiflexible rings mix under equilibrium conditions, differences in rheological properties induce separation under flow. This phenomenon has potential applications in microfluidic devices and could be used to segregate ring polymers of similar mass and chemical composition.

BIAS-FORCE GUIDED SIMULATIONS COMBINED WITH EXPERIMENTAL VALIDATIONS TOWARDS GPR17 MODULATORS IDENTIFICATION

PARTHIBAN MARIMUTHU[a], SANA KARI[b], AKSHAYA MURUGESAN[b]
THIYAGARAJAN RAMESH[b], SRIVATSAN KIDAMBI[c]
JAMOLIDDIN RAZZOKOV[d,e,f,g,h], CHANDRABOSE SELVARAJ[i]
MEENAKSHISUNDARAM KANDHAVELU[b]

[a] Pharmaceutical science laboratory (PSL – Pharmacy) and Structural Bioinformatics Laboratory (SBL– Biochemistry), Faculty of Science and Engineering, Åbo Akademi University, FI-20520 Turku, Finland.
[b] Molecular Signaling Group, Faculty of Medicine and Health Technology, Tampere University and BioMediTech, P.O.Box 553, 33101, Tampere, Finland.
[c] Department of Chemical and Biomolecular Engineering, University of Nebraska-Lincoln, 820 N 16th Street, 207 Othmer Hall, NE, 68588, USA.
[d] Institute of Fundamental and Applied Research, National Research University TIIAME, Kori Niyoziy
39, 100000, Tashkent, Uzbekistan.
[e] College of Engineering, Akfa University, Milliy Bog Street 264, 111221, Tashkent Uzbekistan.
[f] Institute of Material Sciences, Academy of Sciences, Chingiz Aytmatov 2b, 100084, Tashkent, Uzbekistan.
[g] Department of Physics, National University of Uzbekistan, Universitet 4, 100174, Tashkent, Uzbekistan.
[h] Laboratory of Experimental Biophysics, Centre for Advanced Technologies, Universitet 7, 100174, Tashkent, Uzbekistan.
[i] Department of Biotechnology, Division of Research and Innovation, Saveetha School of Engineering, SIMATS, Chennai – 602105, Tamil Nadu, India.

Glioblastoma Multiforme (GBM) is known to be by far the most aggressive brain tumor to affect adults. The median survival rate of GBM patient's is < 15 months, while the GBM cells aggressively develop resistance to chemo- and radiotherapy with their self-renewal capacity which suggests the pressing need to develop novel preventative measures. We have recently proved that GPR17 —an orphan G protein-coupled receptor— is highly expressed on the GBM cell surface and it has a vital role to play in the disease progression. Despite the progress made on GBM downregulation, there still remain difficulties in developing a promising modulator for GPR17, till date. Here, we have performed robust virtual screening combined with biased-force pulling molecular dynamic (MD) simulations to predict high-affinity GPR17 modulators followed by experimental validation. Initially, the database containing 1379 FDA-approved drugs were screened against the orthosteric binding pocket of the GPR17. The external bias-potentials were then applied to the screened hits during the MD simulations which enabled to predict a spectrum of rupture peak force values that were used to select four approved drugs –ZINC000003792417 (Sacubitril), ZINC000014210457 (Victrelis), ZINC000001536109 (Pralatrexate) and ZINC000003925861 (Vorapaxar)– as top hits. The hits selected turns out to demonstrate unique dissociation pathways, interaction pattern, and change in polar network over time. Subsequently the selected hits with GPR17 were measured by inhibiting the forskolin-stimulated cAMP accumulation in GBM cell lines, LN229 and SNB19. The ex vivo validations shows that Sacubitril drug can act as a full agonist, while Vorapaxar functions as a partial agonist for GPR17. The pEC 50 of Sacubitril was identified as 4.841 and 4.661 for LN229 and SNB19, respectively. Small interference of the RNA (siRNA)– silenced the GPR17 to further validate the targeted binding of Sacubitril with GPR17. In the current investigation, we have identified new repurposable GPR17 specific drugs which are likely to increase the opportunity to treat orphan deadly diseases.

References

[1] Chen, Y., et al., The oligodendrocyte-specific G protein-coupled receptor GPR17 is a cell intrinsic timer of myelination. Nat Neurosci, 2009. 12(11): p. 1398-406.
[2] Simon, K., et al., The Orphan G Protein-coupled Receptor GPR17 Negatively Regulates Oligodendrocyte Differentiation via Gαi/o and Its Downstream Effector Molecules. J Biol Chem, 2016. 291(2): p. 705-18.
[3] Marucci, G., et al., The G Protein-Coupled Receptor GPR17: Overview and Update. ChemMedChem, 2016. 11(23): p. 2567-2574.

ASSESSING PASSIVE PERMEABILITY OF DRUG RELEASING LIPOSOMES CONTAINING SATURATED FATTY ACIDS WITH MOLECULAR DYNAMICS SIMULATIONS

SAMANEH DAVOUDI[a], AN GHYSELS[a]

[a] IBiTech - Biommeda group, Ghent University, Corneel Heymanslaan 10, Block B-entrance 36, 9000 Gent, Belgium

liposomes are an advanced delivery systems in cancer treatment[1]. A new generation of pH-sensitive liposomes that use fatty acids (FAs) as a trigger for drug release in tumor tissues has been introduced[2]. FAs are known as permeability enhancers in the intestinal epithelium[3]. They are also known to disrupt lipid tail packing in POPC bilayer and narrow the distance between them[4]. The question then arises whether FAs in liposomes might negatively affect the drug release half-life, the time it takes for the drug to circulate in the bloodstream. The passive permeability of the drug through the liposome's membrane is a crucial parameter for a timely drug release. Therefore, we investigate how two liposome parameters, i.e. the curvature and the lipid composition, affect the passive permeability. Flat bilayers and three liposomes with different diameters are simulated using coarse-grained molecular dynamics. The 23 simulated systems have varying lipid composition of dipalmitoylphosphatidylcholine (DPPC), cholesterol, and two protonated or deprotonated saturated FAs. The water permeability is determined using the counting crossings method. Moreover, two types of permeants with different free energy profile across the membrane, i.e. well and barrier, are added to the simulation to investigate the effect of the permeant's solubility in the membrane[5]. A new definition for liposome permeability was used, which was introduced in our recent paper to allow for comparison with flat bilayers[6].

In our simulations, curvature increases the permeability of water by reducing the membrane thickness. Water is a molecule with a single high free energy barrier in the middle of the bilayer. The two other types of permeants have a free energy profile with a mixture of wells and barriers in the headgroup region of the membrane. Unlike for water, a high curvature decreases the permeability for these two permeants. Overall, the permeability changes are limited to about 12% for the most curved membranes. Next, the composition of the liposomes is varied by incorporating FA and/or cholesterol. FAs decrease both the APL and thickness of the membrane, resulting in an increase in the permeability of all permeants. Cholesterol shows a similar effect on APL, but it has an opposite impact on the membrane thickness and permeability. FAs and cholesterol thus have an opposing effect on the permeability, where cholesterol's effect is slightly stronger. It can be concluded that the passive drug release from a pH-sensitive liposome does not seem to be significantly affected by the presence of FAs.

References

[1] Filipczak N. et al. Recent advancements in liposome technology, Adv. Drug Deliv. Rev., 2020, 4-22.

[2] Bonhenry, D. et al. Effects of phospholipid composition on the transfer of a small cationic peptide across a model biological membrane, J. Chem. Theory Comput., 2013, 5675-5684.

[3] Shakhawath H. et al. Influence of Bile Composition on Membrane Incorporation of Transient Permeability Enhancers, Mol. Pharm., 2020, 4226-4240.

[4] Kneiszl R. et al. In Silico-Based Experiments on Mechanistic Interactions between Several Intestinal Permeation Enhancers with a Lipid Bilayer Model, Mol. Pharm., 2021, 124-137.

[5] Davoudi S. et al. Sampling efficiency of the counting method for permeability calculations estimated with the inhomogeneous solubility–diffusion model, J. Chem. Phys., 2021, 054106.

[6] Davoudi S. et al. Defining permeability of curved membranes in molecular dynamics simulations, Biophys. J., 2023, 1–10.