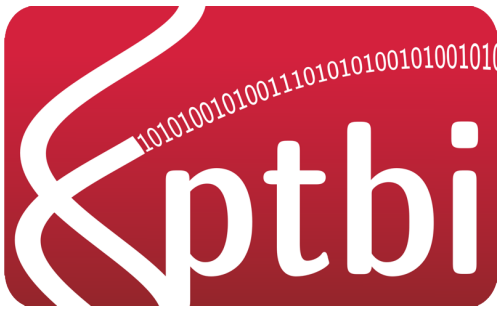
Bioinformatics in Toruń
Book of Abstracts

23–25 June 2022, Toruń, Poland



Program Committee:

- Prof. Dr. Wieslaw Nowak (*Nicolaus Copernicus University, Torun, Poland*)
- Prof. Dr. Jacek Błażewicz (*Poznan University of Technology, Poznan, Poland*)
- Prof. Dr. Jerzy Tiuryn (*Warsaw University, Warsaw, Poland*)
- Prof. Dr. Jarek Meller (*University of Cincinnati, USA*)
- Prof. Dr. Witold Rudnicki (*Bialystok University, Bialystok, Poland*)
- Dr. Łukasz Pepłowski (*Nicolaus Copernicus University, Torun, Poland*)

Local Organizing Committee:

- Prof. Dr. Wieslaw Nowak (*Nicolaus Copernicus University, Torun, Poland*)
- Dr. Łukasz Pepłowski (*Nicolaus Copernicus University, Torun, Poland*)
- Dr. Bartek Wilczyński (*Warsaw University, Warsaw, Poland*)
- Prof. Dr. Witold Rudnicki (*Bialystok University, Bialystok, Poland*)
- Dr. Jakub Rydzewski (*Nicolaus Copernicus University, Torun, Poland*)
- Dr. Katarzyna Walczewska-Szewc (*Nicolaus Copernicus University, Torun, Poland*)

Should you have any questions regarding Book of Abstracts, contact J. Rydzewski: `jr@umk.pl`.

Contents

Lectures

Mental states in brains and computers

Włodek Duch[1]

[1]Neurocognitive Laboratory, Center for Modern Interdisciplinary Technologies, Nicolaus Copernicus, Poland

Full understanding of brain/mental processes requires multi-level phenomics. We would like to create maps between brain states to mental states [1]. Such modeling attempts have been made in the past at high levels of abstraction [2]. Attractor network models are a step closer to neurobiological reality. Using neural models that incorporate some neurobiological mechanisms gives an insight into putative brain processes. Trajectories of brain activations may be linked to mental events. Transitions between brain states depend on the biophysical properties of neurons. Graphs that show such transitions illustrate associative thinking processes, creation of conceptual networks, and help to understand mental disorders. Dynamics of such networks allows to formulate novel hypotheses, taking into account processes contributing to activation of neurons on many levels.

Progress in understanding mechanisms of spatial navigation in hippocampus formation (Nobel 2014) shows the way to create abstract cognitive maps. Analysis of real brain signals to discover fingerprints of brain cognitive activity is still a challenge, but has a great potential to create better brain-computer interfaces, neurofeedback and therapeutic procedures. Using functional magnetic resonance imaging (fMRI) and electroencephalographic (EEG) data, it is possible to assess changes in the activation of large-scale brain networks and to develop biomarkers useful for the diagnosis of mental diseases. These networks change as a result of cognitive load and working memory training, modifying the network of potentially accessible brain states [5]. The challenge is to recognize brain fingerprints using EEG and develop practical methods to extract relevant information from the EEG signals in real time, and use it to parameterize neuromodulation methods (DCS, TMS) that act directly on the structure of brain connections, repairing or optimizing their performance.

[1] Duch W (1996) Computational physics of the mind. Computer Physics Communication 97: 136-153

[2] Duch W and Diercksen GHF (1995) Feature Space Mapping as a universal adaptive system. Computer Physics Communications 87: 341-371

[3] Duch W. (2019), Autism Spectrum Disorder and deep attractors in neurodynamics. Ch. 13, Springer Handbook of Multi-Scale Models of Brain Disorders, pp.135-148.

[4] Duch W. (2021). Memetics and Neural Models of Conspiracy Theories. Patterns. Cell Press

[5] Finc, K, Bonna, K, He, X, Lydon-Staley, D.M, Kühn, S, Duch, W, & Bassett, D. S. (2020). Dynamic reconfiguration of functional brain networks during working memory training. Nature Communications 11, 2435 (IF 11.8)

Email: `wduch@umk.pl`

The challenges and opportunities of adaptive nanopore sequencing

Alex Adams[1]

[1]Nuffield Department of Clinical Medicine, University of Oxford

Nanopore sequencing utilises changes in current flow through nanometer-scale pores in a polymer membrane caused by the passage of a DNA/RNA molecule through the pore, with the nucleotide sequence determined from the current signal. Nanopore sequencing generates sequence data in real-time, with read lengths >1Mb achievable, and has been used in generation of novel genome assemblies, targeted sequencing with and without PCR, metagenomic sequencing, simultaneous identification of nucleotide modifications, and recently, covid testing. A variety of devices are available to suit different scales and budgets, with sizes ranging from benchtop to pocket-sized.

Adaptive sequencing is a developing protocol which allows enrichment or depletion of pre-selected sequences by rapidly aligning the beginning of a strand against a reference in real-time using the GPU, and either continuing sequencing or ejecting the strand from the nanopore. This talk will give an overview of the technology, the strengths and limitations of nanopore adaptive sequencing, and experience of setting up the protocol.

Email: `alex.adams@ndm.ox.ac.uk`

Integrated drug discovery simulations for Covid-19 main protease

Yasuteru Shigeta[1]

[1]Center for Computational Sciences, University of Tsukuba, Tsukuba, Japan

The close-contact indoor social activity combined with the lightning-fast spread of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) fuels the coronavirus disease 2019 (COVID-19) pandemic. This global pandemic directly affected the world economics. Currently, rare diseases or new epidemics lack therapeutic agents regarded as a significant health problem for the world population. To deal with these problems, one must search for a potent compound or drug. However, the traditional procedure takes more than ten years and a million dollars to launch an effective medication for drug discovery that is extremely important in the pharmaceutical industry.

For this issue, we have conducted in silico studies for the drug repositioning together with the experimental assay to confirm the efficacy [1-3]. The screening and molecular docking methods were used to search for the possible inhibitor from the numerous compounds in the online and in-house databases (halogenated derivative library) that have been confirmed the efficacy by experimental and computational studies. Recently, the potent inhibition efficacy of a natural compound, flavonoid, on SARS-CoV2 protease at the step of viral replication is confirmed by in vitro study. Also, the crystal structure of SARS-CoV2 protease in complex with baicalein was available in the database, and we used it as a structural template for screening in this study. Among several flavonoids and their derivatives, a brominated baicalein exhibit the highest affinity.

To get more insight of binding affinity of compounds to an allosteric pocket, a quantum mechanics (QM)-based method, fragment molecular orbital (FMO) combined with a polarizable continuum model (PCM) were used to specify the amino acid residues interacting with the brominated baicalein. Each pair of the fragment was figured the paired interaction energy (PIE and PIEDA) that was used to explain both bonded and non-bonded interaction. According to PIEDA, the favorable region of TH024 is the oxyanion region of the 3CLpro located at residue 142 to 147 and 163. In this pocket, G143 and C145 showed the lowest -25.3 and -26.9 kcal/mol. This indicates that this compound may serve as a potent anti-viral drug toward 3CLpro [3].

Furthermore, the application of Ligand Binding Path Sampling Based on Parallel Cascade Selection Molecular Dynamics (LB-PaCS-MD) method [4] combined with the FMO calculation disclose the ligand binding pathway from aqueous solution into the active site of 3CLpro and customize the binding pocket suitable for a potent compound. Here, rubraxanthone has been confirmed to have an antiviral activity to SAR-CoV-2 Mpro as a mixed inhibition mode with relatively low cytotoxicity and high percent cellular inhibition.

[1] B. Nutho, et al., Biochemistry, 59(18), 1769-1779 (2020).

[2] P. Deetanya, et al., Comput. Struct. Biotech. J., 19, 3364-3371 (2021).

[3] K. Hengphasatporn, et al., J. Chem. Info. Model. 62(6), 1498–1509 (2022).

[4] H. Aida, Y. Shigeta, R. Harada, Materials 15, 1490 (16 pages) (2022).

Email: shigeta@ccs.tsukuba.ac.jp

Comparative genomics of insects to unveil the evolution of gene regulatory networks

Guillem Ylla[1]

[1]Laboratory of Bioinformatics and Genome Biology, Faculty of Biochemistry, Biophysics and Biotechnology, Jagiellonian University, Krakow, Poland

Insects represent 90% of the described animal species and display a huge range of morphologies and lifestyles which allowed them to conquer almost any terrestrial ecosystem. There is also a long history of using insects to study biology, and the genetic model par excellence, the Drosophila melanogaster, is a good example of it. With the emergence of the "omics" technologies, we are now in a privileged position to study genomes and how they are regulated to produce different phenotypes. Arguably, one of the most remarkable examples of the power of genome regulation is the insect metamorphosis, in which a single genome produces strikingly different larva, pupa, and adult characters. Even more intriguing, is how metamorphosis originated and evolved through evolution.

Traditionally, most of the genetic information of insects came from just a few closely related species with complete metamorphosis. In recent years, however, we have sequenced new insect genomes from underrepresented lineages that maintained the ancestral and incomplete type of metamorphoses, such as that of cockroaches and crickets. These genomes, coupled with the "omics" data we generated and bioinformatics techniques, allowed us to unveil novel roles of regulatory elements, like DNA methylation and transcription factors, in insect development and metamorphosis. Furthermore, using hundreds of "omics" datasets from public databases we are able to test whether our findings on specific organisms can be generalized across species.

Email: `guillem.ylla@uj.edu.pl`

Machine learning approaches to regulatory element identification

Bartek Wilczynski[1]

[1]University of Warsaw

In the past 10 years, we have worked on various approaches to gene regulatory element classification. We have been using Bayesian networks (Bonn et al. 2012), Random Forests (Herman-Iżycka et al, 2017), Support vector machines (Podsiadło et al 2013) to predict enhancer and promoter positions in human and model organism genomes. Now we have some new results regarding predicting positions of enhancers and promoters in human gliomas and glioblastoma. This is based on the data from the glioma regulatory element atlas that we have published recently (Stepniak et al. 2021) and we can show relatively high accuracy of predictions (AUC>80%) using convolutional neural networks on this challenging dataset.

References:

Bonn, S., Zinzen, R., Girardot, C. et al. Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development. Nat Genet 44, 148–156 (2012).

Herman-Izycka, J., Wlasnowolski, M. & Wilczynski, B. Taking promoters out of enhancers in sequence based predictions of tissue-specific mammalian enhancers. BMC Med Genomics 10, 34 (2017).

Podsiadło, A., Wrzesień, M., Paja, W. et al. Active enhancer positions can be accurately predicted from chromatin marks and collective sequence motif data. BMC Syst Biol 7, S16 (2013).

Stępniak, K., Machnicka, M.A., Mieczkowski, J. et al. Mapping chromatin accessibility and active regulatory elements reveals pathological mechanisms in human gliomas. Nat Commun 12, 3621 (2021)

Email: `bartek@mimuw.edu.pl`

Hunting for Biomarkers in Omics Data - methods and application

Witold Rudnicki[1]

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

In my talk I will present the recent results obtained in analysis of biomedical data using machine learning methods. I will present the methodological advances developed in our laboratory, in particular those aiming at selection of the robust minimal set of biomarkers.

Email: w.rudnicki@uwb.edu.pl

The role of PE-binding protein 1 in the ferroptosis process

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 and Parkinson disease, acute brain injury, kidney damage, and asthma. Active research shows that ferroptosis may become a new strategy in the treatment of cancers. Its 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 PEBP1 (Cell 2017), is a master promoter of ferroptotic cell-death signaling regulated by several enzymatic mechanisms occurring independently or concertedly. Our objective is to unearth the enzymatic mechanisms underlying the ferroptosis process at the molecular level. Using molecular dynamics simulations, elastic network models, and bioinformatic tools together with the experimental verification, we explained the previously unknown mechanisms and factors that affect ferroptosis, thus providing molecular insights of the catalytic processes involved (JACS, J Clin Invest 2018, JCIM 2019). Our recent studies also revealed a critical role of iNOS/nitric oxide (Nature Chem Biol 2020, IJMS 2021) and phospholipase iPLA2beta (Nature Chem Biol 2021) in the regulation of ferroptosis. We also resolved an apparent paradox related to the most common ferroptosis inhibitor, Ferrostatin 1. We demonstrated that its anti-ferroptotic action is not limited to radical scavenging but also includes suppression of peroxidation of arachidonoyl-phosphatidylethanolamine catalyzed by the 15-lipoxygenase-PEBP1 complex (Redox Biol 2021). Our studies showed that the presence of PEBP1 is essential for the generation of the ferroptotic cell death signal.

Acknowledgments: This work was supported by NIH (HL114453, U01AI156924, U01AI156923, CA165065, NS076511, NS061817, P41 GM103712) and by Polish National Science Centre no. 2019/35/D/ST4/02203.

Email: karolamik@fizyka.umk.pl

Computational resources for studying selective macroautophagy

Vasilis Promponas[1]

[1]Department of Biological Sciences, New Campus, Office B161, University of Cyprus

Macroautophagy (or simply autophagy - literally meaning "self-eating") is a catabolic process conserved throughout the eukaryotes. Autophagy was initially considered a bulk recycling process, but recently there has been much interest in studying selective forms of autophagy that mediate the targeted degradation of excess or damaged (i) organelles, (ii) biomolecules, their complexes or aggregates, and (iii) even invading pathogens.

Within this presentation, I will give a brief overview of the field and present recent advances in the design and implementation of tailored computer algorithms, software tools and online resources that enhance our ability to study this key cellular process, in a possible combination of "wet" and "dry" experiments.

Email: ?

Exploring Dynamic Determinants of Quorum Quenching Activity Employed by N-terminal Serine Hydrolases to Enable Rational Engineering of their Potency as Alternative Antimicrobial Agents for Selective Control of Virulence in Resistant Bacterial Species

Bartlomiej Surpeta[1,2], Michal Grulich[3], Andrea Palyzova[3], Helena Maresova[3], 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, Poznan, Poland
[2]International Institute of Molecular and Cell Biology in Warsaw, Poland
[3]Institute of Microbiology, Academy of Sciences of the Czech Republic, Prague, Czech Republic

Antibiotics have served to control bacterial growth for many decades. By interfering with the essential functions, antibiotics exert intense selective pressure, inducing the development of resistance. Hence, the discovery of novel antimicrobials is crucial. Aside from new generations of antibiotics, alternative strategies targeting non-essential bacterial functions are sought to complement the application of antibiotics. Quorum quenching (QQ) processes disrupt bacterial communication via quorum sensing, which markedly affects their virulence, while its indirect mode of action exerts only limited pressure, avoiding induction of the resistance.

Here, we explored mechanisms used by N-terminal serine hydrolases (NtSHs) to degrade N-acyl homoserine lactones (HSLs), signaling compounds of Gram-negative bacteria. We investigated the prototypical QQ enzyme, acyl-homoserine lactone acylase from Pseudomonas aeruginosa (paPvdQ), and penicillin G acylase from Escherichia coli (ecPGA), which was employed as an inactive negative control. Using molecular dynamics (MD) simulations we analyzed the plasticity and pre-organization of their active sites and binding of HSLs. The quantum mechanics/molecular mechanics MD simulations were applied to study the atomistic details of QQ reactions.

We found that both enzymes degraded HSLs via similar mechanisms and confirmed the unexpected activity of ecPGA experimentally, adding this industrially optimized enzyme to the QQ toolbox. We also identified differences in their catalytic actions arising mainly from the distinct structures of their acyl-binding cavities and the dynamics of their molecular gates. To probe the designability of ecPGA towards improved QQ activity, we have rationally selected mutants targeting the found determinants. The experimental characterization of the most promising constructs with several HSLs confirmed their successful engineering.

Overall, we identified mechanisms of QQ activity employed by several NtSHs, and determinants governing degradation efficiency. We altered these structural and dynamic determinants in ecPGA, establishing their relevance in designing more potent antibacterial agents.

This work was supported by the National Science Centre, Poland (2017/25/B/NZ1/01307 and 2021/41/N/NZ2/01365). The computations were performed at the Poznan Supercomputing and Networking Center.

Email: janbre@amu.edu.pl

Charting and characterizing the conformational landscape of native proteins

Shoshana J. Wodak[1]

[1]VIB-VUB Center for Structural Biology, Pleinlaan 2 1050 Brussels Belgium. ORCID: http://orcid.org/0000-0002-0701-6545

Proteins are dynamic systems that adopt multiple conformational states, a property essential for many biological processes. Over two thirds of experimentally determined structures deposited in the Protein Databank (PDB) represent multiple conformations of the same or related protein, observed in different crystal forms, when interacting with other proteins or other macromolecules, or upon binding small molecule ligands. We argue that charting this conformational diversity across the PDB can be employed to build a useful approximation of the conformational landscape of native proteins. We present a community-wide effort towards this goal and discuss the challenges involved.

Email: ?

Lock-Free de Bruijn graph

Daniel Gorniak[1] **and Robert Nowak**[1]

[1]Warsaw University of Technology, Institute of Computer Science

De Bruijn graph is one of the most important data structures used in de-novo genome assembly algorithms, especially for NGS data. There is a growing need for parallel data structures and algorithms due to the increasing number of cores in modern computers.

We created the lock-free version of the de Bruijn graph, as well as a lock-free algorithm to build such graph from reads. Our algorithm and data structures are developed to use parallel threads of execution and do not use mutexes or other locking mechanisms, instead, we used only compare-and-swap instruction and other atomic operations, available in C++20 standard library. It makes our algorithm very fast and efficiently scaling.

We developed a C++ library and tested its performance to depict its high speed and scalability compared to other available tools. We used E. Coli genome, Baker's yeas genome and human, GRCh38.p14 first chromosome. Our results shows 3 times faster than CuttleFish algorithms, currently the fastests de Bruijn assembler. Moreover the algorithm scale very well up to 32 threads, maximum number of cores available on our servers.

C++ library with source codes is available at github `https://github.com/dandon223/LFdBG` under MIT license.

Email: `r.m.nowak@elka.pw.edu.pl`

Topological analysis as a tool for detection of abnormalities in protein-protein interaction data

Alicja Nowakowska[1] **and Malgorzata Kotulska**[1]

[1]Department of Biomedical Engineering, Faculty of Fundamental Problems of Technology, Wrocław University of Science and Technology, Wrocław, 50-370, Poland

Protein-protein interaction datasets constitute an essential layer in multi-omics approach to biomedical knowledge, which can be modeled as networks. This representation gives insight into molecular pathways, help to uncover novel potential drug targets or predict a therapy outcome. Nevertheless, the data that constitutes such systems is frequently incomplete, error-prone and biased by scientific trends in a hidden way. Implementation of methods for detection of such shortcomings can improve protein-protein interaction data analysis.

We performed topological analysis of a few protein-protein interaction networks (PPINs) from IntAct Molecular Database, regarding cancer, Parkinson's disease (two most common subjects in PPINs analysis), Human Reference Interactome, and the AmyloGraph interactome database based on meta-analysis of 883 interactions between 46 proteins reported in 173 manuscripts. The data collections were shown to be often biased by scientific interests, which highly impacts the networks structure. This may obscure correct systematic biological interpretation of the protein-protein interactions and limit their application potential. What is more, we show that updated data in amended databases may even further deteriorate the database quality. The bigger and the newer does not mean the better. As a solution to this problem, we propose a set of topological methods for the bias detection, which performed in the first step provides more objective biological conclusions regarding protein-protein interactions and their multi-omics consequences.

Availability: A user-friendly tool ETNA (Extensive Tool for Network Analysis) is available on https://github.com/AlicjaNo
The software includes a graphical Colab notebook: https://githubtocolab.com/AlicjaNowakowska/ETNA/blob/main/ETN

Email: ?

News from the metagenomics and metadesign of the subways and urban biomes (MetaSUB)

Paweł Łabaj[1]

[1]Jagiellonian University, Krakow, Poland

We present a global atlas of 4,728 metagenomic samples from mass-transit systems in 60 cities over 3 years, representing the first systematic, worldwide catalog of the urban microbial ecosystem. This atlas provides an annotated, geospatial profile of microbial strains, functional characteristics, antimicrobial resistance (AMR) markers, and genetic elements, including 10,928 viruses, 1,302 bacteria, 2 archaea, and 838,532 CRISPR arrays not found in reference databases. We identified 4,246 known species of urban microorganisms and a consistent set of 31 species found in 97% of samples that were distinct from human commensal organisms. Profiles of AMR genes varied widely in type and density across cities. Cities showed distinct microbial taxonomic signatures that were driven by climate and geographic differences. These results constitute a high-resolution global metagenomic atlas that enables discovery of organisms and genes, highlights potential public health and forensic applications, and provides a culture-independent view of AMR burden in cities.

We believe that microbial genetic mapping of urban environments will give public health officials tools to assess risk, map outbreaks, and genetically characterize problematic species. The paper was highlighted as Best of Cell Press 2021 and we wish to further disseminate our results and emphasize the urgent need for developing global surveillance systems to monitor potential infectious diseases outbreaks and AMR distribution in the context of rapid climate change.

Email: `pawel.labaj@uj.edu.pl`

Use of alignment-free sequence descriptors (AFSDs) in the characterization of emerging global viral pathogens

Subhash C. Basak[1], Ashesh Nandy[2], Marjan Vracko[3], Smarajit Manna[4], Tathagata Dey[2], and Tathagata Dutta[2]

[1]Department of Chemistry and Biochemistry, University of Minnesota Duluth, MN 55811, USA
[2]Centre for Interdisciplinary Research and Education, Jodhpur Park, Kolkata 700068, India
[3]National Institute of Chemistry, Hajdrihova 19 , 1000 Ljubljana, Slovenia
[4]Jagadis Bose National Science Talent Search, Kolkata 700107, India

The twenty first century has seen the emergence of a variety of global pathogens including SARS-CoV-2, SARS (Severe Acute Respiratory Syndrome), MERS (Middle East Respiratory Syndrome) , Zika, the latest being the currently transmitting Monkeypox. Coronaviruses consist of a group of enveloped, positive-sense single-stranded RNA viruses that infect humans as well as other animals. We performed cluster analysis on a set of 573 sequences belonging to SARS, MERS, and SARS-CoV-2. The sequences were represented using a set of computed alignment-free sequence descriptors (AFSDs) and analyzed applying different chemometric methods, e.g., Euclidean/ Mahalanobis distances, principal component analysis and self-organizing maps (Kohonen networks). An AFSD maps the biological sequence into the set R of real numbers. A collection of AFSDs comprise a multidimensional space of computed biodescriptors. Results showed that the sequences are well clustered regarding the type of virus; however, some of those show the tendency to belong to more than one virus type. Further studies with a larger sets of coronavirus sequences are in progress. Such studies can be useful in the comparison and characterization of emerging variants with the already known ones. Results of this study will be discussed in this presentation along with future plans of research in this area.

Email: sbasak@d.umn.edu

Structure of the human nuclear pore complex by AlphaFold and integrative modeling

Jan Kosinski[1]

[1]European Molecular Biology Laboratory (EMBL) Hamburg and Centre for Structural Systems Biology (CSSB) Hamburg

Nuclear pore complexes (NPCs) mediate nucleocytoplasmic transport. Their intricate 120-megadalton architecture remains incompletely understood. Here, we report a 70-megadalton model of the human NPC scaffold with explicit membrane and in multiple conformational states. We combined artificial intelligence (AI)–based structure prediction with in situ and in cellulo cryo–electron tomography and integrative modeling. We show that linker nucleoporins spatially organize the scaffold within and across subcomplexes to establish the higher-order structure. Microsecond-long molecular dynamics simulations suggest that the scaffold is not required to stabilize the inner and outer nuclear membrane fusion but rather widens the central pore. Our work exemplifies how AI-based modeling can be integrated with in situ structural biology to understand subcellular architecture across spatial organization levels.

Email: `jan.kosinski@embl-hamburg.de`

Exploring gene expression features concordant with Topologically Associating Domains

Patrycja Rosa[1,2] **and Aleksander Jankowski**[1]

[1]Faculty of Mathematics, Informatics and Mechanics, University of Warsaw, Warsaw, Poland
[2]Laboratory of Molecular Neurobiology, Nencki Institute of Experimental Biology, Polish Academy of Sciences, Warsaw, Poland

Gene expression profiles differ between different tissues, and also within heterogeneous populations of cells forming a single tissue. The variation in gene expression may be caused by different mechanisms, and can help achieve robustness of gene expression programs. We aimed to test whether the variation in gene expression could possibly be regulated by the spatial chromatin structure, especially by Topologically Associating Domains (TADs).

For our study, we used published scRNA-seq data from Drosophila melanogaster ventral nerve cord. We grouped 24,199 cells into 142 cell clusters, corresponding to specific central nervous system subpopulations. To obtain the information about gene location in TAD structure, and to calculate the distance from TAD boundaries, we integrated them with previously published Hi-C data from Kc167 cells. We calculated the average gene expression level in each cell cluster, and further used the mean and standard deviation of these averages to quantify gene expression variability.

We found that mean and standard deviation of expression are mostly correlated with each other, and which justified further use of their ratio (coefficient of variation) as a measure. Overall, genes located close to the TAD boundary are less variable in their expression, and those transcribed in the direction away from the TAD boundary have higher variability. We further considered pairs of genes at different distance ranges, calculated their spatial autocorrelation, and compared them to a model with permuted gene expression profiles. We observed an enrichment indicative of a relationship between gene expression variation and the gene location in the chromatin structure.

Email: `patrycja.rosa@uw.edu.pl`

De novo assembly as a solution to the Traveling Salesman Problem in bounded-degree graphs using Quantum Annealer.

Katarzyna Nałęcz-Charkiewicz[1]

[1]Warsaw University of Technology

An algorithm of de novo assembly of DNA sequences, formulated as a solution to the Traveling Salesman Problem in bounded-degree graphs will be presented. The solution is designed for long-read sequencing technologies. An approach was applied in which the Traveling Salesman Problem is defined as a Pure Integer Programming task, using one of the so-called compact (with polynomial number of variables and constraints) TSP formulations. The quantum annealing algorithm was used to find the optimal (or close to optimal) path in the graph, using D-Wave's quantum annealer and Constrained Quadratic Model Hybrid Solver, one of the newest solvers in D-Wave hybrid solver service.

Email: k.m.nalecz@gmail.com

Squeezing water to the limits, transport of water through narrow tunnels in enzymes and consequences to their (dis)function

Carlos Sequeiros-Borja[1,2], Aravind Selvaram Thirunavukarasu[1,2], Cedrix Dongmo Foumthuim[1,2], and Jan Brezovsky[1,2]

[1]International Institute of Molecular and Cell Biology in Warsaw
[2]Institute of Molecular Biology and Biotechnology, Faculty of Biology, Adam Mickiewicz University in Poznan

An aqueous environment is necessary for life as we know it, and water is required for almost all biochemical processes at a molecular level. Proteins employ water molecules in different ways: as proton donors or receivers, molecular stabilizers, building blocks for larger molecules and as entropic and enthalpic actors. In this sense, proteins need to transport water molecules across their internal network of tunnels to reach the desired action sites, either inside them or acting as molecular pipes to control cellular osmotic pressure.

Even though water is an essential player in enzymatic activity and stability, its transport has been mostly neglected, with water transport studies mainly focused on the transport across membrane proteins. The transport of molecules through a protein's tunnel network has proven hard to study experimentally, with a few exceptions. Hence, the most popular approach to study such events is Molecular Dynamics (MD) simulations. The primary advantage of MD simulations is that they provide a detailed atomic picture of the transport process itself, the tunnel pathways employed and estimated rates of transport.

In this study, we focused on the transport of water molecules across three members of the subfamily of hydrolases: haloalkane dehalogenase from Rhodococcus rhodochrous (Dha), potato epoxide hydrolase 1 (Epx), and lipase from Candida rugosa (Lip). Employing 5μs of MD simulation per system, we observed that only a few tunnels were responsible for the majority of water transport in Dha, contrasting with a higher diversity of tunnels in Epx and Lip. Interestingly, water molecules could traverse narrow tunnels of subatomic radius, which is surprising given the accepted value of 1.4 Å as water molecule radius. Our analysis of the transport events in such narrow tunnels showed a significantly larger number of hydrogen bonds formed between the water molecules and protein, likely compensating for the steric penalty of the process.

Our results revealed that commonly disregarded narrow tunnels account for 20% of total water transport processes observed, an amount that is not negligible at all. We showed how the insights obtained here could be applied to explain differences in a single-point mutation of the human soluble epoxide hydrolase, which is associated with a higher incidence of ischemic stroke.

Email: carseq@amu.edu.pl

Basecalling DNA sequences using sequence-to-sequence machine learning algorithms by joint processing of raw and event nanopore data

Adam Napieralski[1] **and Robert Nowak**[1]

[1]Institute of Computer Science, Faculty of Electronics and Information Technology, Warsaw University of Technology

Third-generation DNA sequencers provided by Oxford Nanopore Technologies (ONT) produce electrical current signal data that can be used for detecting nucleotides. This problem is called basecalling. Various solutions have already been proposed — the earlier ones were based on HMMs. Then, machine learning algorithms based on neural networks have gained much more attention and achieved better accuracy scores when applied. However, results were still lower than competitive sequencing techniques like e.g. Illumina's.

Basecallers also differ in input data type — currently, most of them work on raw data straight from the sequencer, but the approach of using event data is also known and can be explored. Event data is obtained by preprocessing raw data and dividing it into appropriately determined segments, described by several features computed from raw data values within each segment.

We propose a novel approach that uses joint processing of raw and event data and develop a new basecaller named Ravvent, which incorporates this approach. We implement the tool using sequence-to-sequence machine learning models based on an encoder–decoder architecture of recurrent neural networks (RNNs). It uses twin encoders and an attention mechanism.

We tested the tool on real-life and simulated datasets. We analysed the suitability of different RNN types for use in the basecaller and the accuracy of the models on simulated datasets with increasing complexity. We compared the full model's accuracy and performance results with its components, i.e., processing only raw and only event data. The proposed basecaller was also compared with the existing ones, including Guppy, the current official ONT basecaller.

Email: `adam.napieralski.stud@pw.edu.pl`

fingeRNAt—A novel tool for high-throughput analysis of nucleic acid-ligand interactions

Natalia A. Szulc[1], Zuzanna Mackiewicz[1], Janusz M. Bujnicki[1], and Filip Stefaniak[1]

[1]Laboratory of Bioinformatics and Protein Engineering, International Institute of Molecular and Cell Biology in Warsaw, Warsaw, Poland

Computational methods play a pivotal role in drug discovery and are widely applied in virtual screening, structure optimization, and compound activity profiling. Over the last decades, almost all the attention in medicinal chemistry has been directed to protein-ligand binding, and computational tools have been created with this target in mind. With novel discoveries of functional RNAs and their possible applications, RNAs have gained considerable attention as potential drug targets. However, the availability of bioinformatics tools for nucleic acids is limited.

We will present fingeRNAt — a software tool for detecting non-covalent interactions formed in complexes of nucleic acids with ligands. The program detects nine types of interactions and the scope of contacts can be easily expanded using a simple plugin system. We will present applications of fingeRNAt-generated interaction fingerprints for visual and computational analysis of RNA-ligand complexes, including analysis of interactions formed in experimentally determined RNA-small molecule ligand complexes deposited in the Protein Data Bank. We will also discuss the application of interaction fingerprint-based similarity as an alternative measure to RMSD to recapitulate complexes with similar interactions but different folding.

Email: `fstefaniak@iimcb.gov.pl`

Posters

A methodology to find new enzyme families based on protein structure modeling by artificial intelligence. Protein kinases example

Marcin Gradowski[1], Marianna Krysińska[1], and Krzysztof Pawłowski[1,2,3]

[1]Department of Biochemistry and Microbiology, Warsaw University of Life Sciences — SGGW, Warsaw, Poland
[2]Department of Molecular Biology, University of Texas Southwestern Medical Center, Dallas, TX, USA
[3]Department of Translational Medicine, Lund University, Lund, Sweden

Currently, methods for detecting distant homology based on amino acid sequence have reached their limits. Typically, they rely on important catalytic residues of the enzyme family.

By modeling protein structures by artificial intelligence-based programs (such as AlphaFold2 and RoseTTAFold), we can generate highly accurate protein models. Then, by comparing the models to known structures, we can infer homology by structural similarity and assign novel members to known enzyme families in the absence of detectable sequence similarity.

We will present few examples of uncharacterized proteins, e.g., from human pathogens, whose 3D models exhibit structural similarity to the protein kinase-like superfamily despite lack of significant sequence similarity and lack of conserved active sites.

In conclusion, we will discuss a method for finding novel enzyme families using protein structure models.

Email: mgradowski89@gmail.com

PACT – Prediction of Amyloid Cross-interactions by Threading

Jakub W. Wojciechowski[1] **and Małgorzata Kotulska**[1]

[1]Department of Biomedical Engineering, Wroclaw University of Science and Technology, 50-370, Wrocław, Poland.

Amyloids are protein aggregates most commonly known for their role in the development of severe neurodegenerative diseases such as Alzheimer's or Parkinson's disease. However, the unique features of such structures were utilized by many organisms for a wide range of physiological roles including biofilm formation and hormone storage. More recent studies have shown that in some cases the presence of amyloid aggregates can affect the aggregation kinetics of other proteins. This so called cross-seeding or, more generally, cross-interactions turned out to be crucial for understanding comorbidity of amyloid related diseases, including Alzheimer's disease and type II diabetes. Despite the importance of the process, our understanding of it is still very limited due to costly and time consuming experiments required to study such interactions. To overcome this problem, we have developed PACT method. The method is based on modeling of the heterogenous fibril, formed by two sequences of interest. Such a model is then assessed using DOPE statistical potential implemented in Modeller software. The main assumption of the method is that pairs of interacting amyloids will be more energetically favorable than negative cases. Based on that, it is possible to find an energy threshold for cross-interacting pairs. Importantly, the method can work with long protein fragments and, as a purely physicochemical model, it relies very little on the training data. The method, for the first time, opens the possibility of high throughput study of amyloid interactions.

Email: `jakub.wojciechowski@pwr.edu.pl`

Chaos game representation and its application in bioinformatics

Adrian Kania[1]

[1]Department of Computational Biophysics and Bioinformatics, Faculty of Biochemistry, Biophysics and Biotechnology, Jagiellonian University, Gronostajowa 7, Cracow 30-387, Poland

Chaos game representation is a useful technique for transforming biological sequences into series. Then, they may be analysed by time-series techniques, for example, Discrete Fourier Transform. It is especially interesting when it comes to free-alignment methods. The whole process is not only fast but also provides reliable phylogenetic trees. Another application encompasses complexity analysis. It appears that some lethal mutations may be detected using this approach. There are many possible extensions of the chaos game representation devoted to protein or nucleotide sequences. In my last paper, I tested how the combination of chaos game representation and DNA walk may be utilized to detect some differences among essential and non-essential genes in Bacteria. Despite persistent advances in experimental methods, computer predictions may constitute an important part of this investigation. Additionally, I presented there some characteristic patterns that are expected to be developed in future studies.

Email: adkwazar@gmail.com

Stop the biting: tracking the insecticide-mediated allosteric changes in insect and human voltage-gated sodium channels.

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

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

Mosquitoes are the primary vectors of diseases such as malaria, yellow fever, dengue, West Nile fever, or Zika, registered in more than 100 countries. Those diseases represent 17% of the global infectious diseases, causing almost one million deaths each year. Prevention is largely implemented by the usage of insecticides. However, mosquitoes have developed resistance to the most commonly used insecticides. At a time of global warming, this poses a particularly serious threat. The first step in finding new, effective chemicals that could reduce disease transmission is thorough understanding of molecular mechanisms of insecticide mode of action is required.

Among molecular targets of insecticide action are voltage-gated sodium channels (VGSCs). These transmembrane proteins initiate and propagate action potentials in response to a membrane depolarization in nerves and muscles. The structure of the α-subunit of VGSC consists of a single polypeptide chain that folds into four domains built up from six transmembrane helices each. Helices S1-S4 constitute the voltage-sensing-domain with helix S4 acting as a voltage sensor, while helices S5 and S6 contribute to the ion-conducting pore. Here, we present a detailed analysis of selected insecticide interactions with cockroach (Periplaneta americana) PaNav1 model and human hNav1.7 VGSCs. By analyzing molecular dynamics simulations of insecticide-channel complexes, we look for allosteric conformational changes that may lead to the channels' prolonged activation and inactivation. We also investigate the volume of lateral fenestrations present in VGSCs that could enable insecticide access to the central cavity of the channel. In future work, we plan to design a new, photoswitchable acyl sulfonamide insecticide with an azobenzene functional group that should enable light-modulated VGSC conductance in the insect central nervous system.

Our modeling is the first step in the development of a new class of selective, light-controlled insecticides and may facilitate research in insect biology.

Acknowledgements:

NCN PRELUDIUM-20 Grant number 2021/41/N/NZ3/02165

Calculations have been performed on the Centre of Informatics - Tricity Academic Supercomputer & network (CI TASK)

Email: beata.niklas.x@gmail.com

Links between amyloid-related diseases and bacteria proteome

Alicja Nowakowska[1], **Jakub Wojciechowski**[1], **and Małgorzata Kotulska**[1]

[1]Wrocław University of Science and Technology

Multiple bacteria species are suspected to be involved in onset and progression of many disorders including neurodegenerative ones and Type II diabetes. One of the possible explanations for this phenomena is amyloid interactions. Amyloids are insoluble fibrils with a cross-beta structure. Amyloid deposits are well known hallmarks of Parkinson's and Alzheimer's disease. Nevertheless, they can also have a functional role e.g. in bacteria they are building blocks of biofilms. It is possible that the aggregation of functional amyloids from bacteria induces aggregation of the pathological ones. To shed light on this hypothesis, we performed a search for the functional amyloids in the human gut microbiome. In the first step, we extracted homologs of the 42 known functional amyloids. To discard homologs without amyloid properties, we looked for amyloidogenic motifs in the dataset using MEME software. In addition, we performed annotation analysis with EggNOG and InterProscan. Our results indicated the presence of a large collection of functional amyloids spread across different bacteria families. This shows that the space of potential candidates for amyloid interactions is much wider than was previously thought.

Email: `alicjanowakowska01@gmail.com`

Photodynamics of biliverdin IXα bound to bacteriophytochrome via enhanced sampling

Sylwia Czach[1], Jakub Rydzewski[1], Katarzyna Walczewska-Szewc[1], Wiesław Nowak[1], and Krzysztof Kuczera[2]

[1]Institute of Physics, Nicolaus Copernicus University
[2]Department of Molecular Biosciences, University of Kansas, Lawrence

Photoisomerization is a characteristic feature of light-activated molecular switches consisting of a chromophore bound to a protein. It is also one of the primary mechanisms for converting light into local molecular fragment motion. Complexes exhibiting photoisomerization are present in all living organisms.

For biliverdin IXα (BV) bound to bacteriophytochrome (BphP) (from Deinococcus radiodurans), a conformational change in the D-ring of the pyrrole chromophore triggers conversion between spectrally distinct red (Pr) and far-red (Pfr) conformers.

To overcome the conformational transition energy barrier, we use a molecular dynamics (MD) approach based on enhanced sampling.

We conclude that the results obtained from free energy (FE) barrier calculations between important metastable states are consistent with experimental data. We successfully show that the enhanced dynamics of the BV BphP complex contribute to detecting fundamental conformational changes that propagate through two experimentally determined signal transduction pathways.

Our work provides a better understanding of the mechanisms behind BV photoisomerization. And it gives a more accurate understanding of the intramolecular signal transduction pathways from the chromophore to other regions of the phytochrome protein.

Email: `s.czach@yahoo.com`

Signals Transduction Inside Proteins after Ligand Photoexcitation. Insights from Apomyoglobin Molecular Dynamics Simulations

Łukasz Pepłowski[1]

[1]Institute of Physics, Nicolaus Copernicus University, Toruń, Poland

Achieving light control of matter is a promising goal. Chromophores undergoing conformational changes upon photon absorption trigger useful biological signals: vision, phototropism, etc. In recent years a boom in the optogenetics is observed. Prompted by these developments we ask: to what extent a change in a local dipole moment or a change of a ligand shape in a protein interior may affect the protein dynamics? The answer may help designing new molecular switches or light-controlled drugs.

As model systems we used apomyoglobin H93G (ApoMb) with popular fluorescent charge transfer probes (PRODAN, ALADAN) or photoactive azobenzene derivative JB253 (possible antidiabetic drug) inserted in the heme cavity. PRODAN exhibits fluorescence very sensitive to the environment's polarity. We exploited a concept that its dual emission originates from two states: one weakly polar - planar and the other one strongly polar, twisted [1,2]. We consider this fluorophore as a simple "limiting case" model for computational studies of electronic excitation induced conformational changes in proteins.

Our docking studies show that both PRODAN and JB353 molecules can bind in the ApoMb pocket. Using a simplified sudden excitation approach [3] for the docked structures we monitored a protein structure evolution on hundreds of nanoseconds time scale using classical and excited state MD. Results show, that there are noticeable changes in the local protein dynamics but even a very large charge transfer does not affect globular protein model dynamics dramatically. This study delineates a possible range of conformational changes accessible for a small globular protein triggered by a photon absorption in an endogenous chromophore. Complete studies of light-controlled channels may require exascale computing.

Acknowledgement: NCN Grant no. 2016/23/B/ST4/01770

[1] Cohen BE. et al., Science 2002, 296:1700

[2] Balter A, et al., Chem Phys Lett 1988, 143:565

[3] Rydzewski J, et al. Handbook of Computational Chemistry, Springer, 2016

Email: drpepe@fizyka.umk.pl

Diverse Roles of Membrane in Ferroptic Cell Death. Insights from Molecular Dynamics Study of 15LOX1-PEBP1 Complex

Thiliban Manivarma[1,2], Karolina Mikulska-Rumińska[1,2], and Wiesław Nowak[1,2]

[1]Nicolaus Copernicus University in Torun, Department of Biophysics, Torun, Poland
[2]Centre for Modern Interdisciplinary Technologies, 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 extensive all-atoms molecular dynamics data on 15LOX1-PEBP1 system embedded in a realistic membrane environment. We repeat simulations for membrane-free complex and in that way we pinpoint structural differences in this system induced by interactions with membrane composed out of 1,2-Dioleoyl-sn-glycero-3-phosphocholine (DOPC), 1,2-Dioleoyl-sn-glycero-3-phosphoethanolamine (DOPE) and 1-stearoyl-2-arachidonyl-phosphoethanolamine (SAPE). Moreover, In future we will be planning to use Steered Molecular Dynamics and Enhanced Sampling techniques to study the ligand (substrates) diffusion pathway to the entrance of catalytic site of 15LOX1.

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: thiliban@doktorant.umk.pl

Evaluation of the stereochemical quality of predicted RNA 3D models in the RNA-Puzzles submissions

Francisco Carrascoza[1], Maciej Antczak[1,2], Zhichao Miao[3,4], Eric Westhof[5], and Marta Szachniuk[1,2]

[1]Institute of Computing Science and European Centre for Bioinformatics and Genomics, Poznan University of Technology, 60-965 Poznan, Poland
[2]Institute of Bioorganic Chemistry, Polish Academy of Sciences, 61-704 Poznan, Poland
[3]European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Genome Campus, CB10 1SD, UK
[4]Translational Research Institute of Brain and Brain-Like Intelligence and Department of Anesthesiology, Shanghai Fourth People's Hospital Affiliated to Tongji University School of Medicine, Shanghai 200081, China
[5]Université de Strasbourg, Institut de biologie moléculaire et cellulaire CNRS, Architecture et Réactivité de l'ARN, 12 allée Konrad Roentgen, 67084 Strasbourg, France

In silico prediction is a well-established approach to derive a general shape of an RNA molecule based on its sequence or secondary structure. This paper reports on the stereochemical quality of the RNA three-dimensional models predicted using dedicated computer programs. The stereochemistry of 1,052 RNA 3D structures, including 1,030 models predicted by fully automated and human-guided approaches within 22 RNA-Puzzles challenges and reference structures, is analysed. The evaluation is based on standards of stereochemistry, established for RNA, that the Protein Data Bank requires from deposited experimental structures. Deviations from standard bond lengths and angles, planarity or chirality are quantified. A reduction in the number of such deviations should help in the improvement of RNA 3D structure prediction accuracy.

Email: `franciscocarrascoza@gmail.com`

Towards rational design of biotechnological enzymes: Computational analysis of nitrile hydratase in non-classical solvents

J. Berychowska[1], L. Peplowski[1,2], Z. Cheng[3], Z. Zhou[3,4], and W. Nowak[1,2]

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

[2] Centre for Modern Interdisciplinary Technologies, Nicolaus Copernicus University, Wileńska 4, 87-100 Toruń, Poland

[3]Key Laboratory of Industrial Biotechnology, Ministry of Education, School of Biotechnology, Jiangnan University, Wuxi 214122, China

[4]Jiangnan University (Rugao) Food Biotechnology Research Institute, Rugao 226500, China

Green chemistry uses enzymes, and one of the most successful examples is application of nitrile hydratase (NHase) to convert nitriles into amides [1]. The main question is how to optimize it further for industrial purposes [2, 3]. Some NHases lose their activity in the product concentration from 20% to 40% [4]. Current variant, extracted from Pseudonocardia thermophila, loses its activity when a concentration of product reaches 50%. In order to gain insight into the mechanism of this effect and to propose modifications in the enzyme (1IRE.pdb) we performed MD simulations (3x200 ns for each solution, NAMD, CHARMM27, 300 K) with increasing concentration of acrylamide (0%, 20%, 50% m/m respectively). We could observe that acrylamide molecules enter the active center of the enzyme. We were able to see changes in hydrogen-bond patterns, new distribution of salt bridges and some differences in the first solvation shell induced by the increasing amount of the catalytic product.

Acknowledgement: This research is funded by IDUB N. Copernicus #MEMOBIT grant. ICNT UMK computer facilities are acknowledged.

[1] Z. Cheng, Y. Xia, Z. Zhou, Front Bioeng and Biotech 2020, 8.

[2] Z. Cheng, Y. Lan, J. Guo, D. Ma, S. Jiang, Q. Lai, Z. Zhou, L. Peplowski, Molecules 2020, 25, 4806;

[3] J. Guo, Z. Cheng, B. J. Berdychowska, X. Zhu, L. Wang, L. Peplowski, Z. Zhou, Int J Biol Macromol 2021.

[4] H. Yamada, M. Kobayashi, Biosci Biotechnol Biochem 1996, 60, 1391-1400.

Email: jaber@doktorant.umk.pl