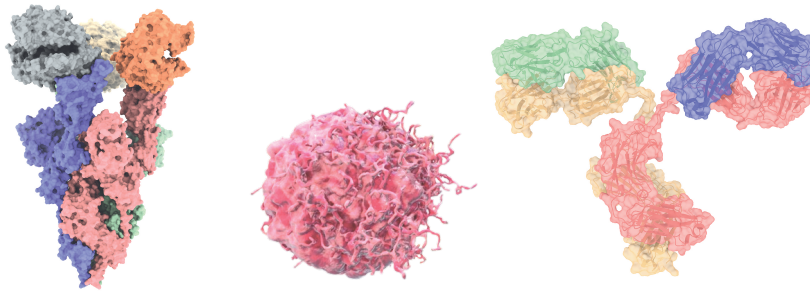


Prague Protein Spring 2023

PRAGUE PROTEIN SPRING

Organizers and Scientific Committee

Jiri Vondrasek
IOCB Prague, Czech Republic
Jiri.Vondrasek@uochb.cas.cz

Mikael Oliveberg
Stockholm University, Sweden
mikael.oliveberg@dbb.su.se

Michele Vendruscolo
University of Cambridge, UK
mv245@cam.ac.uk

www.pragueproteinspring.cz

ISBN 978-80-86241-70-8

Institute of Organic Chemistry and Biochemistry of the Czech Academy of Sciences
ELIXIR Czech Republic
SCIENTIFIKA

No professional language editing of the abstract content was performed.
Scientific and legal responsibility for the abstracts belongs to the authors.

ISBN 978-80-86241-70-8

Summary of content

Scientific Programme

Thursday 4 May 2023

Opening: **Jiří Vondrášek - IOCB Prague, Mikael Oliveberg - Stockholm University, Michele Vendruscolo - University of Cambridge**

Session 1 (18:00 - 19:30)
Chairman: Michele Vendruscolo

18:00 - 18:45 **Alan Fersht - Gonville and Caius College, University of Cambridge, UK**
From engineering mutations to study folding and stability to drugging p53 oncogenic mutants

18:45 - 19:30 **Ken Dill - Stony Brook University, USA**
How protein folding may hold the key to origin of life

19:30 - 21:00 Welcome drink, reception

Friday 5 May 2023

Session 2 (9:00 - 12:30)
Chairman: Gabriella Heller

9:00 - 9:45 **George Makhatadze - Rensselaer Polytechnic Institute, USA**
Novel Tool for Biophysical Studies of Amyloidogenesis

9:45 - 10:30 **Joan-Emma Shea - University of California, Santa Barbara, USA**
Self-Assembly of the Tau Protein: Liquid-Liquid Phase Separation and Fibrillization

10:30 - 11:00 Coffee break

11:00 - 11:45 **Michele Vendruscolo - University of Cambridge, UK**
Targeting protein aggregation in misfolding diseases

11:45 - 12:30 **Gary Pielak - University of North Carolina at Chapel Hill, USA**
Protein stability in living cells & under crowded conditions in vitro

12:30 - 14:00 Lunch

Session 3 (14:00 - 18:15)

Chairman: Joan-Emma Shea

14:00 - 14:45 **Ellinor Haglund - University of Hawai'i at Mānoa, USA**
The folding and function of proteins with a complex topology

14:45 - 15:30 **Wim Vranken - Vrije Universiteit Brussel, BE**
Defining and predicting conformational variability

15:30 - 16:15 **Mihaly Varadi - EMBL-EBI, UK**
Navigating the Flood of Predicted Protein Structures: Challenges and Opportunities for Understanding Disease

16:15 - 16:45 Coffee break

16:45 - 17:30 **Jiří Vondrášek - Institute of Organic Chemistry and Biochemistry of the CAS, CZ**
Domain context determines function in chimeric multidomain proteins

17:30 - 18:15 **Brian Baker - University of Notre Dame, USA**
The structural biophysics of specificity in cellular immunity

Sponsors block (18:15 - 18:55)

18:15 - 18:30 NanoTemper

18:30 - 18:45 Peptone Ltd.

18:45 - 19:30 Dinner

19:00 - 22:00 Poster session

Saturday 6 May 2023

Session 4 (9:00 - 12:30)
Chairman: Ellinor Haglund

9:00 - 9:45 **Silvio Tosatto - Università di Padova, IT**
CAID 2: lessons from the second critical assessment of protein intrinsic disorder prediction

9:45 - 10:30 **Benjamin Schuler - University of Zurich, CH**
Probing the dynamics and interactions of disordered proteins with single-molecule spectroscopy: From disordered complexes to phase separation

10:30 - 11:00 Coffee break

11:00 - 11:45 **Gabriella Heller - University College London, UK**
Drugging disordered proteins using NMR and MD

11:45 - 12:30 **Virginia Burger - New Equilibrium Biosciences, USA**
Targeting intrinsically disordered proteins with small molecule drugs

12:30 - 14:00 Lunch

Session 5 (14:00 - 18:15)
Chairman: Virginia Burger

14:00 - 14:45 **Paul Robustelli - Dartmouth College, USA**
Molecular recognition mechanisms of intrinsically disordered proteins

14:45 - 15:30 **Kamil Tamiola - Peptone Ltd., CH**
Rational modulation of intrinsic disorder in cytokines using generative AI

15:30 - 16:15 **Klára Hlouchová - Charles University, CZ**
Towards artificial proteins and alternative alphabets

16:15 - 16:45 Coffee break

16:45 - 17:30 **Konstantinos Tripsianes - Masaryk University, CZ**
Order-disorder continuum in protein function and disease

17:30 - 18:15 **Mikael Oliveberg - Stockholm University, SE**
Functional optimisation of diffusive protein-protein interactions in live cells

20:00 - 23:00 Conference dinner

Sunday 7 May 2023

Departures

Alan Fersht	University of Cambridge (UK)
Ken Dill	Stony Brook University (USA)
George Makhatadze	Rensselaer Polytechnic Institute (USA)
Joan-Emma Shea	University of California, Santa Barbara (USA)
Ellinor Haglund	University of Hawaii at Manoa (USA)
Benjamin Schuler	University of Zurich (CH)
Wim Vranken	Vrije Universiteit Brussel (BE)
Silvio Tosatto	University of Padova (IT)
Mikael Oliveberg	Stockholm University (SE)
Michele Vendruscolo	University of Cambridge (UK)
Jiří Vondrášek	Institute of Organic Chemistry and Biochemistry, CAS (CZ)
Klára Hlouchová	Charles University (CZ)
Brian Baker	University of Notre Dame (USA)
Paul Robustelli	Dartmouth College (USA)
Kamil Tamiola	Peptone Ltd. (CH)
Gabriella Heller	University College London (UK)
Virginia Burger	New Equilibrium Biosciences (USA)
Gary Pielak	University of North Carolina at Chapel Hill (USA)
Mihaly Varadi	EMBL-EBI (UK)
Konstantinos Tripsianes	Masaryk University (CZ)

Lectures

From engineering mutations to study folding and stability to drugging p53 oncogenic mutants

Alan Fersht - Gonville and Caius College, University of Cambridge, UK

My career as a protein scientist has extended across 7 decades, from the solution of the first crystal structures of proteins in the 1960s to the prediction of structures using AlphaFold in the 2020s, punctuated by the introduction of recombinant DNA technology and protein engineering, computer simulations and sophisticated biophysical methods. I will give a brief account of attitudes of the times and how I have used the procedures in to design molecules to rescue oncogenic mutants of the tumour suppressor p53.

How protein folding may hold the key to origins of life

Ken Dill - Stony Brook University, USA

No one knows how life arose from prebiotic chemistry 3.5 billion years ago. Some people believe it was an RNA world, since some RNAs can self-replicate. But, we have recent theory and modeling that shows that the first molecules may have been peptides and the process may have been folding. For one thing, the needle-in-a-haystack nature of the origins problem is solved by funnels in sequence space because of the energetics of funneled folding in conformational space. For another, the first life required a Darwinian-evolution-like propagation process, which needs to select on function and phenotype. The catalytic power of proteins provides the cooperativities needed to breakover from degradation to long-term persistence. This mechanism makes some testable predictions.

Novel Tool for Biophysical Studies of Amyloidogenesis

George Makhatadze - Rensselaer Polytechnic Institute, USA

Insoluble protein aggregates called amyloid fibrils are linked to various human diseases such as Alzheimer's disease, Parkinson's disease, and type II diabetes. Our research demonstrates that regardless of their amino acid sequence, green fluorescent protein (GFP) can bind the core of amyloid fibrils. The interaction between GFP and amyloid fibrils is specific, as GFP does not interact to the monomeric forms of these peptides/proteins or other protein aggregates. GFP cannot bind to amyloid fibrils formed by large proteins that have a fuzzy coat surrounding the amyloid core. However, if protease digestion removes the fuzzy coat, GFP can detect the amyloid core of these fibrils. By using NMR, we have mapped the binding interface on GFP and validated it through mutagenesis. We proposed a FRET-based sensor for amyloid detection based on GFP's ability to bind amyloid fibrils for practical applications. The feasibility of this approach was established by testing the prototype construct.

Self-Assembly of the Tau Protein: Liquid-Liquid Phase Separation and Fibrillization

Joan-Emma Shea - Department of Chemistry and Biochemistry & Department of Physics, University of California, Santa Barbara, USA

Tau is an intrinsically disordered protein that plays an important role in stabilizing microtubules. Under pathological conditions, this protein can also self-assemble into fibrillar structures, a process that has been associated with a class of neurodegenerative diseases known as Tauopathies. Interestingly, this protein is also capable of assembling into liquid droplets through a process of liquid-liquid phase separation (LLPS). Using a combination of field-theoretic simulations, coarse-grained models, and atomistic simulations, we present an investigation of the mechanisms of fibrillization and phase separation of this protein. We investigate fragments of Tau that have a propensity to either phase separate or form fibrils, enabling us to shed light into the sequence characteristics linked with these two modes of assembly. Finally, we introduce a 19-residue fragment of Tau that is capable of seeding the fibrillization of full-length Tau, and we discuss the effect of point mutations in modulating aggregation in familial forms of Tauopathies.

Targeting protein aggregation in misfolding diseases

Michele Vendruscolo - Department of Chemistry, University of Cambridge, UK

The phenomenon of protein misfolding and aggregation is associated with a wide range of human disorders, including Alzheimer's and Parkinson's diseases. A central role in these conditions is played by protein misfolded oligomers, which are among the most cytotoxic products resulting from the process of protein aggregation. It has been very challenging, however, to target these oligomers with therapeutic compounds, because of their dynamic and transient nature. To overcome this problem, I will describe a kinetic-based approach, which enables the discovery and systematic optimization of compounds that reduce the number of oligomers produced during an aggregation reaction. I will illustrate this strategy for the amyloid beta peptide, which is closely linked to Alzheimer's disease. As this strategy is general, it can be applied to oligomers of any protein in drug discovery programmes.

Protein stability in living cells & under crowded conditions in vitro

Gary Pielak - University of North Carolina at Chapel Hill, USA

The crowded and complex environment in cells is predicted to affect protein behavior compared to dilute buffer. My laboratory and our collaborators are examining crowding effects on the stability of proteins and their complexes in cells and under physiologically-relevant crowded conditions using NMR. A challenge in these endeavors is detecting the test protein in a sea of crowders. 19F-NMR is ideally suited to overcome this challenge and provide high quality data on folded- and unfolded- proteins as well as free and bound forms of complexes. I will focus on equilibrium data acquired in living Escherichia coli cells, Xenopus laevis and Danio rerio oocytes as well as in vitro in concentrated cosolute solutions. The cosolutes include synthetic polymers and their monomers, other proteins and lyophilized cytosol. The results show that crowding affects folding and binding in ways not always correctly predicted by simple models, and I will briefly introduce a new model that explains the data.

The folding and function of proteins with a complex topology

Ellinor Haglund - University of Hawai'i at Mānoa, USA

Folding of proteins into their active 3D-structure occurs spontaneously or is assisted with the help of chaperones within a biologically reasonable time, from micro- to milliseconds. It occurs within different compartments of the cell, controlled by the chemical environment. When folding goes wrong in cells, misfolded and/or aggregated proteins may arise, unable to perform their specific biological function. The correlation between structural motifs and their 3D-structure has been established to influence biology. However, less is known about the biological implications of protein topology, i.e., motifs that can act as a structural switch in response to environmental changes. Leptin is the founding member of the Pierced Lasso Topology (PLT), a newly discovered protein family sharing the unique features of a "knot-like" topology. A PLT is formed when the protein backbone pierces through a covalent loop formed by a single disulfide bond. PLTs are found in all kingdoms of life, with 14-different biological functions, found in different cell compartments. Despite the large number found in nature, where more than 600 proteins have been found with a PLT, a connection between topology and biological function has not yet been determined. Using the hormone leptin as our model system, we investigate the folding, aggregation, and/or misfolding of the hormone leptin, and the association between the threaded topology and the biological function. The results show that the threaded topology shift the transition state ensample and adds ruggedness on the folding free energy landscape with and important role for biological signaling. Interestingly, the chemical environment controls the threaded topology, and thus, we propose that PLTs may act as a molecular switch to control activity in vivo.

Defining and predicting conformational variability

Wim Vranken - Vrije Universiteit Brussel, BE

The dynamics and related conformational changes of proteins are often essential for their function, but are difficult to characterise and interpret. Protein structure is not fixed, with the thermodynamic landscape that determines their conformational behavior determined by their local environment. So how should we define conformation and dynamics? Amino acid residues that do change conformation under native conditions, for example, can not be assigned to a specific secondary structure, but should rather be interpreted in probabilistic terms within the secondary structure space where it transitions. We here propose an alternative approach to describe protein conformation and dynamics based on five different conformational regions (helix, surrounding helix, sheet, surrounding sheet and other) defined from NMR ensemble and chemical shift data. A further conformational variability metric describes how often a residue might move between these regions.

Navigating the Flood of Predicted Protein Structures: Challenges and Opportunities for Understanding Disease

Mihaly Varadi - EMBL-EBI, UK

The massive influx of predicted protein structures in public databases presents challenges and opportunities for researchers worldwide. While experimentally-determined protein structures have long proved invaluable in understanding why a protein is implicated in a disease, the amount of data in the Protein Data Bank is a limitation. Since 2022, over 800,000,000 predicted protein structures have been available in public databases, presenting a significant opportunity to gain new insights. However, traditional tools in structural bioinformatics often struggle to handle this vast amount of data. Moreover, most of the predicted structures lack biological context, limiting their usefulness. In this presentation, we highlight the efforts made by data services teams worldwide to address this situation and help with accessing data, finding relevant information, and making sense of it. We also underscore the limitations and caveats researchers must consider when working with the new AI-based predicted structures. Finally, we will emphasise the importance of molecular structure data, both experimentally determined and computationally predicted, in understanding diseases.

Domain context determines function in chimeric multidomain proteins

Jiří Vondrášek - Institute of Organic Chemistry and Biochemistry of the CAS, CZ

Allosteric modulation of a protein domain function is one of the most intriguing concepts in molecular biology nowadays . PDZ domains represent ideal candidate to decipher mechanism of binding specificity via adjacent domains synergy in multi-domain protein context by means of structural, computational and biophysical methods. We focused on understanding of PDZ allostery brought by adjacent domain together with folding, structure, dynamics and binding properties of the fusion proteins. The protein domains in questions are PDZ3 from Zonula Occludens (ZO-1) protein (homo sapiens) in various 2 domains fusion constructs. We addressed a question of the mechanism in which PDZ3 specificity is controlled by a character, position and size of the attached domain. This can make possible to contribute to the understanding of the domains coevolution phenomenon.

The structural biophysics of specificity in cellular immunity

Brian Baker - University of Notre Dame, USA

T cells orchestrate cellular immunity by recognizing peptide antigens bound and presented by major histocompatibility complex (MHC) proteins. While high specificity is a hallmark of the immune system, T cell receptors (TCRs) are also widely cross-reactive. The presence of both specificity and cross-reactivity in TCR recognition has been challenging to understand. Moreover, TCR cross-reactivity contributes to the development of autoimmunity, transplant rejection, and off-target recognition in T cell therapy. Here we demonstrate that using well known structural and biophysical considerations, the "enigma" between TCR specificity and cross-reactivity can be well understood, providing a platform for engineering, predicting, and optimizing the specificity of T cell immunity.

CAID 2: lessons from the second critical assessment of protein intrinsic disorder prediction

Silvio Tosatto - BioComputing Laboratory, Department of Biomedical Sciences, Università di Padova, IT

Protein intrinsic disorder (ID) is a complex and context-dependent phenomenon that covers a continuum between fully disordered states and folded states with long dynamic regions. The lack of a ground truth that fits all ID flavors and the potential for order-to-disorder transitions depending on specific conditions makes ID prediction challenging. The second round of the Critical Assessment of protein Intrinsic Disorder prediction (CAID 2) challenge aimed to evaluate the performance of different prediction methods across different benchmarks, leveraging the annotation provided by the DisProt database, which stores the coordinates of ID regions when there is experimental evidence in the literature.

The CAID 2 challenge demonstrated varying performance of different prediction methods across different benchmarks, highlighting the need for continued development of more versatile and efficient prediction software. Depending on the application, researchers may need to balance performance with execution time when selecting a predictor. AlphaFold seems to be a good ID predictor but it is better at detecting absence of order rather than ID regions as defined in DisProt.

The CAID 2 predictors can be freely used through the CAID Prediction Portal, and CAID has been integrated into OpenEBench, which will become the official platform for running future CAID challenges.

Probing the dynamics and interactions of disordered proteins with single-molecule spectroscopy: From disordered complexes to phase separation

Benjamin Schuler - Department of Biochemistry, University of Zurich, CH

The functions of proteins have traditionally been linked to their folded structures, but many proteins perform essential functions without being folded. Quantifying the highly dynamic and conformationally diverse ensembles of these intrinsically disordered proteins (IDPs) is an important aspect of understanding their functional mechanisms. I will focus on highly charged IDPs and illustrate how single-molecule spectroscopy combined with simulations can be used to probe their dynamics, interactions, and phase separation.

Drugging disordered proteins using NMR and MD

Gabriella Heller - Department of Structural and Molecular Biology, Division of Biosciences, University College London, UK

Intrinsically disordered proteins (IDPs) and proteins with intrinsically disordered regions (IDRs) are prevalent biomolecules (30-40% of human proteins) that rapidly interconvert between many structural conformations. Because they lack long-lived rigid drug binding sites for small molecules, IDPs and IDRs are considered 'undruggable' by the mainstream pharmaceutical industry despite being associated with many diseases, including cancer, cardiovascular disease, diabetes, neurodegeneration, and diseases caused by viruses (global costs >2T USD). Recent evidence suggests that IDPs and IDRs can indeed interact with small, drug-like molecules, however, there are limited approaches to characterize these interactions. In the first part of my talk, I will describe integrative methods, combining Molecular Dynamics (MD) simulations with experimental data from Nuclear Magnetic Resonance (NMR) spectroscopy and other biophysical experiments to characterize the druggability of the monomeric form of the amyloid-β peptide, whose aggregation is a hallmark of Alzheimer's disease. I will detail the identification and characterization of a small, drug-like molecule that sequesters amyloid-β in its soluble monomeric form and prevents its toxic aggregation. We observe the binding to be extremely dynamic, such that amyloid-β not only remains disordered in the bound form, but also increases its conformational entropy. In the second half of my talk, I will share our findings on NMR-based experimental techniques to quantitatively detect and characterize small-molecule interactions with disordered proteins, using the disordered domains 2 and 3 from the non-structural protein 5A (NS5A-D2D3) in hepatitis C virus as a model system. I will demonstrate that several solution-state NMR observables including protein- and ligand-detected chemical shifts are insensitive to the binding of NS5A-D2D3 with a small molecule. Nevertheless, dynamic NMR methods are highly sensitive to this interaction. As IDPs and IDRs are increasingly recognized to bind small molecules, we anticipate that MD and dynamic NMR experiments will play key roles in identifying and characterizing new interactions. Furthermore, detailed understandings of mechanisms underpinning binding may lead to the ability to engineer improved interactions, yielding drugs targeting these dynamic biomolecules that are prevalent in human disease.

Targeting intrinsically disordered proteins with small molecule drugs

Virginia Burger - Co-Founder and CEO of New Equilibrium Biosciences, USA

The class of intrinsically disordered proteins (IDPs) is commonly associated with disease, comprising over two-thirds of those proteins associated with cancers, as well as many well-established proteins linked to neurodegenerative disorders. This link to disease makes IDPs desirable drug targets, yet very few drug candidates have reached the clinic that target IDPs. Because of the small number of well-validated, published molecules that regulate IDPs, there has been limited research on the chemical features of small molecules that selectively bind specific IDPs. Disordered proteins have distinct ranges of various chemical properties, such as absolute net charge and hydrophobicity, when compared to ordered proteins, indicating that the molecules that bind disordered proteins may also have different ranges of chemical properties than the molecules that bind ordered proteins. To determine chemical features common to molecules binding IDPs as well as chemical features that might enable selective binding to specific IDPs, we performed a large-scale screen of 30 billion small molecules against several constructs of three IDPs via a DNA-encoded library (DEL). We analyze the findings from the screen to determine whether any chemical features are common to IDP binders. An understanding of chemical properties that promote binding to IDPs can inform the design of IDP-specific screening libraries.

Molecular recognition mechanisms of intrinsically disordered proteins

Paul Robustelli - Department of Chemistry Dartmouth College, USA

Intrinsically disordered proteins (IDPs), which represent ~40% of the human proteome, play crucial roles in a variety of biological pathways and biomolecular assemblies and have been implicated in many human diseases. IDPs do not fold into a well-defined three-dimensional structure under physiological conditions. Instead, they populate a dynamic conformational ensemble of rapidly interconverting structures. As a result, IDPs are extremely difficult to experimentally characterize and are largely currently considered 'undruggable' by conventional structure-based drug design methods. Here I will discuss the development of methods to define conformational ensembles of IDPs from experimental data, algorithms to structurally and kinetically cluster heterogenous conformational ensembles of IDPs and the application of these methods to understand how IDPs bind their physiological interaction partners and small molecule drugs.

Rational modulation of intrinsic disorder in cytokines using generative AI

Kamil Tamiola - Founder and CEO at Peptone, Peptone Inc., CH

Cytokines are prime engineering and therapeutic targets given their pivotal roles in immunity. However, the progress and adoption of clinical use of cytokines have been hampered by their short blood half-lives and severe side effects caused by low specificity and tissue targeting challenges. In my talk, I will present an AI driven and unsupervised approach to engineering of a biologically potent yet stable variants of IL-21 cytokine.

Towards artificial proteins and alternative alphabets

Klára Hlouchová - Department of Cell Biology, Faculty of Science, Charles University, CZ

All extant cells known to humankind build proteins from the same 20 coded amino acids. The sequence space of the canonical amino acids provides far more options for de novo gene birth and design and is being extensively exploited. Nevertheless, synthetic biology and the study of origins of life imply that functional proteins can be built using smaller and energetically less costly alphabets and that inclusion of non-canonical amino acids can escape the Central Dogma of all life and be of great potential in biotechnology and therapeutic strategies.

While AI-based predictors have learnt to work with the canonical alphabet, the rules guiding the properties of restricted and extended alphabets remain challenging. To approach this, we study highly combinatorial libraries composed of different amino acid repertoires. I will summarize the knowledge we have gained from the study of early origins of life, describe the biophysical properties of sequences with selected non-canonical amino acids and discuss our approaches towards designing novel xeno-alphabets for building artificial proteins.

Functional optimisation of diffusive protein-protein interactions in live cells

Mikael Oliveberg - Institutionen för biokemi och biofysik, SE

he function and fitness of a living cell is orchestrated by dynamic interactions between intracellular proteins. For any given protein, these interactions include numerous transient encounters with the crowded background during diffusive search, specific binding to functional partners and various catalytic/conformational motions. Using a combination of live-cell NMR analysis, first-principle theory and physicochemical analysis of proteomes across all organisms with sequenced genomes, we have distinguished a second level of generic protein-protein interactions providing a new perspective on cellular function and evolution. The interactions in question are primarily of electrostatic origin, unspecific and only manifest themselves under densely crowded conditions. Their biological role is to assure inter-molecular repulsion to maintain the cytosolic components suitably fluid for cellular function. Consistently, the proteome-charge composition that promotes this repulsion is also under evolutionary control and shows divergent optimization across different types of organisms. The most intriguing aspect of these findings, however, is not that the proteome charge is so precisely tuned that it presents an alternative base for detailed evolutionary classification, but that it challenges our current understanding of the intracellular conditions. Instead of just presenting steric crowding, the 'non-functional' parts of protein surfaces actively control cellular function, and this control renders the intracellular environment more different from standard physiological buffer than previously anticipated.

Order-disorder continuum in protein function and disease

Konstantinos Tripsianes - CEITEC Brno, CZ

According to the "positional information paradigm", if natural selection is acting to preserve a biological function, changes in the sequence that alter that function will be removed from the population over evolutionary time. If allowed to implicate stabilizing selection for the evolution of disordered regions as we do for the folded domains, then the concept of my presentation is as follows: neither disordered regions nor structured domains are independent of one another, but sense and respond to each other and have co-evolved to have some synergistic properties that are required to enable key biological functions. I will focus on the intimate relationship between Ck1ε and DVL proteins, a kinase-substrate affair that controls WNT signal transduction in development and homeostasis. Our mechanistic data highlight a continuum of structural states that bridges two highly different dynamic regimes and regulates protein function. Disturbing the thin line between structure and disorder may result in disease phenotypes of animal models.

Posters

List of posters

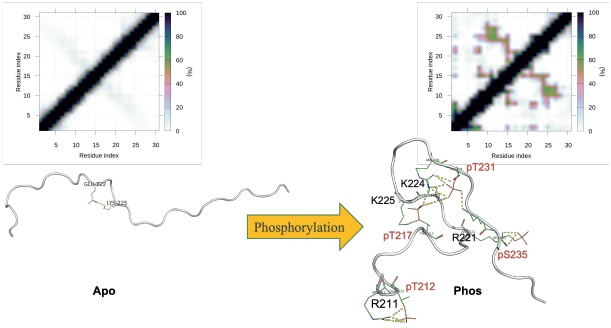
1. **Molecular dynamic simulation study of multiple phosphorylations at the proline rich region of tau(210-240) peptide**
   Krishnendu Bera - CEITEC MU, Department of Chemistry and National Centre for Biomolecular Research, Faculty of Science, Masaryk University
2. **Simulation of oligosaccharide binding to HEV32 domain**
   Jan Beránek - University of Chemistry and Technology, Prague
3. **Predicting ion binding sites in proteins using machine learning**
   Christos Feidakis - Charles University, Faculty of Science, Department of Cell Biology
4. **Experiment Insight in Preferential Interactions in Bio-Soft Matter**
   Jan Heyda - University of Chemistry and Technology, Prague
5. **Unraveling Protein Patterns: The Unexplored Potential of Large Language Models**
   Tomáš Honzík - University of West Bohemia
6. **Engineering of a novel two-enzyme fusion system for PET degradation**
   Jiraskova Katerina - Institute of Organic Chemistry and Biochemistry of the CAS
7. **Protein substrate encapsulation in the GroEL/ES cage causes its destabilization due to a reduced hydrophobic effect**
   Ilia Korobko - Weizmann Institute of Science, Israel
8. **Using a Novel ANN to Design Artificial Promoter-Operator Constructs**
   Lukáš Kuhajda - University of West Bohemia
9. **Inhibitors of the mitochondrial rhomboid protease PARL modulate the PINK1/Parkin axis enhancing mitophagy**
   Denise-Liu` Leone - Institute of Organic Chemistry and Biochemistry of the CAS
10. **Specific ion effects on amino acid dissociation equilibria**
    Varun Mandalaparthy - Technical University Darmstadt
11. **New protein reactive centers against aspartic proteases**
    Martina Mičková - Institute of Organic Chemistry and Biochemistry of the CAS; Faculty of Science, Charles University
12. **Dances with enzymes: Multiscale simulation of glycolytic enzyme assembly formation**
    Tom Miclot - Department of Computational Chemistry, J. Heyrovský Institute of Physical Chemistry

13. **TRANSIENT OLIGOMERIZATION AND AGGREGATION OF HUMAN LEPTIN**
    Grace Orellana - University of Hawaii at Manoa
14. **Engineering of autocatalytic activity of HIV-1 protease**
    Klára Poštulková - Institute of Organic Chemistry and Biochemistry of the CAS
15. **Sequence fingerprint of structured and disordered proteins**
    Patrícia Sotáková - Institute of Organic Chemistry and Biochemistry of the CAS
16. **Protein tunnels, channels and pores, ChannelsDB & MOLEonline**
    Anna Špačková - Placký University Olomouc
17. **Dynamical allostery of multi-domain protein complexes**
    Josef Šulc - Institute of Organic Chemistry and Biochemistry of the CAS
18. **Analysis and Sampling of Molecular Simulations by adversarial Autoencoders**
    Guglielmo Tedeschi - University of Chemistry and Technology, Prague.
19. **Linking the Diffusive Dynamics of Bacterial Proteome to Cell Metabolism and Death**
    Štěpán Timr - J. Heyrovský Institute of Physical Chemistry
20. **To mistranslate or not to live**
    Vjaceslav Tretjacenko - Weizmann Institute of Science
21. **Proteolytic profiles of human AMBN by MMP-20 and KLK-4 proteases**
    Veronika Vetýšková - Institute of Organic Chemistry and Biochemistry of the CAS
22. **Accurate modeling of free energy landscapes enables "Unstructure"-based drug discovery for IDPs**
    Bentley Wingert - New Equilibrium Biosciences

Molecular dynamic simulation study of multiple phosphorylations at the proline rich region of tau(210-240) peptide

Krishnendu Bera[1,2,3], Alessia Lasorsa[4], Isabelle Landrieu[4], Jozef Hritz[1,2] - [1]CEITEC MU, Masaryk University, Brno, Czech Republic. [2]Department of Chemistry, Faculty of Science, Masaryk University, Brno, Czech Republic. [3]National Centre for Biomolecular Research, Faculty of Science, Masaryk University, Brno, Czech Republic.[4]Unité de Glycobiologie Structurale et Fonctionnelle, Lille, France.

Elucidating the conformational dynamics of intrinsically disordered proteins (IDPs) regulated by post translational modifications (PTMs) such as phosphorylation is challenging. Tau is a well-known IDP found hyper-phosphorylated in Alzheimer's disease (AD) in humans [1]. The proline-rich motif of tau(210-240) peptide directly interacts with several of it's binding partners proteins such as BIN1, 14-3-3 etc. All atoms molecular dynamic (MD) simulation studies have been performed for apo and four phosphorylated (212PThr, 217PThr, 231PThr, 235PSer) tau(210-240) peptide using three different temperature variants (278K, 298K and 310K) and two different force field parameters (AMBER99SB-ILDN and CHARMM36m) with TIP4PD water model. These two force fields parameters combine with TIP4PD water model close to experimental observations for multiple IDPs and IDR found from our group previous studies [2, 3]. From our MD simulations we observed, these four-phosphorylations cause increase in compactness of the peptide. The binding of partner proteins like BIN1 with tau may be altered by the strong salt bridges, forming nearby lysine and arginine due to the phosphorylation [4]. Phosphorylation induces a strong structural transition, with tau(210-240) favouring a bent conformation. The MD simulation results were verified using NMR experimental parameters like chemical shift and 3J-coupling. The experimental part has been carried out by our collaborator Prof. Isabelle Landrieu [5].



References:

1. Shimada, T. et al. (2013). BioMed research international, 2013; 2. Zapletal, V., et al. (2020) Biophysical Journal, 118(7), 1621-1633; 3. Pavlíková Přecechtělová, J. et al. (2019) Journal of Chemical Theory and Computation, 15(10), 5642-5658; 4. Chen, Y. et al. (2019) Journal of Molecular Neuroscience, 68, 620-630; 5. Lasorsa A.#, Bera K.# et al. (2023) Biochemistry (Accepted)

Acknowledgments:

Computational resources were supplied by the metacentrum and IT4 Innovations National Supercomputing Center (OPEN-17-7), project (e-INFRA CZ 90140 and LM2018140) provided within the program Projects of Large Research, Development and Innovations Infrastructures. KB and JH were funded by the Ministry of Education, Youth, and Sport of the Czech Republic (MEYS CR), grant number LTAUSA18168 (Inter-Excellence Inter-Action), Czech Science Foundation [GF20-05789L] and European Union Excellence Hub (101087124). KB is also supported by Brno Ph.D. Talent Scholarship – funded by the Brno City Municipality, Brno, Czech Republic.

Simulation of oligosaccharide binding to HEV32 domain

Jan Beránek - University of Chemistry and Technology, Prague

HEV32 is a 32-residue domain of a protein hevein, known for its ability to form complexes with chitin-based polysaccharides. It is a suitable model system for studying protein-saccharide interactions. We studied interaction of HEV32 domain with mono-, di- and trisaccharide derived from chitin using well-tempered funnel metadynamics. For each saccharide, four different combinations of collective variables were used. There were always two collective variables per a system, the first one describing the distance between the saccharide and the binding site, and the second one describing conformational changes in the HEV32 domain. For all systems, 2 μs long simulations were carried out and the free energy surface was obtained. Because of slow convergence in the cases of some combinations of the CVs, the affected simulations were prolonged to 3 μs. Standard binding free energy of all saccharide molecules to HEV32 were predicted from the free energy surfaces. For systems with trisaccharides, the observed in two cases that binding energies were in agreement with the experimentally determined value ≈ –22 kJ/mol [1]. In the other cases, the predicted binding energy was different, and this difference is discussed. Conformational changes taking place in HEV32 during binding of the saccharides was also characterized.

[1] N. Aboitiz, M. Vila-Perelló, P. Groves, J. Asensio, D. Andreu, F. Cañada, J. Jiménez-Barbero, ChemBioChem, 5, 1245-1255 (2004)

Predicting ion binding sites in proteins using machine learning

Christos Feidakis - Charles University, Faculty of Science, Department of Cell Biology

The study of protein–ligand interactions constitutes a prominent field in structural biology. Observing the effects of ligand binding, determining the temporal order of the conformational shifts that occur throughout the binding interaction, or exploring the specificity of a binding site, involve studying several protein–ligand interactions. Ions are particularly prominent among these interactions, with approximately half of all known proteins interacting with small acid radicals and metal ions in order to stabilize their structure and regulate their biological functions. The information within these interactions is key to understanding the underlying mechanisms that are involved in health and disease, and by extension, understanding newly discovered protein structures by assigning functional annotations to them. The advances in protein structure prediction have greatly expanded our protein structure repositories, highlighting the importance of being able to understand and annotate the new structures.

Here we focus on interactions between proteins and ions, and we develop a software tool to predict the position of ion binding sites in proteins by training upon the structural information in known interactions (P2Rank). Our workflow consists of building test and training datasets for each ion ligand, choosing appropriate features for learning, and optimizing a random forest-based, machine learning algorithm to get accurate predictions. In addition, we leverage a newly developed software (AHoJ), to build datasets of apo-holo pairs and test our predictor against naturally unbound structures, providing a realistic perspective to the field of ligand binding site prediction.

Experiment Insight in Preferential Interactions in Bio- Soft Matter

Jan Heyda, Daniel Ondo, Jakub Polák - University of Chemistry and Technology, Prague

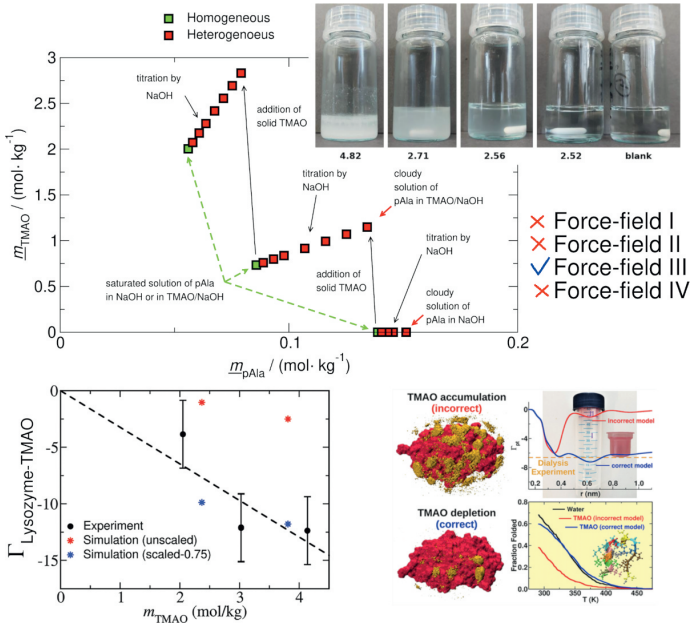
Computer simulations present a versatile tool to quantitatively characterize osmolyte-protein interactions. The subsequent thorough analysis allows describe dominant interactions at atomistic level. However, this approach is only as good as the underlying parametrization of simulation model.

Here, we present ensemble of experimental methods, which may be used to determine preferential interactions of salts and osmolytes with biomolecules – such as dialysis, osmometry, or solubility. These thermodynamic measures are stringent tests of force-field quality and transferability from one system to another.

Our protocols are presented on example of selected peptides: from model pentaAlanine via TrpCage minipeptide to real-size protein Lysozyme. Complemented by MD simulations.

Figure 1:

Examples of application of well targeted experimental techniques for determination of key properties which are decisive for selection or refinement of the osmolyte-water or osmolyte-protein force-fields.



Unraveling Protein Patterns: The Unexplored Potential of Large Language Models

Tomáš Honzík - University of West Bohemia

Large language models (LLMs), such as ChatGPT, have demonstrated remarkable success in various natural language processing tasks. However, their potential applications extend beyond human language, and in this study, we explore their capabilities in deciphering the complex "language" of proteins. Compared to human languages, the language of proteins is significantly more challenging for humans to understand, given its inherent complexity and vastness.

In our approach, we analyze the ESM-2 transformer model's predictions on the UniRef90 dataset, focusing on single amino acid and dipeptide predictions. From this analysis, we generate confusion matrices that reflect the consistency of these predictions with the chemical properties of the amino acids. Our findings suggest that LLMs have the capacity to provide alternative substitution matrices, such as BLOSUM, by capturing inherent patterns in protein sequences.

Furthermore, we demonstrate that the application of Bayes' theorem can improve the LLM prediction performance for masked dipeptide prediction. The Bayesian approach allows us to estimate the joint distribution of the dipeptides, providing a more accurate representation of their co-occurrence patterns in protein sequences.

Through the innovative application of LLMs to analyze protein properties beyond structure, our study highlights the untapped capability of these models in understanding the complex language of proteins. By revealing novel insights into amino acid substitution patterns and dipeptide co-occurrence, our findings have the potential to enhance computational tools used in protein analysis, which could impact various fields, including protein design, bioinformatics, and proteomics.

Engineering of a novel two-enzyme fusion system for PET degradation

Jiraskova Katerina, Postulkova Klara, Galgonek Jakub, Bousova Kristyna, Vondrasek Jiri - Institute of Organic Chemistry and Biochemistry of the CAS

KEYWORDS

polyethylene terephthalate (PET); biodegradation; PET-degrading enzymes (PDEs); PET hydrolyse (PETase); protein engineering; chimeric protein

ABSTRACT

Large amounts of PET waste accumulate in landfills where it is predicted to persist for centuries to millennia. Weathering and fragmentation of larger plastic products in the environment generate microplastics (MPs), which have been found in both water and soil, and more recently have been observed entrained in the air or in marine animals. Due to the possibility of entering the human body through the food chain, their persistence and bioaccumulation properties, MPs are presumed a serious threat to various organisms and human health. Biodegradation of plastics pollution is an emerging strategy offering an environmentally friendly route for waste recycling. Enzymatic breakdown of PET has been studied extensively with various bacterial PET-degrading enzymes identified. A significant milestone was the discovery of a bacterium *Ideonella sakaiensis* secreting enzyme PETase which converted PET with a relatively high activity. A number of its improved variants have been designed however the enzymatic properties remain insufficient for the industrial application. Several studies on dual-enzyme systems revealed synergy for increased PET depolymerization, enhanced further by linking the enzymes. The effectiveness of the fusion enzyme strategy was confirmed by Knott et al. 2020, who improved PET and its intermediate MHET turnover by linking PETase and MHETase into a chimeric protein. The aim of our project is a development of a novel highly effective two-enzyme fusion system capable to hydrolyze PET to monomers important for subsequent reuse in new products. This approach may contribute towards the future concept of a circular PET economy and thus protect human population from negative health effects from microplastics in the environment.

Protein substrate encapsulation in the GroEL/ES cage causes its destabilization due to a reduced hydrophobic effect

Dr. Ilia Korobko - Weizmann Institute of Science, Israel

Chaperonins are large molecular machines found in all kingdoms of life, which assist protein folding via a complex ATP-dependent allosteric mechanism. A paradigmatic chaperonin studied for over three decades is GroEL/ES from E. coli. It had remained unclear, however, whether substrate proteins encapsulated in the GroEL/ES cavity are stabilized or destabilized. Measuring the energy of the encapsulated substrate proteins would require overcoming two significant challenges (i) identifying a perturbation for measuring stability that does not affect the chaperonin itself, and (ii) the substrate protein of interest must remain encapsulated during the measurements. By using a novel strategy that circumvents the obstacles mentioned above, we have been able to measure, for the first time, the stability of a protein inside the cage. We found that a model protein substrate is destabilized by more than 5 kcal mol^{-1} in the GroEL/ES cavity compared to bulk solution. This finding was surprising since steric confinement is expected to have a stabilizing effect. In a further investigation, we introduced point mutations in the hydrophobic core of the model protein substrate removing one methylene group every time. Such mutations are expected to be destabilizing owing to the loss of the free energy of transfer of a methylene group from water to the protein's hydrophobic interior. The effects of most of these mutations were found, however, to be neutral or even stabilizing, thereby indicating a diminished hydrophobic effect in the GroEL/ES cavity. We suggest that the reduced hydrophobic effect is caused by water ordering due to the small number of hydration shells between the cavity and protein substrate surfaces. Our findings are likely to be relevant to other systems that involve encapsulation in confining compartments.

Using a Novel ANN to Design Artificial Promoter-Operator Constructs

Lukáš Kuhajda - University of West Bohemia

In this work, unreliable expert methods for inserting operators into promoters are replaced using neural networks with a newly designed structure.

Promoter sequences have been incorporated into artificial neural networks (ANNs) in recent years, resulting in many specialized models. However, none of these models provide insights into gene expression under variable conditions or allow for the incorporation of novel regulatory elements into the promoter. Creating synthetic genetic circuits involves composing promoters with operators, but the current method still relies on inefficient expert methods that are time-consuming and costly.

This work proposes a new concept of self-supervised ANN learning, where the trained model can directly propose genetic modifications to promoters without further training. The model is trained with two inputs: a short (cut-out) part of the promoter (16-64bp) and the main promoter sequence (400bp), with the cut-out part replaced with a sequence from a foreign promoter. The ANN learns to insert/overwrite the cut-out part into the promoter based on context, as if it naturally belonged there.

To design a laboratory experiment, the model only requires the promoter sequence and operator sequence as inputs, and it outputs a proposal for where to insert the operator into the promoter. The model was tested using the organism *Saccharomyces cerevisiae*. The operator for the repressing bacterial transcription factor *TetR* was successfully inserted into five wild-type yeast promoters, resulting in significant repression in 3 out of 5 cases without disrupting the natural function of the promoter (other than expression strength, nothing was known about the promoters a priori).

This initial test confirms the functionality of the designed and trained model, which is capable of proposing transformations that lead to a change in the original sequence's logic. The model's functionality is being expanded, which could significantly streamline work in the laboratory.

Inhibitors of the mitochondrial rhomboid protease PARL modulate the PINK1/Parkin axis enhancing mitophagy

Denise-Liu` Leone, Kathrin Bach, Kvido Stříšovsky - IOCB, Institute of Organic Chemistry and Biochemistry, Prague

Rhomboids are serine intramembrane proteases that fulfill manifold physiological functions in all domains of life ranging from quorum sensing in bacteria to cell signaling and membrane protein quality control. [1,2] Crystal structure analysis has yielded first glimpses into the catalytic mechanism, and rhomboids have been implicated in various disease contexts. The mitochondrial rhomboid protease PARL regulates mitophagy via the PINK1/Parkin pathway by cleaving PINK1 and PGAM5 in the inner mitochondrial membrane. Mutations in this pathway have been associated with Parkinson's disease, and PARL is a potential drug target for modulating PINK1/Parkin-dependent mitophagy.
We have discovered that peptidyl ketoamides are potent inhibitors of rhomboid proteases [3], and now have developed ketoamides targeting PARL. These compounds comprise peptide derived from PINK1 (a natural substrate of PARL) and a hydrophobic substituent (tail) at the ketoamide nitrogen.[3,4] The efficacy and selectivity of the inhibitors was evaluated in vitro using a novel in vitro fluorescence-based assay and cell-free translated rhomboid. One of the generated ketoamides showed high potency against PARL in cellular assays and induced effects comparable to a PARL knockdown on the PINK1/Parkin pathway without unspecific toxic effects. Hence, ketoamides can be used as a chemical tool for the modulation of mitophagy in the context of cell biological studies and the search for therapeutic strategies for Parkinson's disease. We are currently developing approaches for efficient mapping of sequence preferences of rhomboid proteases, and we explore the chemical space of the tail substituent to elevate the potency and selectivity of ketoamide inhibitors of rhomboid proteases further. [5]

[1] Lemberg MK, Strisovsky K. Mol Cell. (2021), 81(12):2507-2519.
[2] Began J, et al. The EMBO Journal (2020) 39: e102935.
[3] Tichá, A., Stanchev, S., ... Majer, P., & Strisovsky, K. (2017). Cell Chem. Biol., 24(12), 1523-1536.
[4] Poláchová, E., Bach, K., Heuten, E., Stanchev, S., …, Lemberg M. & Stříšovský, K. (2023). J. Med. Chem. 66(1), 251- 265.
[5] Kasperkiewicz P, Poreba M, Snipas SJ, Parker H, Winterbourn C, Drag M. (2014) PNAS 111 (7) 2518-2523.

Specific ion effects on amino acid dissociation equilibria

Varun Mandalaparthy - Technical University Darmstadt

Proteins are essential biological macromolecules whose stability and function are influenced by various physiological factors, including pH and ionic strength. Different ions can have qualitatively different effects on protein stability, as described by the Hofmeister series. Salts generally reduce the pKa of weak acids in solution (i.e. the weak acids dissociate more favorably) by a mechanism involving the screening of the deprotonated state charge. However, the magnitude of this effect is sensitive to changes in the ion identity.

In this study, we utilize Constant pH MD simulations to study the salt-dependent behavior of the amino acid Aspartic acid (Asp) at different pH values. We calculate the pKa of Asp in the presence of a fixed salt concentration. Hofmeister effects were interrogated by systematically varying the size of the ions (σ in the Lennard-Jones (LJ) potential) while keeping the charge (q) and interaction strength (ε in the LJ potential) fixed. This allows us to study strongly hydrated (high charge density) and weakly hydrated (low charge density) ions. Increasing the anion size leads to a monotonic decrease in the pKa of Asp while changes in the cation size present non-monotonic trends. In particular, the cation size effects on the pKa are sharply peaked, resembling an inverted volcano plot. This effect may be understood in context of the Law of Matching Water Affinity which states that binding (and therefore screening) is optimal if the cation and the deprotonated Asp have similar hydration free energies.

Overall, our study allows us to investigate Hofmeister effects on the stability of proteins while considering physiologically relevant pH-dependent variable protonation states of amino acids.

New protein reactive centers against aspartic proteases

Martina Mičková - 1) Institute of Organic Chemistry and Biochemistry of the Czech Academy of Sciences, Flemingovo n. 2, Prague, Czech Republic; 2) Faculty of Science, Charles University, Albertov 6, 128 00 Praha 2

Protein inhibitors of aspartic proteases are rare in nature, and only two of them have been structurally characterized so far in complex with a target protease. Here we present new crystal structures of cathepsin D in complex with three proteinaceous inhibitors: potato inhibitors pAPI-1 and pAPI-2 from the plant Kunitz family, and equistatin domain 2 (Eqd2) from sea anemone belonging to the thyropin (thyroglobulin type-1-like) family. The inhibitors show distinct designs of structural binding motifs that are based on disulfide-stabilized loops forming a network of interactions in the extended non-primed part of the enzyme active site. Our results will help in the development of biomimetic inhibitors of medicinally relevant aspartic proteases.

Dances with enzymes: Multiscale simulation of glycolytic enzyme assembly formation

Tom Miclot - Department of Computational Chemistry, J. Heyrovský Institute of Physical Chemistry, Czech Academy of Sciences, Czech Republic

In living cells, many enzymes are able to interact with each other to form dynamic enzyme assemblies. In particular, this is the case for phosphofructokinases, which have been identified as prominent members of glycolytic enzyme assemblies and shown to form filaments. However, the mechanism of the formation of these structures and their physiological roles remain poorly understood. Here we use a combination of all-atom and coarse-grained molecular dynamics simulations to describe the assembly formation by phosphofructokinase enzymes. By obtaining contact maps between proteins, comparing their interaction networks, we characterize the role of specific amino acids and secondary structures in the mechanism of formation of these superstructures. The work provides a starting point for the elucidation of mechanisms governing the localized metabolism in cells.

TRANSIENT OLIGOMERIZATION AND AGGREGATION OF HUMAN LEPTIN

Grace E. Orellana[1], Jennifer M. Simien[1], Chad A. Brautigam[2], Yao Hu[3], Qian Wang[3] ,Lisa Lang[4], and Ellinor Haglund[1] - [1]Department of Chemistry, University of Hawai'i at Mānoa, Honolulu, USA; [2]Departments of Biophysics and Microbiology, UT Southwestern Medical Center, USA; [3]Department of Physics, University of Science and Technology of China, China; [4]Department of Biochemistry and Biophysics , Stockholm University, Sweden

Obesity is categorized as one the largest public health challenges in the world and the rapid emergence demonstrates that genetic components play a small role in the progression of this disease. Leptin is a pleotropic hormone critical for regulating energy homeostasis, satiety, and nutritional status linked to obesity. Despite its important role, the molecular details of leptin and the leptin receptor (LEP-R) complex remain elusive. In this work, we tested the hypothesized binding sites of leptin to the receptor focusing on binding site I, II, and III utilizing designed antagonist proteins. Our results support binding site I of leptin has a role in biological signaling and there is a possibility of the formation of a higher-order complex or an allosteric rearrangement of the human LEP-R complex. By using all-atomistic simulations, a predicted trimeric complex is unable to bind to LEP-R due to blockage of binding sites I and III. We investigate the formation of transiently formed leptin oligomers *in vitro* using analytical ultracentrifugation and utilize tryptophan fluorescence to characterize the mechanism of aggregation. Interestingly, irreversible aggregates form from wild-type leptin and the question then remains if transient oligomers of leptin are the potential seed for aggregation in diet-induced obesity.

Engineering of autocatalytic activity of HIV-1 protease

Klára Poštulková - Institute of Organic Chemistry and Biochemistry of the CAS

The expanding use of proteases in commercial processes, therapeutics and research has led to the need for different strategies to engineer them. Engineering targets the thermodynamic stability, solubility, catalytic efficiency or substrate specificity. New protease properties can be achieved by mutagenesis or by fusion with different protein domains.

One of the most studied viral proteases is the HIV-1 protease. This protease plays a crucial role in the viral life cycle by cleaving the Gag and Gag-Pol polyproteins to release the mature structural and functional proteins, including itself. The maturation of the protease is called autoprocessing and its mechanism is not yet fully understood. This process can be disrupted by a specific mutations in the sequence of HIV-1 protease and also by fusion with specific proteins at the N-terminus of the protease (Maltose Binding Protein, Thioredoxin).

In our design, we linked HIV-1 protease to Thioredoxin at the N-terminus. Thioredoxin is a commonly used solubilizing protein. Although an inhibitory effect of this protein on protease maturation has been previously described, in our design we could observe the presence of a mature protease after expression.

A more significant inhibition of autoprocessing occurred in the protease carrying the C95K mutation. The cysteine at position 95 is commonly mutated to alanine to increase the solubility of the protease. Mutation to lysine resulted in higher solubility of the protease but almost complete loss of its activity.

One of our next goals is to describe the effect of this position on autoprocessing in mutant variants of HIV-1 protease. Based on a comprehensive analysis of the HIV-1 protease sequence database (HIV Drug Resistance Database, Standford University), we designed a modified HIV-1 protease with more than 30 drug-resistant mutations together.

Sequence fingerprint of structured and disordered proteins

Patrícia Sotáková - Institute of Organic Chemistry and Biochemistry

Disordered proteins are a topic of growing interest. With ongoing research describing the relationship between sequence and structure, this work aims to investigate features in amino acid sequence that could indicate fingerprints of structured or disordered proteins. These fingerprints could deepen our understanding of disordered regions or protein folding. Furthermore, this knowledge could help design new deep learning predictors of protein disorder or protein domain recognition. We performed statistical analysis on sequences obtained from Protein Data Bank and DisProt database, including a comparison of protein sequences with artificial ones generated under the assumption of amino acid pairwise independence. Subsequently, we identified triples of two amino acids and their distance that are significantly different in occurrence to the artificial set. Based on this analysis, we sorted the triples into the following categories: abundant, standard, and depleted. Observed pairs with abnormal frequency in a given distance can be interpreted as a fingerprint of secondary structure, motif, domain, or other unknown identification of disordered proteins depending on the dataset. A simple example of a sequence fingerprint was observed in the PDB dataset. The abundance of histidines in distances 1-5 can be explained by histidine tagging used during protein purification.

Protein tunnels, channels and pores, ChannelsDB & MOLEonline

Anna Špačková[1], Václav Bazgier[1], Radka Svobodová[2], Lukáš Pravda[2], David Sehnal[2], Ondřej Vávra[2], David Bednář[2], Karel Berka[1] - Palacký University Olomouc[1], Masaryk University[2]

Channels in proteins plays significant role in developing new drugs, because of this is important to study these structures. MOLEonline (https://mole.upol.cz/) is available tool for discovering tunnels in proteins structure. Algorithm can find out tunnels, pores and channels on the protein surface. Obtained information can be stored in ChannelsDB database (https://channelsdb.ncbr.muni.cz/). Because of the results from algorithm don't say anything about biological importance, we would like to develop new tool which can recognize it. We are planning based this tool on artificial intelligence together with knowledge of biologically useful channels, and thus create new ontology. This improvement can help with docking molecules into buried active sites and overall in drug discovery.

Dynamical allostery of multi-domain protein complexes

Josef Šulc - Institute of Organic Chemistry and Biochemistry of the CAS

This study investigated the impact of connecting different domains in different contexts to a central binding protein. One of the goals was to determine how these modifications affect the ability of the protein to bind small peptide ligands. Specifically, we focused on the PDZ3 domain connected to TrpCage miniproteins at the N-terminus and C-terminus, using glycine linkers of various lengths.

To investigate this phenomenon, we used a combination of molecular dynamics simulations, enhanced sampling techniques, semi-empirical quantum mechanics, and statistical analysis techniques. We analyzed the molecular dynamics of the PDZ3 domain and its connections to the TrpCage miniproteins in different contexts and quantified the resulting changes in dynamical, conformational, and electrostatic behavior. In doing so, we have developed a novel methodology for analyzing the dynamic behavior of multi-domain proteins.

Our analysis revealed that modifying the central binding protein significantly altered its dynamic and conformational behavior, which in turn affected its ability to bind ligands. Specifically, we found that connecting the PDZ3 domain to the TrpCage miniproteins at different termini results in distinct changes in entropyleading to a significant impact on ligand binding. We also observed that one of the protein variants displays IDP-like behaviour.

Our study highlights the importance of considering the dynamic behavior of multi-domain proteins in the design of new proteins and therapeutic macromolecules. By providing a deeper understanding of the relationship between protein dynamics and function, our findings open up new avenues for designing and optimizing protein-protein interactions, with potential applications in drug development and other biotechnological areas. The novel methodology we developed for analyzing protein dynamics, conformations, and electrostatics could be widely applicable to studying other multi-domain proteins as well.

Analysis and Sampling of Molecular Simulations by adversarial Autoencoders

Guglielmo Tedeschi[1], Vladimír Višňovský[2], Aleš Křenek[2], Vojtech Spiwok[1] - [1]Department of Biochemistry and Microbiology - University of Chemistry and Technology, Prague. [2]Department of Machine Learning and Data Processing - Faculty of Informatics - Masaryk University, Brno.

This research is motivated by acceleration. Molecular simulations make true the possibility to simulate the motion from small molecules to big proteins and their combinations in drug-target complexes. It lets us predict their changing confirmation, their stability and a plenty of other properties thanks to the evolution of molecular structure. However, application of molecular simulations is affected by the large computational costs in computing steps that must be in order of femtoseconds, to assure numerical stability to integrate Newton equation of motion. Taking into account this limitation, a typical molecular dynamics simulation is capable of sampling only a small fraction of the states available to the simulated system, with the likely catch or unlikely loss of some slow or rarely occurring processes, where likelihood depends on the simulation time. There are numerous techniques to address this limitation and to speed up simulations. Metadynamics is an enhancing method based on biasing Hamiltonian of the system that helps to cross barriers and go head through new unexplored free energy surface areas, thanks to some selected internal coordinates, so called collective variables. Choosing correct collective variables to make metadynamics successful is not a trivial task and it depends first of all on the knowledge and expertise of the user. In the last few years there are emerging opportunities for machine learning and artificial neural networks in this field. We decided to develop an adversarial autoencoder[1] as a tool to analyze simulation data and to support users to derive good collective variables to enhance molecular dynamics simulation. The potential of this platform is demonstrated by using Trp-cage unbiased molecular dynamics simulation[2] and ongoing simulations for Villin headpiece by structures generated from Rosetta package[3].

We thank D.E Shaw Research to provide us trpcage trajectory for asmsa training. This work was supported by the Czech Science Foundation (22-29667S).

1. A. Makhzani, J.Shlens, N.Jaitly and I.Goodfellow, I. Adversarial autoencoders. International Conference on Learning Representations, 2016.

2. K. Lindorff-Larsen, S. Piana, R. Dror and D. Shaw, How fast-folding proteins fold. Science 334:517, 2011.

3. J. K. Leman, B. D. Weitzner, S. M. Lewis, J. Adolf-Bryfogle, N. Alam, R. F. Alford, M. Aprahamian, D. Baker, K. A. Barlow, P. Barth, et al. Macromolecular modeling and design in Rosetta: recent methodsand frameworks. Nat. Methods 2020.

Linking the Diffusive Dynamics of Bacterial Proteome to Cell Metabolism and Death

Štěpán Timr - J. Heyrovský Institute of Physical Chemistry, Czech Academy of Sciences

Temperature variations have a big impact on bacterial metabolism and death, yet an exhaustive molecular picture of these processes is still missing. Here, by combining multiscale molecular simulations of the E. coli cytoplasm with neutron scattering measurements, we clearly show that only a minor fraction of the proteome unfolds at the cell death. Moreover, we prove that the dynamical state of the E. coli proteome is an excellent proxy for temperature-dependent bacterial metabolism and death. The proteome diffusive dynamics peaks around the bacterial optimal growth temperature, and a dramatic dynamical slowdown starts just below the cell's death temperature. We show that this slowdown is caused by the unfolding of a small fraction of proteins, which establish an entangling interprotein network, dominated by hydrophobic interactions. Finally, the deduced progress of the proteome unfolding and its diffusive dynamics are both key to correctly reproduce the E. coli growth rate.

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

To mistranslate or not to live

Vjaceslav Tretjacenko - Weizmann Institute of Science

Traditional view on organismal adaptation postulates that mutations in genotype drive the phenotypic variation which allows to sample diverse adaptive strategies towards the environment. In our work we explore different modes of adaptation where protein phenotypic variations arise directly from the frequent ribosomal mistranslations. Combination of mistranslations upon the genes might increase the adaptive capacity of the organism by sampling multiple phenotypic variations at the same time. Additionally, error sampling of phenotypic variants in combination with genetic variation might explain the rise of complex epistatic molecular traits emergence of which would be exceptionally rare otherwise. Here we present the significance of mistranslations on organismal fate on the example of *E.coli* TEM-1 beta lactamase enzyme. Protein was engineered in such a way that only mistranslation will rescue organisms from presented antibiotics. We report that mistranslation frequency directly correlates with organismal fitness allowing us to speculate towards alternative evolutionary pathways supported by the nature of this observation.

Proteolytic profiles of human AMBN by MMP-20 and KLK-4 proteases

Veronika Vetýšková - Institute of Organic Chemistry and Biochemistry of the CAS

Ameloblastin (AMBN) is a protein that plays a crucial part in the formation of tooth enamel. However, recent research suggests that its role in the organism may not be limited to this function alone. The AMBN co-assemble with another protein, amelogenin (AMEL), to form the tooth enamel matrix. The AMBN protein naturally self-assemble to form oligomers via specific sequence in exon 5. Two proteases, enamelysin (MMP-20) and kallikrein 4 (KLK-4), are responsible for processing the enamel proteins, removing the AMEL and non-amelogenin proteins from scaffold to leave fully developed tooth enamel. The human AMBN exists in two different isoforms (ISO I and II), which likely have different functions based on their different proteolytic profiles. We compared the proteolytic profiles of the two AMBN isoforms after proteolysis by MMP-20, KLK-4 and a combination of both. The results that were analysed by mass spectrometry showed significant differences in the cleavage products. These results may help to understand different functions for each isoform in human signalling pathways.

Accurate modeling of free energy landscapes enables "Unstructure"-based drug discovery for IDPs

Bentley Wingert - New Equilibrium Biosciences

Intrinsically disordered proteins (IDPs) make up 30% of human proteins, but over two-thirds of those proteins are implicated in cancers. C-Myc is an example of a fully disordered protein which is implicated in 70% of cancers. The development of drugs targeting IDPs has thus far been extremely challenging for rational or structure-based drug discovery efforts due to their resistance to experimental methods such as X-ray crystallography. Thus, while molecules have been discovered that target Myc and other IDPs, there have been few successful efforts to optimize these drugs, and no drugs have made it to market. Molecular dynamics (MD) simulations enable the visualization of transient binding pockets on IDPs which support structure-based design. Recent developments in MD forcefields, through both extended training data sets and machine-learning, have improved their performance when simulating disordered proteins. New Equilibrium has developed a computational-experimental platform, IDP-SHIFT, that combines machine learning, MD simulations, and "unstructural" biology approaches to identify transient, druggable pockets on IDPs. Using New Equilibrium's high-performance computing cluster, we generated long time-scale MD simulations of the c-Myc transactivation domain (TAD) to determine its free energy landscape. We then identified microstates that were both energetically favorable and contained potential binding pockets and nominated these for virtual screening. Several published compounds have been shown to bind Myc TAD, and we validated our predicted binding pockets by showing that a large-scale virtual screen identified these published compounds in the top 2% of binders. To determine whether these poses enabled further optimization, we then designed compounds with improved docking scores across multiple nominated pockets. These optimized compounds improved in-cell selectivity for Myc, with no loss in potency. We show that through understanding of the free energy landscape of IDPs, it will be possible to effectively regulate the conformational ensemble members most relevant for functional activity.

List of participants

Marco Ancona
anconama@merck.com
MSD Czech Republic

Kathrin Bach
kathrin.bach@uochb.cas.cz
IOCB Prague

Brian Baker
brian-baker@nd.edu
University of Notre Dame

Kateřina Balážová Faltejsková
katerina.faltejskova@uochb.cas.cz
IOCB Prague

Krishnendu Bera
490519@mail.muni.cz
CEITEC MU

Jan Beránek
Jan8be@gmail.com
University of Chemistry
and Technology, Prague

Mateusz Biesaga
mbiesaga@nuagetx.com
Nuage Therapeutics

Kristýna Boušová
kristyna.bousova@uochb.cas.cz
IOCB Prague

Virginia Burger
virginia.burger@newequilibriumbio.com
New Equilibrium Biosciences

Jens Danielsson
jens.danielsson@dbb.su.se
Stockholm University

Ken Dill
dill@laufercenter.org
Stony Brook University

Zsuzsanna Dosztányi
zsuzsanna.dosztanyi@ttk.elte.hu
Eötvös Loránd University

Christos Feidakis
christos.feidakis@natur.cuni.cz
Charles University, Faculty of Science

Alan Fersht
arf25@cam.ac.uk
University of Cambridge

Erdős Gábor
gabor.erdos@ttk.elte.hu
MTA-ELTE Momentum Bioinformatics
Research Group

Valerio Guido Giacobelli
giacobev@natur.cuni.cz
Charles University

Roberta Gioia
roberta.gioia@sibyllabiotech.it
Sibilla Biotech

Ellinor Haglund
ellinorh@hawaii.edu
University of Hawaii at Manoa

Gabriella Heller
g.heller@ucl.ac.uk
University College London

Jan Heyda
jan.heyda@vscht.cz
University of Chemistry
and Technology, Prague

Klára Hlouchová
havova1@natur.cuni.cz
Charles University

Aleš Hnízda
ales.hnizda@lf1.cuni.cz
First Faculty of Medicine CUNI

Tomáš Honzík
thonzik@ntis.zcu.cz
University of West Bohemia

Kateřina Jirásková
katerina.jiraskova@uochb.cas.cz
IOCB Prague

Ilia Korobko
ilia.korobko@weizmann.ac.il
Weizmann Institute of Science

Lukáš Kuhajda
kuhajdal@ntis.zcu.cz
University of West Bohemia

Denise-Liu` Leone
denise.leone@uochb.cas.cz
IOCB Prague

George Makhatadze
makhag@rpi.edu
Rensselaer Polytechnic Institute

Varun Mandalaparthy
mandalaparthy@cpc.tu-darmstadt.de
TU Darmstadt

Martina Mičková
martina.mikulu@uochb.cas.cz
IOCB Prague

Tom Miclot
tom.miclot@jh-inst.cas.cz
J. Heyrovský Institute of Physical Chemistry, CAS

Mikael Oliveberg
mikael.oliveberg@dbb.su.se
Stockholm University

Grace Orellana
georella@hawaii.edu
University of Hawaii at Manoa

Gary Pielak
gary_pielak@unc.edu
University of North Carolina at Chapel Hill

Natália Pižemová
natalia.pizemova@uochb.cas.cz
IOCB Prague

Klára Poštulková
klara.postulkova@uochb.cas.cz
IOCB Prague

Paul Robustelli
paulrobustelli@gmail.com
Dartmouth College

Alma Carolina Sanchez Rocha
sanchezalm@natur.cuni.cz
Charles University

Benjamin Schuler
schuler@bioc.uzh.ch
University of Zurich

Joan-Emma Shea
shea@ucsb.edu
University of California, Santa Barbara

Patrícia Sotáková
patricia.sotakova@gmail.com
IOCB Prague

Anna Špačková
anna.spackova@upol.cz
Plackỳ University Olomouc

Anna Strachotová
anna.strachotova@uochb.cas.cz
IOCB Prague

Kvido Stříšovský
kvido.strisovsky@uochb.cas.cz
IOCB Prague

Josef Šulc
sulc@uochb.cas.cz
IOCB Prague

Kamil Tamiola
kamil@peptone.io
Peptone Ltd.

Guglielmo Tedeschi
tedeschg@vscht.cz
University of Chemistry and Technology, Prague

Štěpán Timr
stepan.timr@jh-inst.cas.cz
J. Heyrovsky Institute of Physical Chemistry, CAS

Silvio Tosatto
silvio.tosatto@unipd.it
University of Padova

Vjaceslav Tretjacenko
vjaceslav.tretjacenko@weizmann.ac.il
Weizmann Institute of Science

Konstantinos Tripsianes
kostas.tripsianes@ceitec.muni.cz
Masaryk University

Mihaly Varadi
mvaradi@ebi.ac.uk
EMBL-EBI

Michele Vendruscolo
mv245@cam.ac.uk
University of Cambridge

Veronika Vetýšková
veronika.vetyskova@uochb.cas.cz
IOCB Prague

Jiří Vondrášek
jiri.vondrasek@uochb.cas.cz
IOCB Prague

Wim Vranken
Wim.Vranken@vub.be
Vrije Universiteit Brussel

Bentley Wingert
bentley.wingert@newequilibriumbio.com
New Equilibrium Biosciences

Notes

Published by Institute of Organic Chemistry and Biochemistry of the Czech Academy of Sciences, ELIXIR Czech Republic and SCIENTIFIKA.

Editors: Jiří Vondrášek, Anna Strachotová, Natália Pižemová
Print: VENICE Praha s.r.o.
Year: 2023
Pages: 66