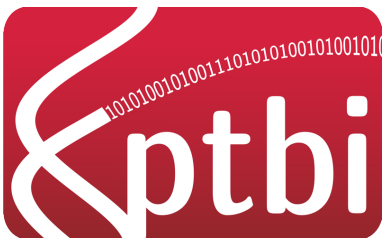
Bioinformatics, Artificial Intelligence and Forensic Science (BARIFORCE)

Book of Abstracts

29 June 2023, Toruń, Poland



Ministry of Education and Science
Republic of Poland

Projekt dofinansowany ze środków budżetu państwa, przyznanych przez Ministra Edukacji i Nauki w ramach Programu "Doskonała Nauka II - wsparcie konferencji naukowych" (nr wniosku: KONF/SP/0440/2023/01)

PROGRAM COMMITEE:

- Prof. dr hab. Wieslaw Nowak – Chair (*NCU, Torun*)
- Prof. dr hab. Jacek Błażewicz (*Politechnika Poznańska*)
- Prof. dr hab. Janusz Bujnicki (*IICMB, Warszawa*)
- Prof. dr hab. Jerzy Tiuryn (*Uniwersytet Warszawski*)
- Prof. dr hab. Jarosław Meller (*University of Cincinnati, USA*)
- Dr hab Witold Rudnicki (*Uniwersytet Białostocki*)
- Dr hab. inż. Paweł Łabaj (*Małopolskie Centrum Biotechnologii Uniwersytetu Jagiellońskiego*)

LOCAL ORGANIZING COMMITTEE:

- Dr. Katarzyna Walczewska-Szewc – Chair (*NCU, Torun*)
- Prof. Wieslaw Nowak – Co-Chair (*NCU*)
- Dr hab. inż. Paweł Łabaj (*Małopolskie Centrum Biotechnologii Uniwersytetu Jagiellońskiego*)
- Dr. Jakub Rydzewski (*NCU*)
- Beata Niklas (*NCU*)
- Sylwia Czach (*NCU*)
- Julia Berdychowska (*NCU*)
- Thiliban Manivarma (*NCU*)
- Tugce Gokdemir (*NCU*)

Contents

Lectures **1**
Marcin Woźniak . . . 2
Session 1: 13:30
Andrzej Polański . . . 3
Session 1: 14:15
Joanna Polańska . . . 4
Session 1: 14:45
Paweł Łabaj . . . 5
Session 2: 15:45
Michał Kowalski . . . 6
Session 2: 16:15
Marcin Gołebiewski . . . 7
Session 2: 16:45
Piotr Łukasiak . . . 8
Session 3: 18:30
TBA . . . 9
Session 3: 19:00
Maciej Antczak . . . 10
Session 4: 9:00
Aleksanda Cupriak . . . 11
Session 4: 9:20
Jan Brezovsky . . . 12
Session 4: 9:40
Adrian Roitberg . . . 13
Session 5: 10:15
Jarosław Meller . . . 14
Session 5: 11:00
Krzysztof Kuczera . . . 15
Session 5: 11:30
Rafał Madaj . . . 16
Session 5: 12:00
Piotr Stomma . . . 17
Session 5: 16:00
Witold Rudnicki . . . 18
Session 5: 16:30
Scott Tighe . . . 19
Session 5: 17:00

Posters **20**
Marek Justyna . . . 21

Beata Sokołowska . . . 22

Sylwia Czach . . . 23

Tuğçe Gökdemir . . . 24

Beata Niklas . . . 25

Thiliban Manivarma . . . 26

Jan Pielesiak . . . 27

Bartosz Adamczyk . . . 28

Tomasz Zok . . . 29

Michał Budnik . . . 30

Wojciech Warmuz . . . 31

Martyna Czarnomska 32
Marlena Osipowicz 33
Małgorzata Orłowska 34
Karolina Mikulska-Ruminska 35
Igor Marchlewski 36
Francisco Carrascoza 37
Julia Berdychowska 38
Natalia Szóstak 39

Lectures

Introduction to forensic genetics and genomics

Marcin Woźniak[1] and Urszula Rogalla-Ładniak[1]

[1]Chair of Forensic Genetics, Collagium Medicum of the Nicolaus Copernicus University, Bydgoszcz, Poland

The use of genetic profiling in forensic science dates back to the 1980s. From the very beginning, the techniques of genetic analysis used for the purposes of the legal system have undergone continuous evolution towards ever higher sensitivity of the methods used and wider possibilities of their application. The advent of the genomic era accelerated the changes and opened unprecedented opportunities for forensic genetics. The aim of the proposed lecture will be a synthetic discussion of the changes that have taken place in forensic genetics due to the increased use of NGS/ bioinformatics and those that will occur in the near future.

Email: marcinw@cm.umk.pl

Propagation of weakly advantageous somatic mutations in tumour growth

Andrzej Polanski[1]

[1]Silesian University of Technology

Somatic mutations in the DNA of cancer cells belong to two classes, rare drivers with strong, causative effect on cancer development, and much more abundant passengers with either small or no effect on cancer development. One frequent opinion in the literature on cancer genomics is that all passenger mutations are fully neutral and have no effect on tumour progression/evolution. However increasing number of studies is bringing experimental evidence that accumulated passenger mutations can impact cancer evolution, parallelly to drivers. Genomic locations and frequencies of somatic passenger mutations can be used to construct molecular signatures to distinguish between cancer types and their progression scenarios. Statistical genomics supported by quantitative molecular functional impact scores applied to positions of passenger mutations contradict hypothesis of neutrality of passenger mutations. Scenario of cancer growth initiated by events of chromosomal instability is requires the role of mildly advantageous mutations.

In the presentation propose a mathematical/simulation model for the scenario where a large number of passenger mutations of weak advantageous effect cause the growth of tumour. We confront the results of modelling to some of the available experimental data on tumour growth, DNA sequencing data on the progression of cancer and experimental data on the speed of tumour growth.

Email: andrzej.polanski@polsl.pl

CIRCA: COVID-19 screening supporting system as an example of the ML-based classification system for heterogeneous dataset

Joanna Polanska[1]

[1]Department of Data Science and Engineering, Silesian University of Technology, Gliwice, Poland

Chest X-rays are a standard diagnostic tool for respiratory diseases. This is also the case for COVID-19. To detect cases more quickly and monitor disease progression more accurately, supervised deep learning techniques can be used. Unfortunately, several systems have shown high accuracy in development, but have not performed well in clinical use. This was due to their poor generalisation, a perennial problem in AI development. The heterogeneity of the training data sets, both technical and clinical, may be a partial explanation for this phenomenon.

We propose a pipeline called CIRCA (https://circa.aei.polsl.pl) for a CXR-based COVID-19 screening support system developed on the basis of six different publicly available datasets. In addition to image standardisation and lung segmentation, CIRCA includes quantitative assessment of data heterogeneity and a hierarchical three-class (normal, COVID-19, non-COVID-19 pneumonia) decision system using convolutional network and radiomic features.

To address the heterogeneity of the datasets, the CXRs were bijectively projected into the 2D domain by performing Uniform Manifold Approximation and Projection (UMAP) embedding on the CNN-based neural features (nUMAP) from the pre-last layer of the pre-trained classification neural network. Comprehensive analysis of the multimodality of the density distribution in the nUMAP domain and its relation to the original image properties was used to identify the main confounders. nUMAP revealed an irrespective of the architecture hidden bias of neural networks towards image resolution, which regular upsampling cannot compensate. The application of advanced deep learning based super-resolution networks can partially mitigate it.

Gaussian mixture modelling applied to nUMAP-based image projection identified three radiologically distinct subtypes for each class, named 1, 2 and 3, where 1 represents class-specific typical images, and 2 and 3 represent atypical images. Proper training design and balancing across subtypes allowed a robust classification system to be built. In the hold-out test set, the subtype-specific cross-dataset classification NPV ranged from 95% to 100%, with PPV from 86% to 100% for all subtypes except N3 (early stage or convalescent) and both C3 and P3 (probable co-occurrence of viral and bacterial pneumonia). Similar results were obtained when an independent set of CXRs was used. The proposed pipeline could be easily adapted to other clinically heterogeneous diseases.

Soil microbiome analysis forensic tool

Paweł Łabaj[1]

[1]Malopolska Centre of Biotechnology, Jagiellonian University, Krakow, Poland

015 just called and wants its technology back – what metagenomics can gain from techniques which are not tredny anymore

Michał Kowalski[1]

[1]Malopolska Centre of Biotechnology, Jagiellonian University, Krakow, Poland

R pipeline for analysis of metagenomic/metatranscriptomic amplicon reads

Marcin Gołębiewski[1,2]

[1]Chair of Plant Physiology and Biotechnology, Faculty of Biological and Veterinary Sciences, Nicolaus Copernicus University
[2]Centre for Modern Interdisciplinary Technologies, Nicolaus Copernicus University

Comparing microbial communities to draw conclusions on state of samples in which they thrive is a common methodology in microbial ecology. Usually the communities are assessed by means of marker sequence amplicon sequencing, and 16S rRNA gene fragments are the most common markers. Obtaining meaningful results requires multiple steps both on the experimental and computational level. I will briefly outline the experimental methodology and present a pipeline implemented in R which can be used to process 16S rRNA gene fragment amplicon sequencing reads coming from commonly used Illumina platforms (MiSeq and NovaSeq).

Email: mgoleb@umk.pl

Genomic map of Poland – computational challenges, opportunities and chances

Piotr Łukasiak[1]

[1]Poznań University of Technology, Poznań, Poland

TBA

TBA[1]

[1]TBA

On entanglements of structure elements in RNA 3D structures

Maciej Antczak[1,2], Mariusz Popend[2], Joanna Sarzynsk[2], Tomasz Zok[1], and Marta Szachniuk[1,2]

[1]Institute of Computing Science, Poznan University of Technology, Poznan, Poland
[2]Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznan, Poland

3D RNA structures contribute to the explanation of their biological functions. However, the 3D structures of the most biologically significant RNAs are currently unknown. Thus, reliable in silico predictions are often crucial for speeding up biochemical experiments or drug design processes.

3D RNA models generated in silico are often inaccurate and exhibit deviations from the native structure, especially if they form complex motifs by long-range interactions. Most state-of-the-art methods predict undesirable entangled RNA fragments (i.e., entanglements) during the modeling of pseudoknots.

We define entanglements of structure elements as spatial arrangements involving two structural elements, where at least one punctures the other [1]. Puncture takes place when one structure element intersects the area within the other. We distinguish two superior classes of entanglements, interlaces and lassos, and subclasses characterized by element types - loops, dinucleotide steps, and open single-stranded fragments.

Here, we briefly describe the RNAspider web server - the first tool to automatically identify, describe, and depict entanglements in 3D RNA structures [2]. Moreover, we discuss the results of entanglement analysis performed for experimentally determined 3D RNA structures stored in the Protein Data Bank [2] and RNA 3D models submitted to the RNA-Puzzles contest [1].

[1] M. Popenda et al., Entanglements of structure elements revealed in RNA 3D models, Nucleic Acids Research, 2021, 49(17), pp. 9625–9632 (doi:10.1093/nar/gkab716).

[2] K. Luwanski et al., RNAspider: a webserver to analyze entanglements in RNA 3D structures, Nucleic Acids Research, 2022, 50(W1), pp. W663–W669 (doi:10.1093/nar/gkac218).

Email: maciej.antczak@cs.put.poznan.pl

Predicting the formation of chromatin loops using genomic data

Aleksandra Cupriak[1] and Aleksander Jankowski[1]

[1]Faculty of Mathematics, Informatics and Mechanics, University of Warsaw

Three-dimensional chromatin structure has an important role in regulating gene expression. One of the elements of this structure are chromatin loops, formed when two genomic loci located at a large distance measured along the DNA strand are spatially close to each other. Chromatin loops allow spatial contact of distal regulatory elements such as enhancers and promoters. Understanding mechanisms leading to the formation of chromatin loops is an important step in disentangling regulatory mechanisms. The goal of this work is to create a classifier that uses genomic data to predict whether two loci in the genome form a chromatin loop.

For prediction, we used annotations of chromatin loops from Hi-C contact maps in 7 selected cell types in humans, and in embryonic and larval cells in D. melanogaster fruit fly. As features we used data from genome-scale experiments on chromatin accessibility (DNase-seq, single cell ATAC-seq) and protein binding (ChIP-seq). Additionally, we used DNA sequence information on transcription factor binding motifs. We also tested different methods for generating negative examples: either by randomly pairing known looping loci, or by sampling regions of accessible chromatin. In our approach, we tried several machine learning models for the task of binary prediction (logistic regression, decision tree, random forest, LightGBM). Finally, we used the trained models to determine which features are the most relevant to predicting the formation of chromatin loops.

The models performed well when the negative examples were generated by sampling regions of accessible chromatin, but the performance dropped when the negative examples were obtained by randomly pairing known looping loci. This suggests that our models accurately identify the regions that constitute the anchors of chromatin loops, but they cannot determine which two anchors would be looping. We found that LightGBM was the best performing classifier, with average AUC-ROC 0.862 for the main model using human data, and 0.947 for fruit fly. Using a cross-validation scheme, we confirmed that the formation of loops in a given cell type can be predicted using a model trained on loops from other cell types. We also confirmed that in human data, CTCF binding and chromatin accessibility are important features affecting loop formation. The results were similar in case of fruit fly, although we could not pinpoint a single protein whose binding could be a satisfactory predictor alone.

Email: ac446520@students.mimuw.edu.pl

Rational engineering of dynamical enzyme pockets for leashing gram-negative pathogenic bacteria

Jan Brezovsky[1,2], Bartlomiej Surpeta[1,2], and Michal Grulich[3]

[1]Laboratory of Biomolecular Interactions and Transport, Institute of Molecular Biology and Biotechnology, Faculty of biology, Adam Mickiewicz University, Poznan
[2]International Institute of Molecular and Cell Biology in Warsaw
[3]Laboratory of Modulation of Gene Expression, Institute of Microbiology of Academy of Sciences of the Czech Republic, Prague, Czech Republic

Rapidly increasing bacterial resistance to antibiotics is a global health problem prioritized by the World Health Organization. Quorum quenching represents one of the alternatives to antibiotic use, in which, the bacterial communication through signaling molecules is disrupted by enzymatic cleavage of these molecules, effectively reducing the expression of genes responsible for virulence factors and biofilm formation.

Here, we have engineered penicillin G acylase for improved degradation of signaling molecules used by several different bacterial species of pathogenic nature. As the substrate binding site is of transient nature, we have employed an ensemble-based computational design of its subsequent molecular gates. Next, we used molecular dynamics simulations to prioritize the most potent candidates for experimental validation, which confirmed computational predictions of enhanced activity and modulated specificity. In-depth experimental and computational characterization allowed us to understand the structural basis for the observed modulation in our constructs. Curiously, we have seen that either too dynamical or too rigid pockets resulted in a compromised activity of the construct with longer signaling molecules, revealing new challenges connected with targeting protein dynamics during their engineering. The predictive power of the presented workflow opens the way for further engineering campaigns toward applicable quorum quenching-based antibacterial agents.

The research was supported by National Science Centre, Poland, grant numbers 2017/26/E/NZ1/00548 and 2021/41/N/NZ2/01365. The computations were performed at the Poznan Supercomputing and Networking Center.

Email: janbre@amu.edu.pl

TBA

Adrian Roitberg[1]

[1]University of Florida, Gainesville, USA

Towards Multiomics Single Cell Resolution Atlas of ccRCC: Combining scRNA-seq and Spatial Transcriptomics to Identify and Characterize Cell Communities Connected to Cancer Progression

Jarowsław Meller[1]

[1]Cincinnati Children's Hospital Medical Center and University of Cincinnati, US

ccRCC tumors are archetypal examples of heterogeneous cancers characterized by a high degree of clonal heterogeneity, posing significant challenges for treatment of this deadly disease. While surgical resection by partial or radical nephrectomy is the first step in treatment of localized or locally advanced ccRCC tumors, there is a ∼50% risk of recurrence after surgical resection within five years with no clear classifiers predicting the risk of recurrence or therapeutic approaches for adjuvant settings.

We hypothesized that spatial organization of cancer cells exhibiting different transcriptomic and metabolic programs contributes to tumor advancement in ccRCC. Thus, determining this organization can provide novel insights into cancer progression, therapeutic vulnerabilities, and drug resistance mechanisms.

We have used 10x Genomics Visium spatial transcriptomics platform to determine spatial organization of 6 stage-3 tumors and map distinct clonal subpopulations identified through meta-analysis of several published scRNA-seq data sets. To that end, we developed a computational pipeline that integrates several established methods, including SpaceRanger, Seurat, and Harmony, and in-house developed modules, to facilitate joint analysis of scRNA-seq and spatial transcriptomic data.

Modeling protein secondary structure formation: Comparing folding pathways of α-helices vs. β-hairpins

Krzysztof Kuczera[1,2]

[1]Department of Chemistry, The University of Kansas, Lawrence, Kansas 66045, USA
[2]Department of Molecular Biosciences, The University of Kansas, Lawrence, Kansas 66045, USA

We have performed molecular dynamics (MD), replica-exchange (REMD) simulations and kinetic modeling to study folding pathways of model peptides forming helices and a hairpin. The simulation results are in good agreement with available experimental data on structure content and folding timescales, while providing novel microscopic insights into the fundamental processes of secondary structure formation. For helix-forming peptides, we investigated effects of sequence and length on folding. In alanine homopeptides (ALA)n of length $n = 5, 8, 15$, and 21 residues we find helix populations and relaxation times increasing from about 6% and 2 ns for ALA5 to about 60% and 500 ns for ALA21, and folding free energies decreasing linearly with the increasing number of residues. Helices tended to fold along multiple pathways, through various intermediates. Statistically the helices were most stable in the central region. However, the individual folding transitions tended to follow a random-walk path, with local dynamics involving correlated transitions of blocks of several consecutive hydrogen bonds. For the 16-residue GB1 hairpin peptide a folded fraction of 40% and global relaxation time of 1.8 were calculated at 320 K. Folding followed the zipper model with nucleation at the central turn followed by consecutive hydrogen bond formation. The transitions were highly cooperative, with all backbone hydrogen bonds and sidechain hydrophobic interactions forming essentially at the same time. Coarse grained kinetic modeling showed transitions to off-path intermediates and identified the transition state for hairpin formation, involving the formation of the central turn and first hairpin hydrogen bond. The 'hairpin' state was also found to be heterogenous, including the fully folded and partially folded in-register hairpins along the zipper pathway.

Side mutants of connexin proteins as a cause of non-functional gap junction beta channels - a structural bioinformatics approach

Rafal Madaj[1], Malgorzata Orlowska[1], Kamil Kaminski[1], Kamil Pawlicki[1], and Stanislaw Dunin-Horkawicz[2]

[1]Institute of Evolutionary Biology, Faculty of Biology, Biological and Chemical Research Centre, University of Warsaw, Żwirki i Wigury 101, 02-089, Warsaw, Poland,
[2]Department of Protein Evolution, Max Planck Institute for Biology Tübingen, Max-Planck-Ring 5, 72076 Tübingen, Germany

Gap Junction Beta proteins (GJB) are group of proteins (known as connexins) responsible for creating channels between adjacent cells. They are a key to cell-to-cell communication in several organs, shortening signaling paths by an order of magnitude. Their mutations cause multiple diseases, such as Charcot-Marie-Tooth type 1X, deafness dystonia syndrome (DFNA11), Erythrokeratodermia variabilis (EKV) or Non-syndromic hearing loss. While the majority of mutations causing these diseases are known, some underlying mechanisms are hitherto unelucidated.

Through structural bioinformatics study, using AlphaFold2, Rosetta and USAlign we managed to identify which mutants within GJB1 and GJB2 proteins are most likely to cause channel malfunctions, yielding in either improper folding or complex instability. Our approach is based on investigation on mutations occurring on monomer-monomer complexing, either in case of homodimer or heterodimeric complex formation. Surprisingly, some mutations have impact only in homodimer formation, being silent in case of heterodimer, despite having similar sequences.

This study may be a base for a connexin-wide study, for identification of underlying mechanism of improper dimer (hence, whole channel) formation.

Email: r.madaj@uw.edu.pl

Interpretable generative model of correlation structure of multidimensional biological datasets

Piotr Stomma[1] and Witold Rudnicki[1,2]

[1]University of Białystok, Institute of Computer Science
[2]University of Białystok, Computational Centre

Background Here we present the algorithm for the creation of high dimensional datasets that have a correlation structure similar to the experimental datasets arising from high-throughput methods of molecular biology, collectively known as omics. Formally, the problem can be stated as follows: given (real) dataset $\mathbf{X} = \{X_1, X_2 \dots X_n\}$, the goal is to get a "clone dataset" **X'**, correlated *like* **X**, where the distribution of all $X'_j \in$ **X'** is known. The standard methods for such a task, which are based on Cholesky factorisation of the original dataset, cannot be applied when the number of variables is higher than the number of samples. In this study, we follow the approach implemented in the WGCNA R package - a tool for weighted gene correlation network analysis. WGCNA also implements a simple method for simulating correlated data. To this end, it computes the principal components of each cluster and then obtains the simulated variables as the largest principal component with added noise. This approach faithfully replicates the basic correlation structure. However, the network topology of real data is significantly different from the topology of simulated data, especially in terms of the distribution of the clustering coefficient.

Methods We improve upon the original WGCNA approach by usign its main idea iteratively. We first build set of clusters at the certain threshold correlation level, and then iteratively apply the same approach within each cluster separately, using increased thresholds at each step. We have used this approach in conjunction with the clique-based clustering algorithm to obtain a simulation of a dataset of 8676 gene expression profiles.

Results and Discussion We have improved the low complexity of the basic variant, while keeping the partition similarity high, and it appears that using our method, as number of thresholds increases, the clustering coefficient distribution converges. Good fit of our model requires accurately capturing the correlation structure, making the model a useful analytical tool. Knowledge of exact distributions of X'_j make this method suitable for validation of classifiers, feature selection or clustering methods. Future work includes: finding optimal threshold value and an optimal balance between accurately replicating correlation network vs complexity of the resulting model.

Email: 16i@protonmail.com

Biomarkers discovery with machine learning and feature discovery

Witold R. Rudnicki 1[2], Radosław Piliszek[2], Oliwia Podbielska[1], Patrycja Wszeborowska[1], and Piotr Stomma[1]

[1]Institute of Computer Science, University of Białystok, Ciołkowskiego 1M, 15-245 Białystok
[2]Computational Centre, University of Białystok, Ciołkowskiego 1M, 15-245 Białystok

Background Modern technologies in molecular biology, so-called omics, are routinely used in studies of the biological background of multiple diseases, in particular cancer. They deliver detailed data with the state of the molecular system described with tens of thousands of variables, such as, for example, the gene expression level in the case of transcriptomics. However, the number of samples is limited both due to the high cost and difficulty of the experimental studies or, in the case of rare diseases, by a small number of cases overall. What is more, in most cases, the data is most often collected without any particular research hypothesis, apart from the hope that analysis of this data will lead to a deeper understanding of problems under scrutiny. Special methods are developed to deal with such data, including multiple stages of data cleaning, filtering and analysis. In particular, machine learning is used to build predictive models without understanding the underlying mechanisms.

Methods I will present an overview of the protocol for the discovery of biomarkers that involve multiple methods of feature selection, extensive application of cross-validation and application of the Random Forest classification algorithm. In particular, I will shortly describe our Robust Agglomerative Feature Selection algorithm, which is based on a new information-theoretic measure of dissimilarity between variables that takes into account the decision variable.

Results and Discussion I will present the result obtained in the analysis of multiple types of cancer. In particular, I will show the 7-gene signature for high-risk bladder cancer, discuss insight gained into molecular differences between post-transplant lymphatic disorder in EBV+ and EBV- patients, as well as molecular signatures between high- and low-risk patients in glioblastoma and clear cell renal cancer.

Email: W.Rudnicki@icm.edu.pl

Technical considerations for metagenomic analysis of low biomass environments

Scott Tighe[1]

[1]Vermont Integrative Genomics Lab, Burlingtion Vermont, USA

Posters

RNA 3D fold generation with denoising diffusion probabilistic models.

Marek Justyna[1], Maciej Antczak[1,2,3], and Marta Szachniuk[1,2,3]

[1]Institute of Computing Science, Poznan University of Technology, Poznan, Poland
[2]Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznan, Poland
[3]European Center for Bioinformatics and Genomics, Poznan University of Technology, Poznan, Poland

Denoising Diffusion Probabilistic Models (DDPM) [1] belong to a class of deep generative models that are currently very successful in image synthesis. They have also been applied in many other fields [2], for example, in biology for the generation of protein sequences and structures [3-4] or in drug design [5-6]. Here, we discuss the preliminary results of the approach adapting DDPM for RNA sequence and 3D fold generation. We transform the 3D structure of the RNA backbone into the torsion angle space. Next, we convert the NxM matrix into a colored-scale heat map; N is the number of nucleotides and M is the number of torsion angles in the RNA backbone. The data, represented as a matrix, are then modified by gradually adding a gaussian noise in the diffusion process. Finally, we train the Denoising model to reverse this process. Following these steps, we built the model with generative properties, able to generate new sequences and its torsion angles from the gaussian noise.

References

[1] Ho, J., Jain, A., & Abbeel, P. (2020). Denoising diffusion probabilistic models. Advances in Neural Information Processing Systems, 33, 6840-6851.

[2] Yang, L., Zhang, Z., Song, Y., et al. (2022). Diffusion models: A comprehensive survey of methods and applications. arXiv preprint arXiv:2209.00796.

[3] Anand, N., & Achim, T. (2022). Protein structure and sequence generation with equivariant denoising diffusion probabilistic models. arXiv preprint arXiv:2205.15019.

[4] Wu, K. E., Yang, K. et al. (2022). Protein structure generation via folding diffusion. arXiv preprint arXiv:2209.15611.

[5] Xu, M., Yu, L., et al. (2022). Geodiff: A geometric diffusion model for molecular conformation generation. arXiv preprint arXiv:2203.02923.

[6] Hoogeboom, E., Satorras, V. G., et al. (2022, June). Equivariant diffusion for molecule generation in 3d. In International Conference on Machine Learning (pp. 8867-8887). PMLR.

Email: mjustyna@cs.put.poznan.pl

Novel Technologies and Their Use in Neurodevelopmental Disorders

Beata Sokołowska[1], Ewa Sokołowska[2], Stanisław J. Chrapusta[3], Paweł Kowalczyk[4], and Dorota Sulejczak[3]

[1]Bioinformatics Laboratory, Mossakowski Medical Research Institute, Polish Academy of Sciences, Warsaw, Poland,
[2]Department of Developmental Psychology, Faculty of Social Sciences, Institute of Psychology, The John Paul II Catholic University of Lublin, Lublin, Poland,
[3]Department of Experimental Pharmacology, Mossakowski Medical Research Institute, Polish Academy of Sciences, Warsaw, Poland,
[4]Department of Animal Nutrition, The Kielanowski Institute of Animal Physiology and Nutrition, Polish Academy of Sciences, Jabłonna, Poland

The field of digital neuroscience and digital medicine deals with the use of innovative technologies as novel tools for scientific research and accurate measurement and intervention in the service of human health. Modeling/conceptual research and experimental/observational clinical studies (both human and animal) demonstrate a wide and growing range of applications of modern technologies, including not only virtual/augmented reality (VR/AR), but also artificial intelligence (AI), robotics and future metaverse concepts. IT/ICT may facilitate the development of new methods of diagnosis, assessment, treatment, and care coordination to provide a better quality of life for people with various disabilities and deficits.

This paper presents promising applications of VR environments (VREs) in neurodevelopmental disorders that have been designed according to expert guidelines. Many VR-based studies focus on attention deficit hyperactivity disorder (ADHD) and other related disorders, and propose VREs as an alternative to conventional approaches in the learning process of children with special educational needs, mainly in primary education. Importantly, VR diagnostic tests for ADHD are more effective than traditional neuropsychological approaches. Other interesting VR studies concern autism spectrum disorders (ASD). Individuals with ASD are characterized by impaired communication, poor social interactions, and deficits in the explicit expression of their affective states. Adaptive, physiology-sensitive, VR- and AI-based training systems may help improve the skills of trainees in all these fields.

It should be noted that the key issue of the transferability of learned social communication skills from VR-based environments to real-world situations remains to be resolved. However, the data obtained so far indicate that this is feasible.

Email: beta.sokolowska@imdik.pan.pl

Revised CHARMM force field parameters for the ground state and signalling state of the photoactive yellow protein photocycle

Sylwia Czach[1], Katarzyna Walczewska-Szewc[1], Jakub Rydzewski[1], and Wiesław Nowak[1]

[1]Institute of Physics, Faculty of Physics, Astronomy and Informatics, Nicolaus Copernicus University in Toruń, ul. Grudziądzka 5, 87-100 Toruń, Poland

Photoactive yellow protein (PYP) is a model blue-light-sensitive photoreceptor, responsible for the process of negative phototaxis, present in many bacterial species. PYP is a member of the PAS (Per-Arnt-Sim) domain superfamily. This domain is highly conserved The core consists of a five-stranded antiparallel β-sheet and several α-helices. The agent responsible for the photoactivation of PYP is the p-Coumaric acid (pCA) chromophore linked by a thio-ester bond to cysteine at position 69.

PYP undergoes a series of conformational changes due to the signal generated after photoexcitation of pCa. After photoexcitation, there is a conformational change of the chromophore, from the ground - trans to the excited - cis state. This transition induces a signal transduction pathway for conformational changes of the entire protein. These conformational changes follow a known photocycle, which starts the protein in the ground state (the dark state, pG). The final step of the PYP photocycle is the signal state (last lighted state, pB).

PYP due to the presence of the conserved PAS domain and its small size, provides a model system for research, both theoretical and experimental. However, the computational study of this system is hampered by the presence of p-Coumaric acid, which is a non-standard structural element for conventional force fields. In order to study the system by molecular dynamics, the existing force field must first be modified with parameters for the covalently bound ligand.

In this work, a modified CHARMM force field is presented. External servers and quantum-chemical calculation programmes were used to parameterise the chromophore. The correctness of the obtained parameters was checked by performing a series of molecular dynamics simulations of the system in solvent using GROMACS.

Email: s.czach@yahoo.com

Taking on Timescale Challenge of Rare Events: Theoretical Investigation of Protein Folding by Metadynamics

Tuğçe Gökdemir[1], Jakub Rydzewski[1], and Wiesław Nowak[1]

[1]Institute of Physics, Faculty of Physics, Astronomy and Informatics, Nicolaus Copernicus University, Grudziądzka 5, 87-100 Toruń, Poland

Molecular dynamics (MD) is a powerful tool to observe a diverse range of systems from biology to physics at the atomistic level. However, the reachable timescale of MD simulations is limited, and investigating long-timescale phenomena is difficult. In rare events (e.g., protein folding, chemical reactions, etc.), systems have several metastable states separated by high energy barriers that limit the sampling. To overcome this timescale challenge, several enhanced sampling algorithms have been proposed. Metadynamics (MetaD) is shown to be one of the promising enhanced sampling methods for the acceleration of MD simulations. In MetaD, a few degrees of freedom, called collective variables (CVs), are selected, and the free energy surface is obtained as a function of these CVs. Adding an external history-dependent bias potential in the CV space improves the sampling. In this study, the radius of gyration and end-to-end distance of carbon atoms have been chosen as CVs, and MetaD has been carried out on a chignolin protein. We demonstrate that MetaD can be used effectively in theoretical studies of protein folding.

Email: tugcegokdemir@doktorant.umk.pl

Toward overcoming pyrethroid resistance in mosquito control with sodium channel blocker insecticides

Beata Niklas[1], Jakub Rydzewski[1], Bruno Lapied[2], and Wiesław Nowak[1]

[1]Institute of Physics, Faculty of Physics, Astronomy and Informatics, Nicolaus Copernicus University, Grudziadzka 5, 87-100 Torun, Poland
[2]University Angers, INRAE, SIFCIR, SFR QUASAV, F-49045 Angers, France

Diseases spread by mosquitoes lead to death of 700,000 people each year. The main way to reduce transmission is vector control by biting prevention with chemicals. However, the most commonly used insecticides lose efficacy due to the growing resistance.

Voltage-gated sodium channels (VGSCs), membrane proteins responsible for the depolarizing phase of an action potential, are targeted by pyrethroids and sodium channel blocker insecticides (SCBIs). Although SCBIs – indoxacarb (a pre-insecticide bioactivated to DCJW in insects) and metaflumizone – are used in agriculture only, they emerge as promising candidates in mosquito control. Therefore, a thorough understanding of molecular mechanisms of SCBIs action is urgently needed to break the resistance and stop disease transmission.

In this study, by performing an extensive combination of equilibrium and enhanced sampling molecular dynamics simulations (3.2 μs in total), we found the DIII-DIV fenestration to be the most probable entry route of DCJW to the central cavity of mosquito VGSC. Results explain the role of the F1852T mutation found in resistant insects and the increased toxicity of DCJW compared to its bulkier parent compound, indoxacarb. We also delineated residues that contribute to both SCBIs and non-ester pyrethroid etofenprox binding and thus could be involved in the target site cross-resistance.

Email: beata.niklas@doktorant.umk.pl

Predicting Hot Spots using a Neural Network approach in 15LOX1-PEBP1 complex responsible for Ferroptosis

Thiliban Manivarma[1], Karolina Mikulska Ruminska[1], and Wieslaw Nowak[1]

[1]Nicolaus Copernicus University in Torun, Department of Biophysics, Torun, Poland

Ferroptosis is recently discovered mechanism of cells death induced by unbalanced regulation one out of three major metabolic pathways involving iron, thiols or lipids, Ferroptosis is connected with such diseases as Alzheimer's disease, Parkinson's disease, sepsis, kidney injury, cancer etc. The process involves enzymatic and non-enzymatic means where the specific mechanisms are still unclear. The enzymatic form of ferroptosis involves iron dependent peroxidation of lipids performed by lipoxygenase (LOX). Presence of products (lipid peroxides) leads to the cell death. LOX is membrane-associated enzyme which imports lipids from the membrane. For the activity, a promiscuous Phosphatidylethanolamine Binding Protein 1 (PEBP1) is recruited. In a recent paper, 15LOX1-PEBP1 complex was modelled using coarse grained approach in a simplified water environment (1). Here, we present artificial intelligence (AI) – Neural network model to find a set of crucial hotspots on 15LOX1-PEBP1 complex. The presented work is consist of Feature extraction, Deep learning classification, Model evaluation. The impact of this AI model would help us to target the protein-protein interface of 15LOX1-PEBP1 for further drug discovery process in Ferroptosis.

Acknowledgment: Support of National Science Centre, Poland, No 2019/35/D/ST4/02203 (KMR, TM), is acknowledged. IDUB NCU #MEMOBIT grant (WN)and ICNT UMK computer facilities are acknowledged as well.

1. Wenzel, S. E., et al. 2017. PEBP1 Wardens Ferroptosis by Enabling Lipoxygenase Generation of Lipid Death Signals. Cell. 171(3):628-641.e626.

Email: thilibanaussie@gmail.com

Analysis of non-canonical interactions and motifs in 2D/3D RNA structures - research plan

Jan Pielesiak[1]

[1]Institute of Computing Science, Poznan University of Technology, Poznan, Poland

RNA is a key particle in the life of all organisms and viruses, and it fulfills a myriad of roles. Thanks to its elasticity and the chemical construction of nucleotides, the common factor connecting different families of RNA is its unique structure. Base pairs in RNA are divided into two groups: canonical - similar to ones seen in DNA, and non-canonical, which are specific only to RNA. Non-canonical base pairs are the key to the creation of structural connections in RNA, which in turn are the biggest factor in the final shape of the particle. The existence of a huge number of different non-canonical types of RNA interactions and the rarity of occurrence of some of them poses a big challenge for its researchers.

The poster presents the plan for the development of new methods for complex structural motifs in RNA containing non-canonical interactions and multi-loops. The main purpose is to discover patterns of base pairing based on experimental 3D models of RNA structures. The secondary purpose is to create a graph model representing those patterns and use it to compare or sample RNA fragments.

A database filled automatically with non-canonical RNA structures saved as graphs will be created. This database will be the basis for training Graph Neural Networks.

Email: tdm137@gmail.com

WebTetrado: a webserver to explore quadruplexes in nucleic acid 3D structures

Bartosz Adamczyk[1], Michał Żurkowski[1], Marta Szachniuk[1,2], and Tomasz Żok[1]

[1]Institute of Computing Science, Poznan University of Technology, Poznan, Poland
[2]Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznan, Poland

Analyzing molecular structures is a time-consuming process. It requires various bioinformatic tools to perform calculations. It is also the case when exploring quadruplexes, specifc motifs occurring in nucleic acid structures. Automating this process with reliable computational methods solves the problem and improves the exploration experience.

We present WebTetrado, a web application that provides everything needed to discover and explore the properties of quadruplexes and tetrads. A set of backend functions read the 3D DNA/RNA structures in PDB or mmCIF format. Next, they identify base pair patterns to and classify tetrads and quadruplexes. WebTetrado computes and visualizes G4 structure parameters such as rise, twist, planarity, and chi angle. The system assigns tetrads to the ONZ and Webba da Silva classifcations and displays the results using the Mol* structure viewer. The web application is freely available at https://webtetrado.cs.put.poznan.pl.

This research was supported by grant 2019/35/B/ST6/03074 from the National Science Centre, Poland.

Email: bartosz.adamczyk@student.put.poznan.pl

ONQUADRO: a database of experimentally determined 3D quadruplex structures

Tomasz Zok[1], Natalia Kraszewska[1], Joanna Miskiewicz[1], Paulina Pielacinska[1], Michal Zurkowski[1], and Marta Szachniuk[1,2]

[1]Institute of Computing Science, Poznan University of Technology, Poznan, Poland
[2]Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznan, Poland

About 14 000 experimentally solved 3D structures of nucleic acids are currently available in the Protein Data Bank. Some of them contain quadruplexes, but researchers looking for them are limited to test-search in abstracts for terms like *G:G:G:G*, *G4*, or words with *tetra-*, *quadro-* or even *octa-* prefixes.

To address these issues, we present **a new database and webserver** named ONQUADRO. It is an automatically updated resource gathering **detailed information** about quadruplexes among 3D structures of nucleic acids. ONQUADRO displays each quadruplex on **several visualizations** and offers **three classification schemes:**

- Webba da Silva's loop-based,
- Webba da Silva's GBA-based, and
- ONZM depending on the nucleobase interactions.

The database also collects *tracts*, *tetrads*, and *loops* and calculates commonly used parameters such as **rise**, **twist**, or **planarity deviation**. In addition to detailed information about each quadruplex, users of ONQUADRO can also get to know statistics regarding the whole collection and its growth in time. We believe ONQUADRO is a hugely beneficial resource to the G4 scientific community.

Email: `mzurkowski@cs.put.poznan.pl`

DiffuRNA: A diffusion model for RNA structure prediction

Michał Budnik[1,2], Jakub Wawrzyniak[1], Łukasz Grala[1], Miłosz Kadziński[2], and Natalia Szóstak[1,3]

[1]TIDK Sp. z o.o., Poznan
[2]Institute of Computing Science, Poznan University of Technology
[3]Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznan

Ribonucleic acid (RNA) is known as a polymer molecule that converts genetic information from DNA into proteins. However, a type of RNA known as non-coding RNA has been found to be involved in more complex processes vital to our lives and health. Targeting these naturally occurring ncRNAs shows diagnostic and therapeutic potential for treatment of many diseases. These functions are determined by the three-dimensional shape that given sequence folds into. Due to high costs and required time of wet lab methods, *in silico* structure prediction has become a major area of interest in bioinformatics. Various algorithms have been proposed and tested for solving both secondary and tertiary structure prediction problems. These methods, though achieving satisfactory results for secondary structure prediction, over the years failed to achieve significant improvements in the 3D structure prediction. As an alternative, with development in both optimization methods and computational possibilities, machine learning based methods started emerging.

Following the successes in a closely related problem of protein folding and recent developments in deep learning, we propose a concept of *de novo* approach for RNA structure prediction called DiffuRNA. The solution concept is based on the diffusion model, which consists of two main steps to learn the data structure. First, the forward process, formally defined as a Markov Chain, gradually introduces Gaussian noise to the sample for T successive steps. In our case, the sample is a 3D representation of an RNA molecule in atomic resolution. The second step is the denoising process. Starting from the isotropic Gaussian noise, it reconstructs the input data for the same number of steps. The reconstruction process will be guided by providing conditioning information of the RNA sequence. The data for training and testing include high-resolution, non-redundant RNA structures obtained by experimental methods.

A major potential limitation of this method is the computational complexity of the model, which may be addressed by encoding the RNA structure to a latent space. Additionally, approaches using diffusion models in protein prediction have displayed a limited ability to generate realistic protein backbones. This may be improved upon with support of methods that operate on rigid-frame representations of residues rotational and translational equivariance.

Email: `michal.budnik@doctorate.put.poznan.pl`

Re-designing key interactions in cohesin-dockerin complex from thermophilic organism to enable its use in synthetic biology applications towards green recycling of polymers

Wojciech Warmuz[1,2] and Jan Brezovsky[1,2]

[1]Laboratory of Biomolecular Interactions and Transport, Department of Gene Expression, Institute of Molecular Biology and Biotechnology, Faculty of Biology, Adam Mickiewicz University in Poznan, Poznan, Poland
[2]Laboratory of Biomolecular Interactions and Transport, International Institute of Molecular and Cell Biology in Warsaw, Warsaw, Poland

Synthetic biology has gained increasing popularity as a potential solution to address limited supplies and environmental changes. Lignocellulosic biomass holds promise as a renewable resource for the production of cellulosic ethanol, a viable alternative to conventional fuels. Microorganisms such as fungi and bacteria can be employed for this purpose. However, many suitable bacterial hosts lack an efficient extracellular depolymerizing apparatus, limiting their ability to decompose complex materials independently [1]. To overcome this challenge, certain bacteria employ cellulosomes, which are cohesin-dockerin complexes composed of scaffolding proteins that enable the attachment of multiple enzymes to the cell surface. To enhance efficiency and enable industrial-scale application, elevated temperatures and thermostable enzymes are advantageous for lignocellulosic biomass hydrolysis. To achieve tighter binding within complexes, proteins derived from thermophilic organisms, which exhibit enhanced stability under harsh conditions, can be utilized.

Since there are very few stable cognate cohesin-dockerin complexes to provide sufficient binding strengths, we have investigated interactions in cohesin-dockerin pair from thermophilic organism Hungateiclostridium clariflavum (Hc) that exhibit intriguing cross-interactions with the pair from Acetivibrio cellulolyticus (Ac), a natural producer of cellulosomes. Due to the lack of crystal structures for Hc proteins, Alphafold2 model was used as input. With the static structure modeled, using FoldX software, both binding mode systems were analyzed to determine residues crucial for the stability of complex. Later, computational saturation mutagenesis of interface residues was performed to propose new designs with potentially enhanced stability. Given the limited understanding of bacterial structures and their inability to withstand harsh conditions present in during biomass processing, insights into the properties and behaviors of these new structures are crucial for advancing research related to cellulose recycling.

[1] Dvořák et al., 2020: Surface Display of Designer Protein Scaffolds on Genome-Reduced Strains of Pseudomonas putida

Email: wojwar2@st.amu.edu.pl

Visualisation of a friction ridge using molecular spectroscopy

Martyna Czarnomska[1], Emilia Gruszczyńska[2], Aneta Lewkowicz[1], Zygmunt Gryczyński[3], Piotr Bojarski[1], and Sławomir Steinborn[2]

[1]Department of Mathematics, Physics and Informatics, University of Gdańsk
[2]Faculty of Law and Administration, University of Gdańsk
[3]Department of Physics and Astronomy, Texas Christian Universit

Molecular spectroscopy was used to reveal dactyloscopic traces. The dactyloscopic trace left at the scene of an incident is commonly used to identify individuals, as fingerprints are unique, immutable, and indelible. In the present study, we propose a pioneering approach to the analysis of this evidence, taking into account the possibility of visualizing fingerprints, DNA traces, and illegal substances present.

Research methods used to visualize minutiae include stereo fluorescence microscopy, direct PCR for trace DNA analysis, horizontal gel electrophoresis using Diamond™ Nucleic Acid Dye, inverted fluorescence microscopy, and Raman spectroscopy for the identification of illegal substances.

This research was supported by a grant National Science Centre 2021/41/B/HS5/03250 (K.W., M.K., A.L., M.Ch., M.Cz., E.G.).

Email: e.gruszczynska.456@studms.ug.edu.pl

Genome wide prediction of regulatory elements using neural networks

Marlena Osipowicz[1], Julia Smolik[1], Magdalena A. Machnicka[1], and Bartek Wilczyński[1]

[1]Institute of Informatics, Faculty of Mathematics, Informatics, and Mechanics, University of Warsaw, Poland

The identification of regulatory DNA elements such as promoters and enhancers, as well as the location of transcription factor binding sites, is crucial for understanding the mechanisms underlying the regulation of gene expression in eukaryotes. Mutations in DNA regulatory elements can disrupt their binding affinities with transcription factors or alter their interactions with other regulatory elements, leading to abnormal patterns of gene expression. Understanding the impact of such mutations and predicting the activity of regulatory regions could be crucial to unraveling the genetic basis of various disorders in the human genome.

In recent years, the advent of high-throughput sequencing technologies has generated vast amounts of genomic data, enabling the development of computational approaches to predict regulatory elements. We have trained four neural networks to predict regulatory activity, based on a recently published atlas of regulatory elements active in human brain tumors (Stepniak et al., Nature Comm. 2021) and unpublished data on the location of single nucleotide polymorphisms in the same group of patients. The networks provide better classification accuracy than random forest models and achieve AUCs of 80-90%. We conducted a detailed analysis of the filters included in the trained neural networks and found their similarity to DNA sequence motifs recognized by transcription factors. We also proposed methods for detecting correlations between motifs encoded in filters and DNA structural features described by a set of parameters (Zhou et al., NAR 2013). The results obtained may contribute to the understanding of the mechanisms of gene expression regulation and expand the possibilities of applying deep machine learning to the study of epigenetic processes.

Email: julia.smolik.99@gmail.com

Bioinformatics approach for programmed cell death receptor from Nostoc punctiforme

Małgorzata Orłowska[1], Rafał Madaj[1], Jędrzej Kubica[1,2], Kamil Krakowski[1], and Stanisław Dunin-Horkawicz[1,3]

[1]Laboratory of Structural Bioinformatics, Institute of Evolutionary Biology, University of Warsaw, Poland
[2]Laboratory of Theory of Biopolimers, Faculty of Chemistry, University of Warsaw, Poland
[3]Department of Protein Evolution, Max Planck Institute for Biology, Tübingen, Germany

Programmed cell death (PCD) is a mechanism linked to multicellularity. Existing research suggests that extensively studied PCD proteins found in eukaryotes share homology with certain prokaryotic proteins that hitherto have not been fully characterized. Putative PCD apparatus consists of two co-expressed proteins - Wrap1 and Npun_R6613, which complex structure nor coevolution has yet been determined.

Therefore, using bioinformatics approach - AlphaFold2 folding, protein-protein molecular docking, protein interface and sequence analysis we managed to determine structure of Wrap1-Npun_R6613 complex and annotated coevolution of selected residues of both proteins. Our study indicate that it exists as tetradecameric structure, with Wrap1 bound to the membrane, while Npun_R6613 counterpart is non-covalently bound through several interface residues.

It extends present knowledge of mechanisms and evolution of programmed cell death in bacteria. The next step is the experimental part, which are going to be conducted on yeast.

Email: m.orlowska19@uw.edu.pl

Towards understanding the molecular mechanism of inhibiting ferroptotic cell death by targeting lipoxygenases

Karolina Mikulska-Ruminska[1]

[1]Faculty of Physics, Astronomy and Informatics, Nicolaus Copernicus University in Torun, PL 87100, Poland

Recent years have brought attention to ferroptosis, an iron- and lipid peroxidation-dependent form of regulated cell death implicated in a broad range of diseases, including Alzheimer's and Parkinson's disease, acute brain injury, sepsis, or asthma. Ferroptosis characteristic feature is the enhanced lipid peroxidation where abstraction of H-atoms from polyunsaturated phospholipids drives the entire peroxidation process causing membrane damage. We demonstrated that a protein complex composed of 15-lipoxygenase and PE-binding protein 1 (Cell 2017) is a master promoter of ferroptotic cell-death signaling regulated by several enzymatic mechanisms.

Our objective is to block the enzymatic mechanisms underlying the ferroptosis process at the molecular level. Using computational molecular approaches such as molecular dynamics simulations, elastic network models, and bioinformatics together with the experimental verification, we explained the previously unknown mechanisms and factors that affect or block ferroptosis, thus providing molecular insights of the catalytic processes involved (JACS, J Clin Invest 2018, JCIM 2019).

Our recent studies revealed a critical role of iNOS/nitric oxide (Nature Chem Biol 2020, IJMS 2021) and phospholipase iPLA2β (Nature Chem Biol 2021) in the regulation of ferroptosis. We also resolved a paradox related to the most common ferroptosis inhibitor, Ferrostatin 1 (Redox Biol 2021), and we proposed new inhibitors of the human complex which can effectively block the ferroptotic cell death signal (PNAS 2023). Our studies showed that computational studies could be effectively applied to explain and block fundamental biological processes.

Acknowledgments: Work supported by Polish National Science Centre no. 2019/35/D/ST4/02203.

Email: karolamik@fizyka.umk.pl

Understanding the role of the spacer polypeptide for quorum quenching function of MacQ acylase.

Igor Marchlewski[1,2], Bartłomiej Surpeta[1,2], and Jan Brezovsky[1,2]

[1]International Institute of Molecular and Cell Biology in Warsaw, Warsaw, Poland
[2]Laboratory of Biomolecular Interactions and Transport, Department of Gene Expression, Institute of Molecular Biology and Biotechnology, Faculty of Biology, Adam Mickiewicz University, Poznan, Poland

Antibiotic resistance is a growing global concern, hence alternative methods to combat bacteria are urgently required. One of them is the use of quorum quenching (QQ) enzymes, which degrade bacterial signaling molecules, preventing the activation of virulence-related gene expression. In this study, we focus on QQ enzyme - MacQ from Acidovorax sp. MR-S7 – a bifunctional acylase with broad ligand specificity towards bacterial signaling molecules - N-acyl homoserine lactones (AHLs) and various antibiotics [1]. MacQ displays unusual features in contrast to other members of Ntn serine hydrolase family that it belongs to. First, spacer polypeptide (SP) that is formed during auto-proteolytic activation was shown to not dissociate from the activated dimer. Furthermore, MacQ presents capsule-shaped tertiary structure, composed of two heterodimers both with SP bound, representing the second unique feature. Therefore, in this study we modeled the complete structure of MacQ capsule with and without SPs, as well as analogous system variants composed of single dimer typical for its family. We performed molecular dynamics (MD) to investigate the possible effects of the complex composition and the presence of highly flexible SP for the enzyme's structure and its catalytic machinery. For the dimer, the favorable preorganization of active site residues was more frequent with the absence of SP. On the other hand, the capsule composed of two dimers was more prone to dissociate without the SPs bound . Finally, we prepared the MacQ capsule-AHLs complex in both variants, with and without SPs. Their MD simulations indicated either negligible or unfavorable effect of the SPs presence for the stability of the AHLs in the state properly pre-organized for the catalysis, depending on the water model used. Nevertheless, even with the diminishing effect of the SP observed for simulations in TIP3P water model, the overall stabilization of the ligand was better as compared to E. coli penicillin G acylase or Pseudomonas AHL acylase enzymes, which were recently established from the perspective of its QQ activity and related molecular determinants [2]. We hypothesize, that modulation of the MacQ to shift its form to SP free dimer, may represent a promising direction to further improve MacQ acylase towards potential application as an alternative or support for conventional antibiotics.

1. Yasutake, Y. et al. Sci Rep 7, 8946 (2017).
2. Surpeta, B. et al. ACS Catal. 12, 6359–6374 (2022).

Email: igomar.put@gmail.com

Ab Initio Study of Glycine Formation In Condensed Phase

Francisco Carrascoza[1,2], Piotr Lukasiak[1,2], Wieslaw Nowak[3], and Jacek Blazewicz[1,4]

[1]Institute of Computer Science, Poznan University of Technology, Poznan, Poland
[2]European Centre for Bioinformatics and Genetics ECBiG, Poznan, Poland
[3]Institute of Physics, Faculty of Physics, Astronomy and Informatics, N. Copernicus University, Torun, Poland
[4]Institute of Bioorganic Chemistry, Poznan Academy of Sciences, Poznan, Poland

Glycine, a key component in protein and prebiotic substance formation, has intrigued researchers due to its spontaneous formation mechanisms under prebiotic Earth conditions and within the interstellar medium (ISM). However, the understanding of these mechanisms remains a topic of debate, influenced by Earth's evolving geochemical environment. Detecting glycine in the ISM poses challenges, although its potential formation in water-rich ice grains within dense molecular clouds is plausible.

In this study, we present ab initio molecular dynamics (MD) simulations, coupled with modern free energy calculations, to model the chemical reaction involving carbon monoxide, formaldimine, and water, leading to glycine production. Our investigations focus on exploring the conditions favoring glycine formation in the condensed phase at temperatures ranging from 50K to 300K. Additionally, we examine the influence of different electric fields on the reaction process.

Remarkably, our findings reveal that glycine can be formed with significantly low energy barriers of 0.5 kcal/mol at 50K. This study aims to assess whether this reaction represents a viable mechanism for glycine formation in both the ISM and on planet Earth. By doing so, we contribute to the ongoing quest for a consensus among various proposals concerning glycine's origin. Furthermore, our work highlights the importance of employing metadynamics and Car-Parrinello MD methods as powerful tools for unraveling intricate, multistep reaction pathways that may hold significant implications for astronomical phenomena.

This work was supported by DS grant 09/91/DSPB/0649), IV Priority Axis of the Operational Program Smart Growth 2014–2020. Computations were performed at Poznan Super Computing and Networking Centre with grant 462. Partial support from the #MEMOBIT NCU IDUB grant (WN) is also acknowledged.

Email: franciscocarrascoza@gmail.com

On the impact of high product concentration on biotechnological enzyme nitrile hydratase structure and dynamics

Julia Berdychowska[1], Łukasz Pepłowski[1], and Wiesław Nowak[1]

[1]Nicolaus Copernicus University

Nitrile hydratase (NHase), an enzyme used in industry to convert nitriles into amides was studied by using molecular dynamics (MD) simulations on hundreds of ns timescales in various solvents. It is known that the enzyme becomes deactivated in high concentration of the product - acrylamide. The mechanism of such deactivation is unclear. Models of the full enzyme in varying high amide concentration were prepared to monitor possible solvent induced structural rearrangements and to decipher the molecular mechanism of this phenomenon. Sets of 500 ns MD trajectories (CHARMM, NAMD) were compared. Statistics is increased since NHase is a tetramer that contains two identical dimers. Numerous events of substrate and product entering into the active site were observed. The pathways of nitrile entering and amide leaving the active site were traced and amino acids interacting with them were identified. We have noticed some substantial conformational changes that are correlated with the high amide concentration, probably they are responsible for decreasing catalytic activity of NHase. This observation may help to design more stable variants.

Email: jaber@doktorant.umk.pl

Associations between Host Factors and the Gut Mycobiota

Natalia Szóstak[1], Luiza Handschuh[1], Anna Samelak-Czajka[1], Katarzyna Tomela[1,2], Marcin Schmidt[1,3], Łukasz Pruss[4,5], Kaja Milanowska-Zabel[4], Piotr Kozlowski[1], and Anna Philips[1]

[1]Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznan, Poland
[2]Department of Cancer Immunology, Chair of Medical Biotechnology, Poznan University of Medical Sciences, Poznan, Poland
[3]Department of Food Biotechnology and Microbiology, Poznan University of Life Sciences, Poznan, Poland
[4]Ardigen S.A., Cracow, Poland
[5]Department of Biochemistry, Molecular Biology and Biotechnology, Faculty of Chemistry, Wroclaw University of Science and Technology, Wroclaw, Poland

Recent research has uncovered a significant role the gut fungal community plays in maintaining human well-being. We delved into the composition and diversity of gut fungi among individuals from the European population. We examined the relationship between gut mycobiota and various host-related sociodemographic, lifestyle, health, and dietary factors. The study encompassed 923 participants. Fecal DNA samples were analyzed by whole-metagenome high-throughput sequencing. Subsequently, we identified the taxonomic profiles of the fungal species and conducted computational and statistical analyses to uncover any associations with 53 factors related to the individuals.

The fungal communities in our samples exhibited a notable prevalence of Saccharomyces, Candida, and Sporisorium. We discovered that ten factors demonstrated significant correlations with the overall variation in the mycobiota, many of which were related to diet. These included the consumption of chips, meat, sodas, sweeteners, processed food, and alcohol, as well as factors like age and marital status. We also observed differences in the diversity and composition of the fungal communities in relation to other factors such as body mass index (BMI), occupation, autoimmune diseases, and the use of probiotics. Through differential abundance analysis, we identified specific fungal species that exhibited distinct patterns of change under particular conditions.

Our findings indicate that yeast species, including Saccharomyces, Malassezia, and Candida, dominate the human gut mycobiota. Although there was considerable variability between individuals, certain fungal species were consistently present across the majority of samples, suggesting the existence of a core gut mycobiota. Furthermore, we demonstrated that host-related factors such as diet, age, and marital status contribute to the variability observed in the gut mycobiota. To the best of our knowledge, this study represents the first comprehensive and large-scale investigation of the European population, examining the associations between gut mycobiota and such an extensive and diverse host-related set of factors.