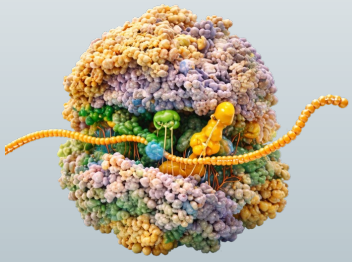
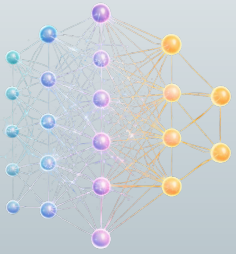
PRAGUE
PROTEIN
SPRING


Prague Protein Spring 2026

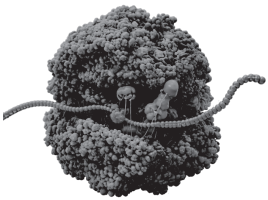
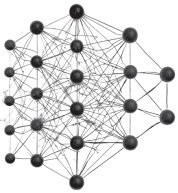
Protein Science in the Post-AlphaFold Era:
New Challenges – New Paradigms

7 - 10 May 2026, Prague, Czech Republic

Organized by

IOCB
PRAGUE

elixir
CZECH
REPUBLIC


Prague Protein Spring 2026

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 978-80-86241-73-9

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

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

© 2026 Institute of Organic Chemistry and Biochemistry of the Czech Academy of Sciences
© 2026 ELIXIR Czech Republic

ISBN 978-80-86241-73-9

Summary of content

Scientific Programme

Thursday 7 May 2026

Opening:	**Jan Konvalinka IOCB Prague director**
	Jiří Vondrášek IOCB Prague
	Mikael Oliveberg Stockholm University
	Michele Vendruscolo University of Cambridge

Opening Session: Where Do We Go from Here?
Chair: Mikael Oliveberg - Stockholm University, SE

Opening Keynote (19:00 - 20:00)

19:00 - 20:00	**David Baker - University of Washington, USA**
	Protein design using deep learning
20:00 - 20:45	**Michele Vendruscolo, University of Cambridge, UK**
	Drug design for undruggable targets using AI
20:45 - 22:30	Welcome Reception / Welcome Drink

Friday 8 May 2026

BLOCK 1: AlphaFold Independent – The Physics and Evolution of Protein Folding
Chair: Ellinor Haglund - University of Hawai'i of Mānoai, USA

08:30 - 09:15	**Eugene Shakhnovich - University of Harvard, USA**
	Protein evolution with and without chaperones
09:15 - 10:00	**Patricia Clark - University of Notre Dame, USA**
	Silent No More: Synonymous Substitutions Regulate Protein Folding and Gene Expression in E. coli
10:00 - 10:30	**Gabor Erdos - ELU Budapest, HU**
	Zero-Shot Prediction of Thermodynamic Properties of Proteins
10:30 - 11:00	Coffee break

11:00 - 11:45	**Stephen Fried - John Hopkins University, USA** Protein Folding in the AI Era: What's Left to Discover
11:45 - 12:25	**Ayala Shiber - Israel Institute of Technology, IL** Evolution of Protein Folding and Assembly Pathways: Deciphering the Dynamics of Divergent Co-translational Assembly Pathways, in Atomic Resolution
12:25 - 13:00	**Iva Pritisanac - Helmholtz Munich, Institute of Computational Biology, DE** A functional map of the human intrinsically disordered proteome
13:00 - 14:30	Lunch Break

BLOCK 2: After AlphaFold – What Machines Know and What They Don't
Chair: Michele Vendruscolo - University of Cambridge, UK

14:30 - 15:10	**Nikolay Dokholyan - University of Virginia, USA** Early Molecular Events in Neurodegeneration
15:10 - 15:50	**George Makhatadze - Rensealler Polytechnic Institute, USA** Electrostatics in the Age of AI Structure Prediction
15:50 - 16:20	**Alena Khmelinskaia - Ludwig-Maximilians-Universität München, DE** Function-driven design of protein assemblies
16:20 - 16:50	Coffee break
16:50 - 17:30	**Sameer Velankar - EMBL-EBI, UK** When Proteins Became Predictable: AlphaFold and the Transformation of Biology
17:30 - 18:10	**Gennady Verkhivker - Chapman University, USA** Predicting Allostery in the AI Era: Protein Language Models, Agents and Biophysics

Friday 8 May 2026

BLOCK 2: After AlphaFold – What Machines Know and What They Don't
Chair: Michele Vendruscolo - University of Cambridge, UK

18:10 - 18:40	**Eliška Fürst - Max Delbrück Center Berlin, DE** Protein structures in context with proteome-wide biophysics
18:40 - 20:00	Supper
19:00 - 23:00	Poster session

Saturday 9 May 2026

BLOCK 3: Designing Life – Proteins, Chemistry and New Functions
Chair: Agnes Toth-Petroczy - Max Planck Institute, DE

09:00 - 09:45	**Lukasz Joachimiak - University of Texas, USA** Decoding Structural Polymorphism of Tau Amyloids in Disease
09:45 - 10:15	**Ahrum Son - Scripps Research Institute, USA** Protein Structural Changes in Alzheimer's Disease: From Mouse Tissue to Human Blood
10:15 - 10:45	Coffee Break
10:45 - 11:45	**Basile Wicky - ETH Zürich, CH** Interfacing with biology using protein design
11:45 - 12:10	**Ellinor Haglund - University of Hawai'i of Mānoa, USA** Structural plasticity in chemokines
12:10 - 14:15	Lunch Break

BLOCK 4: Beyond the Structure – Disorder, Motion and the Full Picture
Chair: Mikael Oliveberg - Stockholm University, SE

14:15 - 14:55	**Agnes Toth-Petroczy - Max Planck Institute, Dresden, DE** Systems biology of disordered proteins and biomolecular condensates
14:55 - 15:30	**Tiffany J. Callahan - Manas AI, USA** From Protein Conformational Landscapes to Cellular States
15:30 - 16:00	**Tomas Pluskal - IOCB Prague, CZ** Decoding the chemical universe of plants
16:00 - 16:30	Coffee Break
16:30 - 17:15	**Andrea Soranno - Washington University in St. Louis, USA** Single-molecule spectroscopy of disordered regions and nucleic acids in bacterial transcription
17:15 - 18:00	**Sebastian Hiller - University Basel, CH** Integrative structural biology with solution NMR spectroscopy
18:00 - 18:30	Closing Remarks and Discussion
19:00 - 22:00	Conference dinner

Sunday 10 May 2026

Departures

David Baker	University of Washington, USA
Michele Vendruscolo	University of Cambridge, UK
Eugene Shakhnovich	University of Harvard, USA
Patricia Clark	University of Notre Dame, USA
Gabor Erdos	ELU Budapest, HU
Stephen Fried	John Hopkins University, USA
Ayala Shiber	Israel Institute of Technology, IL
Iva Pritisanac	Helmholtz Munich, Institute of Computational Biology, DE
Nikolay Dokholyan	University of Virginia, USA
George Makhatadze	Rensealler Polytechnic Institute, USA
Alena Khmelinskaia	Ludwig-Maximilians-Universität München, DE
Sameer Velankar	EMBL-EBI, UK
Gennady Verkhivker	Chapman University, USA
Eliška Fürst	Max Delbrück Center Berlin, DE
Lukasz Joachimiak	University of Texas, USA
Ahrum Son	Scripps Research Institute, USA
Basile Wicky	ETH Zürich, CH
Ellinor Haglund	University of Hawai'i of Mānoa, USA
Agnes Toth-Petroczy	Max Planck Institute, Dresden, DE
Tiffany J. Callahan	Manas AI, USA
Tomas Pluskal	IOCB Prague, CZ
Andrea Soranno	Washington University in St. Louis, USA
Sebastian Hiller	University Basel, CH

Lectures

Protein design using deep learning

David Baker - University of Washington, USA

Proteins mediate the critical processes of life and beautifully solve the challenges faced during the evolution of modern organisms. The focus of our lab is the design of a new generation of proteins that address current-day problems not faced during evolution. In contrast to traditional protein engineering efforts which modify naturally occurring proteins, we design new proteins from scratch to optimally solve these problems. We develop and use deep learning methods to design these new proteins with new functions, produce synthetic genes encoding the designs, and characterize them experimentally. In this talk, I will describe the design of proteins to address current challenges in health, technology, and sustainability.

Drug design for undruggable targets using AI

Michele Vendruscolo - University of Cambridge, UK

Protein-ligand interactions play central roles in biological processes and are of key importance in drug design. Deep learning approaches hold the promise of becoming cost-effective alternatives to high-throughput experimental methods for ligand identification. This is particularly the case for disordered proteins, where there are few experimental options for accurate binding measurements that can be scaled up to cover large chemical libraries. To predict the binding affinity between disordered proteins and small molecules, I will describe a deep learning framework based on the transformer architecture. To show the applicability of this approach, we explored the binding of small molecules to the Alzheimer's Aβ peptide, identifying compounds that delayed its aggregation. Overall, the results that we obtained illustrate the potential of deep learning methods in accurately predicting the interactions of small molecules with disordered proteins, thus uncovering molecular mechanisms and facilitating the initial steps in drug discovery for these otherwise largely undruggable targets.

Protein evolution with and without chaperones

Eugene Shakhnovich - University of Harvard, USA

I will present recent theoretical and experimental developments aimed at understanding the two-way link between protein biophysics and protein evolution. The novel multiscale models integrate the molecular effects of mutations on physical properties of proteins, most notably their stability, into physically intuitive yet detailed genotype-phenotype relationship (GPR) assumptions. One of the key predictions from biophysics-based evolutionary models is that maintaining protein stability represents a fundamental constraint on evolutionary dynamics in many scenarios such as viral escape and antibiotic resistance. Foldase chaperones can relax this constraint opening sequence space to exploration of more diverse evolutionary pathways.

Next, I will describe experimental efforts to test these predictions in evolutionary dynamics of antibiotic resistance under the stress of antibiotic Trimethoprim (TMP) which inhibits essential E coli enzyme DHFR. Comparing strains evolving under endogenous conditions and under overexpression of key chaperone GroEL we found that many resistance mutations under GroEL+ were upstream of the gene leading to overexpression of DHFR whose detrimental effects were buffered by GroEL. In agreement with theoretical predictions, many missense resistant variants under GroEL+ were indeed more MG-like than under endogenous GroEL. However, in variance with all theoretical predictions, one GroEL+ resistant variant M20I was more stable and bound TMP more strongly than WT. Careful analysis of the mechanism of resistance of this variant lead us to the discovery of a new mechanism of GroEL action whereby the chaperone recognizes enhanced dynamics of the protein.

Silent No More: Synonymous Substitutions Regulate Protein Folding and Gene Expression in E. coli

Patricia Clark - University of Notre Dame, USA

Historically, substitutions between synonymous codons were considered phenotypically silent because such substitutions do not alter the encoded amino acid residue. If true, the degeneracy of the genetic code would enable every protein sequence to be encoded by an astronomical number of potential mRNA sequences with no phenotypic consequences. However, research from a large number of groups over the past 10+ years has now convincingly shown that, rather than being silent, synonymous substitutions can significantly perturb essentially every component mechanism associated with gene expression, including transcriptional regulation, mRNA trafficking and half-life, translational efficiency, frameshifting, and co-translational folding. Effects on co-translational folding, for example, can alter the folded structure of the encoded protein, making it more prone to degradation. To date, most studies investigating the effects of synonymous codon substitutions have focused on effects on the gene bearing these substitutions, however there is some evidence for longer-range effects of synonymous substitutions on the expression of other, neighboring genes. For example, synonymous substitutions can alter transcriptional activity of cryptic promoters positioned within genes, which can regulate (up or down) the expression of a neighboring gene. Thus, rather than serving only to encode the amino acid sequence of a proteins, synonymous codon usage within an mRNA sequence is likely under selection to satisfy several regulatory constraints, including those internal and external to the encoded gene.

Zero-Shot Prediction of Thermodynamic Propertiesof Proteins

Gábor Erdős - Eötvös Loránd University, HU

The thermodynamic properties of proteins are fundamental to understanding their function, dysfunction, and evolution. However, experimental characterization of these properties, especially for intrinsically disordered proteins (IDPs) that exist as dynamic conformational ensembles, remains a significant challenge. Computational methods have emerged as a powerful alternative, yet they often require extensive training on protein-specific data, limiting their ability to generalize.

Here, we present a novel transformer based message parsing graph neural network (trMPNN) architecture for the zero-shot prediction of protein thermodynamic properties. By representing proteins as graphs our model learns the underlying physicochemical principles governing protein thermodynamics while retaining speed that allows for the analysis of complete proteomes. The network was trained by maximizing the probability of the native structure against a set of decoys, guided by the Boltzmann distribution, allowing it to learn a transferable energy function. This approach enables accurate predictions on proteins not seen during training, overcoming a major limitation of previous methods.

We demonstrate the power of our network highlighting two key areas. First, we show its ability to predict ensemble-averaged thermodynamic properties of IDPs, providing insights into their unique conformational landscapes. Second, we showcase its accuracy in predicting absolute protein stability (ΔG values), a critical factor in protein engineering and disease pathogenesis. Our model achieves state-of-the-art performance in both tasks, with predictions in excellent agreement with experimental data.

The zero-shot capability of our GNN opens up exciting avenues for high-throughput screening of protein stability, the design of novel proteins with desired thermodynamic properties, and a deeper understanding of the complex interplay between sequence, structure, and thermodynamics in the proteome.

Protein Folding in the AI Era: What's Left to Discover

Stephen Fried - John Hopkins University, USA

The success of AI-based structure prediction has raised a provocative question: has protein folding been solved? Our work argues that, in many ways, the opposite is true. While tools like AlphaFold excel at predicting final structures, they are blind to how proteins actually fold, misfold, and change inside cells – processes that remain central to biology and disease.

In the Fried lab, we develop high-throughput structural proteomics methods, which we have applied to study protein folding at the scale of entire proteomes. This shift in scale reveals behaviors that fundamentally challenge textbook views of folding. We uncovered cases in which a misfolded protein state is *more* stable than the native structure, and situations where molecular chaperones – long thought to universally assist folding – can instead obstruct it.

These phenomena could not have been inferred from AI models trained on existing structural data. Strikingly, we also found evidence that subtle forms of misfolding may accumulate with age, quietly contributing to cognitive decline and potentially to neurodegenerative disease.

Together, our work shows that AI has not closed the book on protein folding. Instead, it has opened the door to discovering new physical principles and new links between proteostasis and human aging – revealing a rich frontier that lies beyond predicted structures.

Evolution of Protein Folding and Assembly Pathways: Deciphering the Dynamics of Divergent Co-translational Assembly Pathways, in Atomic Resolution

Ayala Shiber - Israel Institute of Technology, IL

The ribosome serves as a central hub for cellular protein-protein interactions, coordinating the dynamic interplay of nascent polypeptide chains. To investigate the principles underlying co-translational folding and complex assembly, we integrate selective ribosome profiling, imaging, N-terminomics, and all-atom molecular dynamics simulations of various stages of co-translational assembly interactions. By examining conserved N-terminal acetyltransferases (NATs), we reveal distinct assembly pathways in which highly homologous subunits perform antagonistic roles. We identify a small set of residues as assembly "hotspots" that initiate interactions upon emergence from the ribosomal exit tunnel. These hotspots exhibit high binding energy, stabilizing the entire interface. Adjacent alpha-helices containing hotspots display marked thermolability, undergoing folding and unfolding in simulations-behaviour modulated by partner subunit binding to prevent misfolding. These domains display distinct co-translational folding landscapes, analysed by molecular dynamics. In vivo, mutations in these hotspots impair co-translational complex formation and promote aggregation. Evolutionary analysis further shows that disease-associated missense variants in NATs map to predicted hotspot clusters, disrupting their function. Extending our analysis to the entire proteome using AlphaFold-Multimer and MD-derived interface energy calculations, we observe consistent patterns. We show that the spatial distribution of interface energy is a key determinant and predictive feature of co-translational assembly as well as co-translational thermolability.

A functional map of the human intrinsically disordered proteome

Iva Pritisanac - Helmholtz Munich, Institute of Computational Biology, DE

Much of the human proteome lacks stable structure and consists of intrinsically disordered regions (IDRs). IDRs have key roles in cellular signaling, gene expression, and cellular organization, but their rapid sequence evolution has made them notoriously difficult to study using standard tools. Our work offers new insights into human disordered proteins by focusing on conserved bulk molecular features of the sequence. By mapping these features across thousands of human IDRs, we reveal which conserved aspects of disorder contribute to specific protein functions or interaction networks, and which are associated with disease-risk genes. This resource charts the hidden logic of information encoded in disorder, a long-standing frontier in proteome science.

Early Molecular Events in Neurodegeneration

Nikolay Dokholyan - University of Virginia, USA

The aggregation and misfolding of proteins such as Cu/Zn superoxide dismutase (SOD1) in amyotrophic lateral sclerosis (ALS) and amyloid beta (Aβ) peptide in Alzheimer's disease (AD) are central to their respective neurodegenerative mechanisms. In ALS, post-translational modifications like glutathionylation destabilize SOD1 dimers, leading to toxic oligomers recognized by the C4F6 antibody, with trimeric oligomers driving neuronal death, while larger aggregates may provide neuroprotection. In AD, we reveal that monosialotetrahexosylganglioside (GM1), a crucial lipid in neuronal membranes, triggers Aβ42 oligomerization, stabilizes toxic species via arginine-5 binding, and facilitates ion channel activity linked to toxicity. Notably, GM1 also interacts with apolipoprotein E (APOE), particularly the risk-associated APOE4 allele, altering its structure and enhancing its affinity for GM1 over cholesterol. This interaction promotes GM1 clustering in lipid rafts, influencing its transport and facilitating conditions conducive to Aβ aggregation. By linking APOE's regulatory role in GM1 dynamics to Aβ aggregation and exploring toxic protein oligomers in ALS, our findings illuminate shared pathways in neurodegenerative diseases and suggest novel therapeutic avenues.

Electrostatics in the Age of AI Structure Prediction

George Makhatadze - Rensealler Polytechnic Institute, USA

AI-based tools for protein structure prediction have transformed structural biology, yet their limitations remain incompletely understood. This study reveals a recurring tendency of modern AI structure predictors: their limited capacity to consistently enforce the physico-chemical rules governing the placement of ionizable residues. Although highly accurate for natural proteins, models such as AlphaFold2 frequently generate physically implausible structures for variant sequences, including the burial of charged residues in hydrophobic cores. These inconsistencies reveal that current AI models prioritize statistical patterns over the underlying thermodynamic principles of folding. By demonstrating that short molecular dynamics simulations reliably identify these inaccuracies, we provide a practical validation strategy that strengthens confidence in AI-generated structures and underscores the need to couple machine learning with physics-based refinement.

Function-driven design of protein assemblies

Alena Khmelinskaia - Ludwig-Maximilians-Universität München, DE

Recent computational methods have been developed for designing novel protein assemblies with atomic-level accuracy. Yet, when compared to their natural counterparts, the structural and functional space covered by
de novo designed assemblies remains limited. I will share with you our ongoing efforts in diversifying the structural repertoire of protein assemblies and developing strategies to dynamically control protein assembly state. First, I will describe our approaches to the diversify assembly geometries beyond simple polyhedral geometries, such as linked architectures assembled from rigid building blocks following quasi-equivalence principles. Then, I will present our generalizable interface-seeded design approach for the generation of environemnt responsive oligomers driven by ion-mediate, small molecule-dependent or phosphorylation-triggered protein-protein interfaces. By leveraging novel architectures and a diversity of endogenous and exogenous signals, we aim to generate orthogonal, programmable control elements for synthetic biology.

When Proteins Became Predictable: AlphaFold and the Transformation of Biology

Sameer Velankar - EMBL-EBI, UK

Advances in protein structure prediction marked a turning point for life science research. Access to AlphaFold was a transformative moment. This breakthrough did not emerge in isolation. These methods were built on years of effort to make 3D-structure and sequence data openly accessible through the Protein Data Bank (PDB) and UniProt, the sustained role of CASP as a community benchmarking framework, and a clearly defined scientific question: can we predict protein structure accurately from sequence? Together, open data, benchmarking, and a well-defined challenge created the right conditions for applying AI methods.

That transformation accelerated when Google DeepMind and EMBL-EBI collaborated to make millions of predicted structures openly accessible through the AlphaFold Database. This changed the scale at which structural knowledge can support discovery across the life sciences, including for proteins not amenable to traditional experimental structural studies and for global challenges such as neglected diseases, where structural data had previously been unavailable. Access to a predicted structure for a protein or complex is not enough. It's important to use it correctly, so education, guidance, interpretation, and learning resources are essential.

At the same time, advances in structure prediction and design, together with rapid progress in experimental techniques, have reshaped structural biology and structural bioinformatics. They increasingly support studies of protein dynamics, proteins in cellular contexts (visual proteomics), and intrinsic disorder, and have driven the development of scalable structure analysis tools such as Foldseek and large-scale annotation resources such as The Encyclopedia of Domains. Predicted structures must also be validated and enriched to make them useful for further impact. Proteomics approaches, such as cross-linking and residue-level functional annotations from the literature, provide important sources of validation and functional insight. I will discuss how we at PDBe are building a modular, community-driven ecosystem through the PDB, PDBe Knowledge Base, the AlphaFold Database, and 3D-Beacons.

Predicting Allostery in the AI Era: Protein Language Models, Agents and Biophysics

Gennady Verkhivker - Chapman University, USA

The transformative success of artificial intelligence (AI) in structural biology has culminated in the development of sophisticated co-folding models capable of predicting complex biomolecular assemblies. However, the "holy grail" of computational biology—the reliable prediction of allosteric regulation—remains elusive. In this talk, we present a comprehensive interrogation of the current AI landscape, specifically evaluating five state-of-the-art co-folding tools—AlphaFold3, Boltz-2, Protenix, Chai-1, and DynamicBind—across large-scale, rigorously stratified datasets of orthosteric and allosteric protein-ligand complexes. Our systematic analysis unmasks a striking "allosteric blind spot": while these models achieve near-experimental accuracy in modeling orthosteric binding modes, they exhibit a universal, architecture-independent collapse in predictive performance when challenged with allosteric sites. To decode the biophysical grammar underlying this failure, we introduce a dual Explainable AI (XAI) framework. This approach integrates hierarchically fine-tuned Protein Language Models (PLMs) and structure-based machine learning with physics-derived landscape frustration analysis. We demonstrate that the "detectability" of a binding site is intrinsically linked to its energetic landscape. Orthosteric sites typically reside in minimally frustrated, evolutionarily rigid basins that generate strong, unambiguous signals for AI models. Conversely, allosteric sites are characterized by persistent neutral frustration—a rugged energetic regime that facilitates the conformational plasticity and regulatory versatility required for signal transduction, but which effectively masks these sites from standard data-driven algorithms.

Looking toward autonomous discovery, we introduce AlloAgent, an AI-driven research agent designed to navigate the complexity of the allosteric energy landscape. By integrating a suite of 15 specialized computational tools and curated databases, AlloAgent automates the transition from static structural snapshots to mechanistic hypotheses, coordinating multi-step simulation workflows and mutation impact assessments. This work suggests that the next frontier of the "AI Era" lies in moving beyond purely statistical architectures toward physics-informed, agentic frameworks. By grounding AI in the fundamental principles of energy landscape theory, we can begin to democratize the discovery of cryptic regulatory pockets and accelerate the design of next-generation allosteric modulators for precision medicine.

Looking forward to the next frontier, we also discuss the evolution from "Protein-Only" models to a "Protein-Ligand Unified AI" that moves beyond binary classification toward a Trimodal Alignment of Sequence, Structure, and Chemistry. By utilizing the "ScaleAI" data strategy—incorporating massive synthetic and experimental datasets such as SAIR (5.2M structures) and PLINDER—we aim to map the "Action Space" of 100M+ small molecules onto the protein "Structure Space."

Protein structures in context with proteome-wide biophysics

Eliška Fürst - Max Delbrück Center Berlin, DE

Different environmental cues such as stress, nutrients or drugs, trigger rapid adaptive responses that allow to maintain cellular homeostasis. One of the fastest cellular responses to the environment is the binding of small molecules to proteins. These molecular interactions produce allosteric effects, which means that they trigger a variation of protein activity as a consequence of a structural conformational changes that occurs instantly. Allosteric interactions are thus essential for life and can modulate both the metabolic status of the cells and gene expression.

My group focuses on the development of new MS-bases proteomics technologies to study global structural changes in the proteome. In my talk I will describe the application of proteomic techniques to discover novel protein-small molecules interactions and its more recent applications for the study of the protein small-molecule interactome. I will talk about how we discovered new principles of small-molecule mediated allosteric regulation and chemical communication among cells, and how we plan to further expand this to study small molecule regulated networks in health and disease.

Decoding Structural Polymorphism of Tau Amyloids in Disease

Lukasz Joachimiak - University of Texas, USA

Tau, a microtubule-associated protein, aggregates into β-sheet-rich amyloids in over 25 neurodegenerative diseases collectively known as tauopathies. Recent breakthroughs in cryo-electron microscopy have uncovered remarkable structural diversity among tau fibrils in different diseases. Yet, a fundamental question remains: how does the intrinsically disordered tau protein adopt distinct amyloid folds across pathological states? In this talk, I will share findings from biochemical, structural, and computational studies that begin to unravel the energetic principles that underly the "folding logic" of tau amyloid formation. By dissecting these mechanisms, our goal is to inform the design of conformation-specific diagnostics and therapeutics, tools that could transform how we detect and treat tauopathies with molecular precision.

Protein Structural Changes in Alzheimer's Disease: From Mouse Tissue to Human Blood

Ahrum Son - Scripps Research Institute, USA

Protein misfolding and proteostasis dysfunction are hallmarks of Alzheimer's disease (AD), yet in vivo measurement of protein conformational changes has been technically challenging. Here, we introduce a novel whole-animal in vivo protein footprinting method using perfusion-based dimethyl labeling of exposed lysine residues, enabling structural profiling of intact proteins across multiple tissues. Applying this method to an AD mouse model, we identified proteins exhibiting AD-specific structural changes independent of aging, highlighting that conformational alterations precede changes in expression. As translational extension, we profiled structural changes in plasma proteins from 520 individuals across the AD spectrum—including normal, mild cognitive impairment (MCI), and AD patients—using mass spectrometry and machine learning. We identified a 3-protein panel capable of classifying disease stages with high accuracy (83.44%) and AUROC values exceeding 0.93 in pairwise comparisons. This panel also achieved 86.0% accuracy in longitudinal classification. Together, these studies demonstrate the power of structural proteomics in both animal models and human plasma for uncovering early and systemic biomarkers of AD. Our approaches provide a new framework for studying proteome-wide conformational dynamics in vivo and hold promise for improving early diagnosis and therapeutic monitoring of neurodegenerative diseases.

Interfacing with biology using protein design

Basile Wicky - ETH Zürich, CH

Recent advances in de novo protein design are transforming biology into an increasingly generative discipline. Fueled by progress in machine learning, we can now generate novel proteins in silico with increasing precision—expanding the design space beyond what evolution has explored. In the first part of my talk, I will describe how protein structure prediction networks, originally built for inference, can be re-purposed as generative models through strategies like hallucination.

Beyond generating isolated proteins, we are now exploring how to compose them into functional systems. I will share recent work where we use designed protein networks to implement Boolean logic directly in mammalian cells. Looking forward, we envision extending this approach toward in situ cellular classification at the molecular level—treating designed proteins as modular building blocks for biological computation. These developments point to the potential of generative protein models as a new way to interact with and program biological systems.

BIO: Basile Wicky is currently assistant professor for Biomolecular Design at the Department of Biosystems Science and Engineering, ETH Zurich, where his group works on developing and deploying methods in protein design for application in synthetic biology and cellular engineering. He was previously a postdoc with David Baker (University of Washington), and did his PhD in molecular biophysics at the University of Cambridge.

Structural plasticity in chemokines

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

Disulfide bonds stabilize protein structure, guide folding, enable redox regulation, and create opportunities for functional and structural plasticity. While native disulfide bonds define the canonical functional fold of proteins, non-native disulfides can stabilize alternative conformational states that may modulate, alter, or even impair biological activity. The regulatory and pathological consequences of these alternative disulfide configurations remain poorly understood. In this work, we use the chemokine Interleukine-8 (IL-8) as our model system containing four cysteines forming two disulfide bonds. Chemokines are small, disulfide-rich signaling proteins that regulate immune cell trafficking through receptor activation and gradient formation. Conserved disulfide bonds stabilize the chemokine fold, yet alternative disulfide connectivities can arise under oxidative stress. A combination of molecular dynamics (MD) simulations with in vitro and in cell biological assays was used to understand the role of native versus non-native disulfides in IL-8. Our results show that native disulfide connectivity encodes structural precision beyond thermodynamic stability, fine-tuning core packing and conformational integrity to support proper chemokine function. By defining how alternative disulfide patterns reshape structure without globally destabilizing the fold, this work provides new mechanistic insight into disulfide-driven functional plasticity under physiological and oxidative stress conditions.

Systems biology of disordered proteins and biomolecular condensates

Agnes Toth-Petroczy - Max Planck Institute, Dresden, DE

My lab is fascinated by how the protein diversity observed in nature is generated and distributed across protein families with focus on intrinsically disordered regions (IDRs). Recently, we have developed computational tools to enable comprehensive sequence-based analysis of the IDRome to identify evolutionarily-related and functionally analogous sets of IDRs (SHARK-dive, bio-shark.org), and uncover conserved sequence elements that may drive function (SHARK-capture). We have a special focus on proteins involved in biomolecular condensates. We made some initial progress by developing an unbiased classifier which predicts condensate forming proteins irrespective of their disorder content across organisms (PICNIC, picnic-bio.org). This was enabled by our curated condensate database, CD-CODE (cd-code.org). We also benchmarked variant effect predictors and found that they have lower sensitivity in disordered regions (Luppino et al, BMC Gen 2025), highlighting the difficulty of interpreting the functional effect of mutations in IDRs. Our projects aim to contribute towards the understanding of sequence-function relationships of proteins at proteome-wide scale.

From Protein Conformational Landscapes to Cellular States

Tiffany J. Callahan - Manas AI, USA

Understanding how molecular interactions give rise to cellular behavior and disease requires connecting molecular biophysics with systems-level representations, a gap that remains only partially addressed by current approaches. While larger datasets and more expressive models have improved predictive benchmarks, they have not closed the gap to mechanistic understanding. We begin at the molecular scale. Using molecular dynamics simulations of the Wnt-Wls system, we analyze conformational ensembles to identify the physical and geometric determinants of binding stability and specificity. An interpretable machine learning model trained on these features recovers known interaction mechanisms and highlights additional determinants, providing a generalizable vocabulary for protein–protein recognition beyond static, residue-level descriptions. We then connect this vocabulary to disease using knowledge graph–based representations that integrate molecular interaction features with heterogeneous biological data. These models link interaction-level properties to cellular processes and patient-level phenotypes, enabling structured representations of rare disease mechanisms. Rare Mendelian diseases, including Sickle Cell Disease, Cystic Fibrosis, and Phenylketonuria, illustrate distinct modes of molecular-to-system propagation: altered protein assembly, disrupted folding and trafficking, and loss of enzymatic function, respectively, each driving downstream cellular and organismal pathology. The Wnt-Wls system serves as a tractable model for extracting these determinants, while the knowledge graph provides a scaffold for connecting them across scales. Together, these components define a multi-scale approach for reasoning about how molecular perturbations propagate through biological systems. We present evidence from each scale and outline the challenges in building predictive, testable, and clinically relevant frameworks.

Decoding the chemical universe of plants

Tomas Pluskal - IOCB Prague, CZ

Although plants are an incredibly rich source of pharmaceutically relevant specialized metabolites, biosynthetic pathway elucidation in non-model plant species has proven challenging. Unlike bacteria and many fungi that contain biosynthetic operons, the genes of a given plant typically scatter randomly across the genome, making pathway discovery via genome mining nearly impossible. My lab is developing generalized workflows for connecting biosynthetic enzyme sequences (obtained using RNAseq) to their downstream metabolites (detected in LC-MS data). To this end, we have developed EnzymeExplorer, a targeted machine learning pipeline for predicting the enzymatic functions of terpene synthases directly from their amino acid sequences (Samusevich et al, bioRxiv 2025) and DreaMS, a self-supervised foundation machine learning model for tandem mass spectrometry, which outperforms state-of-the-art methods in a range of different prediction tasks (Bushuiev et al., Nature Biotechnology 2025). We are further building on these foundations towards the final goal: a full computational characterization of the chemodiversity and biosynthetic potential of each plant species using easy-to-obtain experimental datasets.

Single-molecule spectroscopy of disordered regions and nucleic acids in bacterial transcription

Andrea Soranno - Washington University in St. Louis, USA

Our lab uses single-molecule spectroscopy to study how intrinsically disordered regions (IDRs) regulate biological function and disease, from cardiac proteins to viral genome packaging and bacterial transcription. In this talk, I focus on sigma factors (σ), which enable RNA polymerase (RNAP) to initiate transcription. In Mycobacterium tuberculosis, the housekeeping σA contains a ~200-amino-acid N-terminal IDR with highly segregated charges. Removing this IDR strongly reduces transcription under multi-round conditions, suggesting a role in RNAP holoenzyme formation and recycling. Using fluorescence correlation spectroscopy (FCS) and single-molecule FRET, we find that the IDR enhances RNAP binding, adopts salt-dependent conformations, and remains highly dynamic in both free and bound states. Notably, it expands upon RNAP binding but re-compacts in the transcriptionally competent complex. Our results show that σA–RNAP–DNA assembly competes with non-specific DNA interactions and suggest that conserved IDR sequence features across Actinobacteria provide a general mechanism for modulating bacterial transcription.

Integrative structural biology with solution NMR spectroscopy

Sebastian Hiller - University Basel, CH

Structural biology is arguably at the height of its time. The integrated use of experimental and AI methods resolves problems at atomic level that have long been out of reach. Thereby, solution NMR spectroscopy is ideal to connect static structures towards their functional dynamics.

I will describe recent successes to employ solution NMR spectroscopy in such integrated setups, emphasizing the interplay between the different methods. Our journey starts with protein biogenesis in the endoplasmic reticulum (ER), where newly synthesized nascent chains are refolded by a network of molecular chaperones. We discover biomolecular condensates as the organizing principle of this chaperone network and report detailed functional and structural studies.

We then resolve the complete functional cycle of an ATP-driven molecular machine, the Hsp70 chaperone BiP, at atomic level. We create a non-equilibrium steady-state under turnover conditions inside the NMR to resolve that BiP undergoes a branched functional cycle that is regulated by two autoinhibition switches.

Finally, we leverage protein design to establish an experimental pipeline for high-throughput characterization of protein structure and dynamics by NMR. With this setup, a single operator can produce and analyze hundreds of proteins per week at minimal cost, unlocking a new regime of statistical structural biology, where sequence–structure–dynamics relationships are gained from experimental ensemble studies of suitably designed proteins.

Posters

List of posters

1. **Beyond Static Predictions: Evaluating Transient Interaction Landscapes of ATG16L1 in the Post-AlphaFold Era**
   Prokopis C. Andrikopoulos, Moushmi Goswami - First Faculty of Medicine, Charles University
2. **Conformational Ensembles and Early Intermolecular Contacts of Alpha-Synuclein and Familial Mutants**
   Adolfo Bastida-Pascual - Departamento de Química Física. Universidad de Murcia. Spain.
3. **AlphaFold structures show a subset of transport tunnels known from experimental structures**
   Karel Berka - Palacky University Olomouc
4. **5 years of Alphafoldology**
   Karel Berka - Palacky University Olomouc
5. **Comparative study of binding-induced changes across different PDZ domains**
   Borbála Jakab - Pázmány Péter Catholic University Faculty of Information Technology and Bionics
6. **Mapping the structural diversity in random protein sequence space**
   Filip Buchel - Faculty of Science, Charles University
7. **Disulfide topology governs biophysical behavior in chemokines**
   David Budean - University of Hawai'i of Mānoa
8. **No structure, no alignment, no problem: homology and conservation detection in intrinsically disordered and difficult-to- align regions**
   Chi Fung Willis Chow - Max Planck Institute of Molecular Cell Biology and Genetics (MPI- CBG)
9. **Structural basis of ATP-mediated inhibition of mycobacterial GMP reductase**
   Michal Doležal - Institute of Organic Chemistry and Biochemistry of the CAS
10. **SLiMshot: De Novo Design of Dual-Target Protein Binders for Short Linear Interacting Motifs**
    VU Duc Duy - Institut Pasteur, Computational Structural Biology Unit, Université Paris Cité

11. **MuSProt: a multimodal multistate protein dataset towards programmable protein design**
    Wenrui Fan - Centre for Machine Intelligence, University of Sheffield
12. **Phosphorylation Site Prediction Using Protein Language Models**
    Samuel Fančí - Charles University
13. **Addressing AI's Existential Crisis in the Protein Modelling Multiverse**
    Christos Feidakis - Institute of Organic Chemistry and Biochemistry of the CAS
14. **Ancient amino acid sets enable stable protein folds**
    Valerio Guido Giacobelli - Department of Cell Biology, Charles University
15. **ATLAS: Automated Topology Learning And Structure-solving workflow**
    Claudia Leticia Gomez Flores - Max-Planck-Institute for Solid State Research
16. **Multiscale Simulation of Phosphofructokinase-1 Assemblies: From Transient Interactions to Large-Scale Assembly Formation**
    Mehrnoosh Khodam Hazrati - J. Heyrovsky Institute of Physical Chemistry, Czech Academy of Sciences
17. **Improved Prediction of Single Amino Acid Variant Effects: Resolving Ambiguities and Expanding Applications**
    Tamas Hegedűs - Semmelweis University
18. **Effect of terminal modifications on aggregation properties and in vitro cytotoxicity of Staphylococcus aureus phenol soluble modulin β1 (PSMβ1)**
    Aleksandra Kalitnik - Department of Biomedical Engineering, Faculty of Fundamental Problems of Technology, Wrocław University of Science and Technology
19. **Vesiculin Reveals a Central Role of the IGF2 C-Domain in Receptor Specificity Across the Insulin–IGF System**
    Jakub Kaminsky - Institute of Organic Chemistry and Biochemistry of the CAS
20. **Structure-based metagenomic discovery and stabilization of functionally diverse enzymes for environmental sustainability**
    Olga Khersonsky - Weizmann Institute of Science

21. **Design of New Metal-Binding Peptides from Protein Fragments**
    Ján Michael Kormaník - Institute of Organic Chemistry and Biochemistry of the CAS
22. **Engineering Thermostable Proteins Using Large Language Models**
    Lukas Kuhajda - University of West Bohemia in Pilsen
23. **ECLIPSE: Exploring the functional dark proteomes of ESKAPE pathogens through sequence similarity connectomes of Protein Universe Atlas**
    Surabhi Lata - Department of Molecular Structural Biology, Helmholtz Centre for Infection Research (HZI)
24. **Structure-Based Pharmacophore Modeling of DprE1 and Protein– Protein Complex Modeling of DprE1–DprE2 for Anti-Tuberculosis Drug Discovery**
    Csongor Mátyás - Babeș-Bolyai University, Enzymology and Applied Biocatalysis, Faculty of Chemistry and Chemical Engineering
25. **Native-State Heterogeneity in Leptin Reveals Hidden Structural and Oligomeric States**
    Grace Orellana - University of Hawai'i of Mānoa
26. **Exploring the physiological amylome and the molecular mechanisms of its regulation**
    Dorota Maria Pfizenmaier - Institute for Biochemistry, Department of Biology, ETH Zürich
27. **Ovo, an Open-Source Ecosystem for De Novo Protein Design**
    David Příhoda - MSD Czech Republic
28. **Advancing Protein Engineering: Reliable Functional Evaluation and Selective Binder Activation**
    Jakub Ptacek - Institute of Organic Chemistry and Biochemistry of the CAS
29. **Design of proteins by temperting of message passing graph neural network**
    Vojtěch Spiwok - University of Chemistry and Technology, Prague
30. **Large-scale experimental exploration of protein space**
    Elsa Stern - University of Basel
31. **Short Peptides as Model Systems for Mapping Sequence– Thermodynamics Relationships**
    Josef Šulc - Institute of Organic Chemistry and Biochemistry of the CAS

32. **Drosophila as a Model for Structural and Functional Conservation of Leptin Signaling**
    Amanda Taitano - University of Hawai'i of Mānoa
33. **CVFormer: Data-Driven Collective Variables via Transformer Autoencoders**
    Guglielmo Tedeschi - Department of Biochemistry and Microbiology, University of Chemistry and Technology, Prague
34. **Encoded and non-genetic alternative protein variants expand human functional proteome**
    Vjaceslav Tretjacenko - Weizmann Institute
35. **Discovering Distal Mutations that Shape Enzyme Dynamics Using Ultra-High-Throughput Droplet Screening**
    Pavel Vanacek - FNUSA ICRC
36. **ML-Based Conformational Sampling and Ranking Do Not Improve Rigid Protein–Protein Docking**
    Semen Yesylevskyy - Institute of Organic Chemistry and Biochemistry of the CAS, Receptor.AI Inc.

Beyond Static Predictions: Evaluating Transient Interaction Landscapes of ATG16L1 in the Post-AlphaFold Era

Prokopis C. Andrikopoulos, Moushmi Goswami, Jiří Zahradník - First Faculty of Medicine, Charles University, BIOCEV

Recent advances in deep-learning–based protein structure prediction have revolutionised structural biology, enabling rapid modelling of protein complexes that previously lacked experimental structural information. However, the reliability of such predictions remains challenging for transient and weak protein-protein interactions, especially in cases where structurally resolved complexes are absent from training datasets. Here, we investigate the interaction landscape of the ATG16L1 WD40 domain, a key hub in non-canonical autophagy signalling, by systematically evaluating its interactions with multiple experimentally reported binding partners. These include V-ATPase subunit H, IL10RB, IL2RG, TMEM59, NLRP2, MDA5 and A20 among many others. Complex predictions were generated using AlphaFold2, AlphaFold3 and Boltz-2 for cross-method comparison of predicted interfaces and docking orientations. While several interactors converged on a common WD40 interaction surface, significant variability in docking geometry and confidence scores was observed across prediction frameworks, therefore showing the challenge of modelling interactions that are supported by biochemical or mutational evidence only, rather than structural data. In addition, these WD40 interactions are experimentally known to be transient and motif-mediated, with limited representation of corresponding complexes in the Protein Data Bank. This scarcity of training set introduces uncertainty in deep-learning predictions and can lead to apparently plausible but biologically ambiguous models. To address these limitations, we extend our analysis beyond structural predictions by performing molecular dynamics simulations and binding-energy evaluations to assess the stability and energetic plausibility of predicted interaction modes. Together, this work illustrates both the power and the limitations of current structure prediction tools in probing interaction networks governed by flexible peptide motifs. Our results emphasize the importance of integrating deep-learning predictions with energetic and dynamic analyses to interpret protein interaction mechanisms in the emerging post-AlphaFold era.

Conformational Ensembles and Early Intermolecular Contacts of Alpha-Synuclein and Familial Mutants

Adolfo Bastida-Pascual - Departamento de Química Física, Universidad de Murcia, Spain; J. Zúñiga - Departamento de Química Física, Universidad de Murcia, Spain; B. Miguel - Departamento de Ingeniería Química y Ambiental, Universidad Politécnica de Cartagena, Spain; M.A. Soler - Dipartimento di Scienze Molecolari e Nanosistemi, Universita' Ca' Foscari Venezia, Italy; F. Fogolari - Dipartimento di Scienze Matematiche, Informatiche e Fisiche, Universita' di Udine, Italy; S. Fortuna - Dipartimento di Scienze Matematiche, Informatiche e Fisiche, Universita' di Udine, Italy.

Abstrakt:Alpha-synuclein is a 140-residue intrinsically disordered protein whose aggregation into amyloid fibrils is associated with Parkinson's disease and other synucleinopathies. Small oligomeric species are considered key neurotoxic intermediates in this process. Several familial single-residue mutations (A30P, E46K, H50Q, G51D, A53E and A53T) are known to accelerate disease onset, suggesting that subtle changes in the conformational ensemble may modulate early aggregation events. In this work we generated conformational ensembles of alpha-synuclein and its mutants using the Probabilistic Molecular Dynamics Chain Growth (PMD-CG) method [1]. This approach combines Flexible-Meccano [2] and Hierarchical Chain Growth [3] with statistical information derived from tripeptide molecular dynamics trajectories. An important advantage of PMD-CG in mutagenesis studies is its computational efficiency: once the ensemble of the wild-type protein is constructed, generating ensembles for mutants only requires additional simulations of a small set of tripeptides. The resulting ensembles were validated by comparison with experimental NMR J-couplings, chemical shifts, and small-angle X-ray scattering (SAXS) data, showing improved agreement with experiments compared with direct MD simulations. To explore the earliest stages of aggregation, pairs of alpha-synuclein molecules (WT and mutants) were analyzed using protein–protein docking. The results reveal significant mutation-dependent changes in intermolecular contact patterns, particularly in regions involved in aggregation, suggesting possible mechanisms by which single-residue mutations modulate the nucleation process. These results highlight the importance of ensemble-based approaches for intrinsically disordered proteins in the post-AlphaFold era of protein science. [1] Bastida et al., PCCP 26, 23213 (2024) [2] Bernadó et al., JACS, 127, 17968–17969 (2005) [3] Pietrek et al., J. Chem. Theory Comput., 16, 725–737 (2020)

AlphaFold structures show a subset of transport tunnels known from experimental structures

Anna Špačková[1,2], Václav Bazgier[1], Karel Berka[1]
1 Department of Physical Chemistry, Faculty of Science, Palacký University, Olomouc, Czech Republic; 2 IT4Innovations, VSB – Technical University of Ostrava, Ostrava-Poruba, Czech Republic

Transport tunnels connect buried active sites with the surrounding solvent and control substrate access in numerous enzymes. Although both experimentally determined structures and artificial intelligence-folded models are widely available, the conservation of transport pathways between these representations remains unclear. Tunnels in experimental structures and artificial intelligence–folded models were compared within groups defined by UniProt accession. Tunnels were matched based on residue-level composition and overlap-based similarity metrics. The comparison revealed that tunnels identified in artificial intelligence-folded structures frequently correspond to dominant pathways observed in experimental structures. In contrast, experimental structures often contain additional alternative tunnels not captured in predicted models. This directional asymmetry indicates that artificial intelligence-folded structures represent conserved primary access routes, while experimental structures reveal greater variability. Analysis of cytochrome P450 enzymes revealed strong conservation of dominant tunnels across both structural representations. Artificial intelligence–folded structures capture experimentally supported access pathways, whereas experimental structures reveal additional alternative transport routes.

5 years of Alphafoldology

Karel Berka[1], Marián Novotný[2]

1 Dpt Physical Chemistry, Faculty of Science, Palacký University Olomouc; 2 Dpt Cell Biology, Faculty of Science, Charles University Prague

AlphaFold 2 was first presented at CASP 14 in December 2020.1 It was a major success in protein structure prediction, leading to a Nobel Prize in Chemistry last year. Especially since AlphaFoldDB and AlphaFold 2 source code opened in July 2021, the field of structural biology has moved towards the full structural description of all possible proteins with known and, nowadays even of unknown sequences. The speed and breadth of post-AlphaFold tools and services development, for which we coined the term „Alphafoldology" in September 2021, has been breathtaking since then. In the lecture, I will review the current state-of-the-art of the Alphafoldology field and what usage it enables for biology and chemistry.

This work was supported by ELIXIR CZ (reg. No. LM2023055).

1 Jumper, J., Evans, R., Pritzel, A. et al. Highly accurate protein structure prediction with AlphaFold. Nature 2020, 596, 583–589. https://doi.org/10.1038/s41586-021-03819-2

Comparative study of binding-induced changes across different PDZ domains

Borbála Jakab, Mounaf Al Makhlouf, Zoltán Gáspári - Pázmány Péter Catholic University Faculty of Information Technology and Bionics

One of the most prevalent components of human proteins is the PDZ domain. The 80–110 residues that make up PDZ domains have a very similar fold and function because they usually identify partner protein C-terminal regions. Because they mediate protein-protein interactions and are crucial for cellular signaling, these domains are crucial for controlling cellular activity. Despite extensive research on a number of well-known PDZ domains, a comprehensive examination of the dynamic features of their ligand binding is lacking. We investigate the motions of PDZ domains by conducting a thorough comparison analysis of free and bound domain structures using PCA techniques in order to find similar dynamics among different domains. This study was carried out by gradually increasing the diversity of the structures analyzed. We have identified distinct domain motions, and found that a number of PDZ domains may respond to ligand binding with opposing motions, indicating the existence of a distinct and complex dynamic framework for the PDZ domain family. With these findings, we intend to provide a potential universal model for bound/unbound transitions across PDZ domains.

Mapping the structural diversity in random protein sequence space

Filip Buchel[1], Tereza Neuwirthova[1]; Theodora Tureckiova[1]; Gustavo Fuertes[2]; Ales Benda[5]; Dalibor Panek[5]; Mohammed AlQuraishi[3]; Klara Hlouchova[1,4]
1 Faculty of Science, Charles University; 2 Institute of Biotechnology of the Czech Academy of Sciences; 3 Department of Systems Biology, Columbia University; 4 Institute of Organic Chemistry and Biochemistry of the Czech Academy of Sciences; 5 Imaging Methods Core Facility at BIOCEV

The universe of possible protein sequences is astronomically large, yet our understanding of the sequence-structure relationship is confined to the fraction currently used by life. Systematic mapping of the space surrounding the clusters of natural sequences is crucial for understanding protein evolution and designing novel proteins. Here, we illustrate the structural landscape of a random sequence library by screening one million variants using an in vivo proximity biosensor. We reveal that this space is structurally heterogeneous, populated by disordered chains and stress-inducing aggregates, along with "benign" compact sequences that resemble globular proteins and evade cellular chaperone responses. By training a neural network on these phenotypes, we show that structural potential is learnable and generalizes to natural proteomes. These findings demonstrate that biology-like topologies are accessible from random sequences with surprising frequency, providing data required to expand generative protein design beyond evolutionary priors.

Disulfide topology governs biophysical behavior in chemokines

David Budean[1], Dave Jameson[1], Paulo Henrique[2], Ronaldo Oliveira[2]
1 University of Hawaii; 2 Universidade Federal do Triângulo Mineiro

Chemokines are small signaling proteins that guide white blood cells (leukocytes) to sites of inflammation. These ligands are essential to immune function, activating intracellular signaling pathways through interactions with cell surface G protein–coupled receptors. A defining structural feature of chemokines is the presence of two disulfide bonds: one forms a closed loop, while the other threads through this loop, generating a pierced-lasso topology (PLT). This topology is central to chemokine structure and function. However, it remains unclear whether the native PLT represents a strictly defined conformation stabilized by thermodynamic preference, or whether alternative disulfide-bonded isoforms open up the possibility for structural heterogeneity within the native ensemble. To address this, we investigate nature's selectivity for the native topology using a combination of spectroscopic techniques and molecular dynamics simulations. Our results suggest that PLT formation primarily influences folding kinetics rather than thermodynamic stability. Furthermore, non-native disulfide formation exposes residues within the hydrophobic core, reducing packing efficiency and destabilizing the folded structure.

No structure, no alignment, no problem: homology and conservation detection in intrinsically disordered and difficult-to-align regions

Chi Fung Willis Chow[1], Swantje Lenz[1,2], Anna Hadarovich[1,2], Soumyadeep Ghosh[1,2], Maxim Scheremetjew[1,2], Soumyadeep Ghosh[1,2], Doris Richter[1], Ceciel Jegers[3,4], Alexander von Appen[1], Simon Alberti[3,4], Agnes Toth-Petroczy[1,2]

1 Max Planck Institute of Molecular Cell Biology and Genetics, Dresden; 2 Center for Systems Biology Dresden; 3 Cluster of Excellence Physics of Life, TU Dresden; 4 Biotechnology Center, Center for Molecular and Cellular Bioengineering, TU Dresden

Intrinsically disordered regions (IDRs) are highly flexible regions in proteins lacking a stable 3D structure and thus are ineffectively modelled by AlphaFold. IDRs are ubiquitous in the protein universe, and are critical for protein function by harboring key regulatory motifs and mediating protein-protein interactions. Nonetheless, understanding of IDR sequence-function relationships remains poor. A key bottleneck to effective IDR sequence-function analysis emerges from the rapid evolution of IDRs. Since they are not under the constraints of preserving structure, they sample sequence space at a higher rate, thereby leading to reduced sequence identity between homologs. Compounded with potential ex nihilo evolution of short sequence motifs, this renders alignment ineffective in identifying homologs and limits the efficacy of alignment-based tools in detecting conserved sequence features within. To tackle these issues, we developed Similarity/Homology Assessment by Relating K-mers (SHARK), an alignment-free sequence comparison algorithm which improves on existing alignment-free word-based algorithms by incorporating amino acid physicochemical similarities on a continuous scale. The core SHARK algorithm is further developed into the homology detector SHARK-dive, and the motif detector SHARK-capture. SHARK-dive is a machine-learned homology classifier aimed specifically for IDRs and other difficult-to-align sequences. Importantly, it was trained on a set of unalignable orthologous sequences using a curation strategy which aims to minimize reliance on alignments. SHARK-dive outperformed the widely used local alignment-based homology search tools BLAST and HMMER in a systematic benchmark, was able to distinguish between functionally homologous and unrelated IDRs reported in the literature, and can be used to identify potentially interesting IDR homologs. SHARK-capture is a motif detection tool for IDRs which offers consistently strong performance across both site and residues levels. Using SHARK-capture, we identified a highly conserved RDYR motif in the C-terminus of ATPase helicase Ded1p orthologs which was subsequently shown to promote ATPase activity, demonstrating its in providing confident experimental hypotheses of functional IDR regions. The development of these tools represents an initial step in facilitating systematic IDR sequence-function relationship investigation, thereby enabling the development of an IDR-specific homology resource and a global investigation of the IDR sequence landscape.

Structural basis of ATP-mediated inhibition of mycobacterial GMP reductase

Michal Doležal, Zdeněk Knejzlík, Tomáš Kouba, Anatolij Filimoněnko, Hana Šváchová, Martin Klíma, Iva Pichová - IOCB Prague

GMP reductase (GMPR) catalyzes NADPH-dependent conversion of GMP to IMP, a key metabolite in the biosynthesis of all purine nucleotides. This reaction allows mycobacteria and most other organisms to utilize guanine nucleotides for the production of adenine nucleotides without the need for de novo synthesis. In our studies of purine metabolism in mycobacteria, we use Mycobacterium smegmatis (Msm) as a model for the infectious Mycobacterium tuberculosis (Mtb), the causative agent of tuberculosis in humans. In a previously published study1, we demonstrated that the enzymatic activity of MsmGMPR is allosterically regulated by ATP and GTP. While ATP inhibits the enzymatic activity of MsmGMPR, GTP counteracts this inhibition, and thus restores the enzymatic activity. Here, we combine X-ray crystallography, cryo-electron microscopy, and biochemical binding assays to elucidate the molecular basis of MsmGMPR regulation by ATP and GTP2. MsmGMPR forms tetramers with four-fold axis which further assemble into octamers with D4 symmetry. The two tetramers in the octamer adopt either compressed or extended conformation. ATP and GTP compete for a binding site located at the interface of the two tetramers. We show that ATP stabilizes a compressed conformation that inhibits the enzyme by restricting access to the active site and preventing NADPH binding. In contrast, GTP counteracts ATP binding, promoting an active conformation that enables catalysis. Our results provide insight into how MsmGMPR senses and responds to the cellular purine nucleotide balance, revealing a novel mode of allosteric regulation by a CBS domain

SLiMshot: De Novo Design of Dual-Target Protein Binders for Short Linear Interacting Motifs

VU Duc Duy[1], Orso Subrini[2,3], Nicolas Wolff[2], and Massimiliano Bonomi[1]

1 Institut Pasteur, Computational Structural Biology Unit, Université Paris Cité; 2 Institut Pasteur, Signaling and Receptor Dynamics Unit, Université Paris Cité; 3 CortecNet, Les Ulis

Short linear interacting motifs (SLiMs) are conserved amino-acid segments in intrinsically disordered regions that mediate protein–protein interactions. The structural organization of SLiMs can be divided into three regions: the core, the N-terminal and C-terminal flanking regions (also referred to as the N- and C-context), whose variation provides binding context and determines interaction specificity. Designing proteins that interact with SLiMs with defined specificity holds great potential to probe the functional roles of the SLiMs as well as enabling new therapeutic strategies. Here, we introduce SLiMshot, a pipeline for de novo design of protein binders capable of recognizing two SLiMs. The dual-SLiM protein binder is generated by co-designing both the core-SLiM and the context binding sites using a two-track diffusion in both structure and sequence space, with sequences from the two tracks mixed at each diffusion step. Designs are subsequently validated and filtered using AlphaFold2 and PyRosetta, resulting in an end-to-end, user-friendly design workflow. As a proof of concept, we designed binders targeting the BH3 motifs of PUMA and BIM protein and validated them experimentally. Of 15 designs, all highly expressed, 12 designs bind at least one SLiM, and 7 designs bind both motifs, with the strongest binder having single-digit nanomolar affinities. SLiMshot thus enables design of binders with defined specificity, pushing further the potential application of protein design.

MuSProt: a multimodal multistate protein dataset towards programmable protein design

Wenrui Fan, Tianyi Jiang, Zheqing Zhu, Richard Mead, Guillame Hautbergue, Wei Sang, Haiping Lu - Centre for Machine Intelligence, University of Sheffield

Multistate proteins perform many biological functions by adopting multiple conformations and transitioning between them, making them attractive targets for programmable protein design. However, progress in multistate generative modeling is limited by the lack of large-scale datasets. Here, we present MuSProt, a multimodal multistate protein dataset that converts dispersed single-state protein snapshots into an AI-ready resource for multistate protein modeling. MuSProt represents protein dynamics in an epitomic form as (source state, transition, target state) triplets, integrates multimodal information from resources including the Protein Data Bank (PDB) and UniProt, refines coarse annotations into state-specific functions, and associates transitions with labels that summarize their physical plausibility. The dataset is organized as a hierarchical knowledge graph and maintained by an autonomous update pipeline with version control and update logs. MuSProt can be accessed through https://omaib.github.io/datasets/protein.

Phosphorylation Site Prediction Using Protein Language Models

Samuel Fanči, David Hoksza, Marian Novotný - Charles University

Phosphorylation is one of the most widely studied types of post-translational modifications (PTMs). It occurs mainly at three residues: serine (S), threonine (T) and tyrosine (Y). Detection of phosphorylation sites is useful for many downstream tasks in cellular biology and medicine. However, current experimental detection methods, such as mass spectrometry or immunoprecipitation, are often expensive, both in terms of cost and time. Established phosphorylation prediction tools use mostly window-based methods, which may be unable to capture long-distance dependencies between residues. Our goal was to leverage information-rich embeddings from state-of-the-art protein language models (PLMs) to create a more robust phosphorylation prediction method. Our models were trained on two main datasets, using data gathered from PhosphoSitePlus and dbPTM. To obtain a robust performance estimate, we clustered the sequential data and removed sequences above 30% similarity, and trained the models using 5-fold cross-validation. We performed a train-test sequence similarity analysis on datasets of previous site prediction models and found signs of potential data leakage. Our models were trained to predict sites for individual residues (S, T, Y) and select residue combinations (ST, STY). We used base PLM model fine-tuning combined with sequence perturbation to create our classifiers. The perturbation part of the model dynamically swaps residues in the input sequence, either randomly or according to a substitution matrix. We compared our results to SOTA models UniPTM and DeepPSP and observed an average relative increase in performance of 4.4%. On the dbPTM dataset, we achieved a test AUPRC score of 0.52, 0.42 and 0.35 for the S, T and Y models respectively. The combined residue models had a AUPRC of 0.48 for the ST and 0.45 for the STY model. The single-residue models outperformed combined models on their respective residues.

Addressing AI's Existential Crisis in the Protein Modelling Multiverse

Christos Feidakis[1], Radoslav Krivak[1], Jiri Vondrasek[1], Marian Novotny[2], Derrick Agwora[2], Chloé Weiler[3], David Hoksza[4]

1 Institute of Organic Chemistry and Biochemistry of the CAS; 2 Department of Cell Biology, Faculty of Science, Charles University; 3 Faculty of Engineering Sciences, Heidelberg University; 4 Department of Software Engineering, Faculty of Mathematics and Physics, Charles University

Thanks to advances in predictive modelling, most proteins in UniProtKB1 now have a predicted structure in the AlphaFold Protein Structure Database2 (AF-DB). Structural data can capture the conformational diversity a protein can exhibit in response to physiological perturbations, like protein-ligand interactions. A single model can therefore reflect any one of these native states – or perhaps none – but not all of them. A question that emerges is "which native state does a protein model represent?" Here, we classify a selection of predicted structures from AF-DB, based on their structural similarity to two categories of native states with regard to ligand binding: the ligand-bound (holo) and the ligand-unbound state (apo). Using experimentally determined apo and holo states of biologically relevant small-ligand binding sites from the Protein Data Bank3, we map the conformational diversity of these regions4 and classify AF-DB models as apo or holo according to their relative "proximity" toward each category. 1. UniProt Consortium (2021). UniProt: the universal protein knowledgebase in 2021. Nucleic Acids Res. 49, D480–D489. 2. Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G., Yuan, D., Stroe, O., Wood, G., Laydon, A., et al. (2022). AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. Nucleic Acids Res. 50, D439–D444. 3. Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E. (2000). The Protein Data Bank. Nucleic Acids Res. 28, 235–242. 4. Feidakis, C.P., Krivak, R., Hoksza, D., and Novotny, M. (2024). AHoJ-DB: A PDB-wide Assignment of apo & holo Relationships Based on Individual Protein–Ligand Interactions. J. Mol. Biol. 436, 168545.

Ancient amino acid sets enable stable protein folds

Valerio Guido Giacobelli[1], S. Andersson[2], P. Srb[3], T. Neuwirthová[1], Z. Ruszová[1], J. Marhoul[1], S. Pšenička[1], A. Knetl[1], L. Bednárová[3], V. Veverka[1,3], K. Hlouchová[1,3], I. André[2]
1 Department of Cell Biology, Charles University; 2 Department of Biochemistry and Structural Biology, Lund University; 3 Institute of Organic Chemistry and Biochemistry of the CAS

Early proteins likely arose from a chemically limited set of amino acids available through prebiotic chemistry, raising a central question in molecular evolution: could such primitive compositions yield stable, functional folds? Using de novo design, we constructed three ancient protein architectures using a reduced, evolution-inspired alphabet of ten amino acids, e.g. lacking all basic and aromatic residues. The resulting structures adopted their intended topologies and showed exceptional resistance to thermal and chemical denaturation. Computational simulations further revealed that proteins built from this restricted alphabet were as mutation-resilient as those using all twenty canonical residues. Besides their evolutionary implications, our results provide a foundation for minimalist protein design and generation of simplified, robust systems in minimal cell engineering.

ATLAS: Automated Topology Learning And Structure-solving workflow

Claudia Leticia Gomez Flores[1], R. Forth[1], S.Huang[1], M. Wadsack[1], O. Durif[1], J. Ham[1], D. Kumar-Gopalakrishnan[1], K. Kern[1,2], K. Anggara[1]

1 Max-Planck-Institute for Solid State Research; 2 Institut de Physique, École Polytechnique Fédérale de Lausanne

Post-translational modifications (PTMs) expand protein function beyond what primary structure (sequence) alone encodes. Glycosylation is a case where these modifications are especially consequential: the sugar code carried by glycoproteins is invisible to sequence-based prediction methods, yet directly controls molecular recognition, signalling, and self-assembly. Resolving how glycan decorations shape protein structure and function remains an open problem in the post-AlphaFold era. Glycoproteins and other glycoconjugates are flexible polymers that populate broad conformational distributions, which ensemble-averaging techniques such as cryo-EM are ill-suited to capture at the single-chain level. Electrospray Ion Beam Deposition combined with Scanning Tunnelling Microscopy (ESIBD+STM) addresses this gap, by enabling sub-nanometer imaging of individual surface-adsorbed glycoconjugates and resolving glycosidic linkages and backbone architectures at single-molecule resolution. Surface deposition acts as a controlled 2D confinement, restricting accessible conformational space while preserving chemical identity, and provides a single-chain-level readout of structural heterogeneity, which is uniquely sensitive to how sugar decoration patterns shape local and global conformation. Recovering full 3D molecular conformations from 2D topographic height maps remains a computational bottleneck. We present ATLAS, a workflow that resolves this challenge. Starting from monomer positions extracted from STM image analysis and a structured connectivity file, ATLAS proceeds through monomer generation, positioning, and force-field optimisation to produce ensembles of chemically valid 3D candidate structures ready for DFT validation. The height map acts as a biasing potential steering conformational sampling toward geometries consistent with observed topography, while enforcing correct bond lengths, valence angles, and ring conformations. Together, ESIBD+STM and ATLAS aim to close the gap between experimental imaging and structural interpretation at the single-molecule level, directly linking monomer-level chemical identity and attachment geometry to polymer-scale conformational behaviour, opening a route toward high-throughput structure-function mapping in glycoproteins.

Multiscale Simulation of Phosphofructokinase-1 Assemblies: From Transient Interactions to Large-Scale Assembly Formation

Mehrnoosh Khodam Hazrati, Tom Miclot, Stepan Timr - J. Heyrovsky Institute of Physical Chemistry, Czech Academy of Sciences

Human phosphofructokinase-1 (PFK1)—a key glycolytic enzyme—forms filaments and localizes into large-scale assemblies that are thought to play a major role in the regulation of glycolysis. However, the molecular interactions driving this assembly and the precise mechanisms by which it regulates the pathway remain poorly understood. In this work, we combine three levels of description—atomistic, residue-level coarse-grained, and highly coarse-grained—to characterize interactions between PFK1 tetramers and to elucidate factors governing PFK1 assembly formation. Atomistic molecular dynamics simulations of PFK1 filament interfaces reveal specific side-chain interactions that are critical for filament stability. These insights enable us to improve the description of filament formation in residue-level coarse-grained models. Using the Martini 3 and OPEPv7 coarse-grained models, we further identify key regions mediating transient PFK1--PFK1 interactions and show that these include filament-forming interfaces. Finally, we construct a highly coarse-grained model that integrates information from the more detailed simulations. Using this model, we investigate the role of membranes in PFK1 filament formation and describe how filaments may affect the recruitment of other constituents into large-scale glycolytic assemblies.

Improved Prediction of Single Amino Acid Variant Effects: Resolving Ambiguities and Expanding Applications

Tamas Hegedűs, Orsolya Gereben, Hedvig Tordai, Lana Khamisi - Department of Biophysics and Radiation Biology, Semmelweis University

Predicting whether single amino acid variants (SAVs) in proteins lead to pathogenic outcomes is a critical challenge in molecular biology and precision medicine. Experimental determination of the effects of all possible mutations or those observed in pathogenic individuals is infeasible. While existing state-of-the-art tools such as AlphaMissense show promise, their performance remains insufficient for diagnostic applications, they are often challenging to run locally. To address these limitations, we developed pLM-SAV, a simple yet effective predictor leveraging protein language models (pLMs). Our method computes delta-embeddings by subtracting the embedding of the mutant sequence from that of the wild type sequence. These delta-embedding vectors serve as input for a convolutional neural network used for training and prediction. To prevent data leakage, we trained our model on a well-characterized, labeled set of Eff10k and evaluated it on a non-homologous subset of ClinVar data. Our results demonstrate that this approach performs exceptionally well on the Eff10k test folds and reasonably on ClinVar test sets. Notably, pLM-SAV excels in resolving ambiguous predictions by AlphaMissense. We also found that an ensemble method, REVEL, outperforms both AlphaMissense and pLM-SAV, thus, we integrated these REVEL-enhanced predictions into our widely used AlphaMissense web application, https://alphamissens.hegelab.org. Our results demonstrate that an SAV predictor trained on labeled data can achieve high predictive performance. We anticipate that incorporating delta-embeddings into other mutation effect predictors or mutant structure prediction methods will further enhance their accuracy and utility in diverse biological contexts.

Effect of terminal modifications on aggregation properties and in vitro cytotoxicity of Staphylococcus aureus phenol soluble modulin β1 (PSMβ1)

Aleksandra Kalitnik[1], Oliwia Polańska[1], Alicja W. Wojciechowska[1], Anna Szczepańska[2], Dylan Valli[3], Przemysław Sareło[1,4], Monika Szefczyk[2], Michał Maj[3], Małgorzata Kotulska[1]

1 Department of Biomedical Engineering, Faculty of Fundamental Problems of Technology, Wrocław University of Science and Technology; 2 Department of Bioorganic Chemistry, Faculty of Chemistry, Wrocław University of Science and Technology; 3 Department of Chemistry –Ångström Laboratory, Uppsala University; 4 Pre-clinical Research Center, Wrocław Medical University

We evaluated the effects of terminal modifications on the aggregation behaviour and in vitro cytotoxicity of Staphylococcus aureus phenol-soluble modulin β1 (PSMβ1). In our work, we used the wild-type peptide (PSMβ1) and dfPSMβ1, its N-deformylated and C-amidated synthetic counterpart. Using a combination of biophysical and biological assays, we comparatively characterised fibrillation kinetics, secondary structure, aggregate morphology, and cytotoxic effects of these two peptides. We demonstrated that terminal modifications markedly influenced the fibrillation kinetics of PSMβ1 and the morphology of formed amyloid aggregates. The wild-type peptide exhibited classical, concentration-dependent amyloid fibrillation with sigmoidal kinetics, including a pronounced lag phase and rapid elongation at higher concentrations. In contrast, dfPSMβ1 lacked a sigmoidal growth profile and showed immediate but shallow ThT fluorescence increases, indicative of a non-classical aggregation pathway. ATR-FTIR analysis revealed β-sheet-dominated amide I profiles for both peptides. However, upon incubation, PSMβ1 displayed a pronounced shift toward low-frequency β-sheet bands characteristic of strongly coupled, ordered amyloid fibrils, whereas dfPSMβ1 retained higher helix/unordered contributions and β-sheets at higher wavenumbers, consistent with greater structural disorder. Cryo-TEM further demonstrated striking morphological differences of formed amyloid aggregates. PSMβ1 formed long, thin, unbranched fibrils with a regular helical twist, while dfPSMβ1 produced shorter, heterogeneous, and entangled fibrils along with dense rod- or plate-like aggregates. The cytotoxic effects of both peptides were assessed using human umbilical vein endothelial cells (HUVECs). The results revealed that dfPSMβ1 was multiple times less harmful to HUVEC cells than its wild-type counterpart. Overall, these results highlight the critical role of terminal modifications in modulating the aggregation pathways, structural outcomes, and toxicity of PSMβ1. Although the core sequences are identical, altered termini determine whether the peptide grows into elongated, orderly fibrils or architecturally diverse aggregates, and these supramolecular structural variations also correlate with cytotoxic potential.

Vesiculin Reveals a Central Role of the IGF2 C-Domain in Receptor Specificity Across the Insulin–IGF System

Jakub Kaminsky[1], Karolína Mrzílková[2], Marta Lubos[1], Klaudie Hladoníková[1], Irena Selicharová[1], Jiří Jiráček[1], Lenka Žáková[1]

1 Institute of Organic Chemistry and Biochemistry of the CAS; 2 Charles University

Insulin-like growth factor 2 (IGF2) interacts with multiple receptors of the insulin family, yet the structural determinants governing its receptor specificity remain incompletely understood. Here, we show that the deletion of four internal amino acids from IGF2 reveals an unexpectedly dominant role for the C-domain in regulating receptor recognition. We investigated vesiculin, a naturally occurring des[37–40]IGF2 isoform, by combining total chemical synthesis, quantitative binding and signaling assays, circular dichroism spectroscopy, and calculations based on microsecond-scale molecular dynamics simulations. Despite its near-identical sequence to IGF2, vesiculin displays a strikingly altered affinity profile: binding to IGF2R collapses to ~1–4% of wild-type levels, whereas affinity for IGF1R remains largely preserved and interactions with IGFBP3 are dramatically weakened. Our simulations offer an explanation for why such a slight change can cause significant shifts in affinities. In wild-type IGF2, the C-domain appears to act as a dynamic interaction hub, stabilizing binding through cooperative contacts with IGF2R domains D6 and D8. Removal of residues Arg37–Arg40 disrupts this network, forcing vesiculin into suboptimal binding modes and exposing a much more extensive and cooperative receptor interface than previously recognized. This effect propagates across the receptor family. Vesiculin exhibits reduced activation of insulin receptor isoforms, consistent with the loss of key C-domain contacts with α-CT and L2 domains, whereas IGF1R binding remains comparatively resilient. Taken together, vesiculin functions as a powerful perturbation that redefines the IGF2-binding landscape. Our results help to establish the C-domain as a central determinant of receptor specificity and uncover a hidden layer of multi-domain cooperativity in IGF2R recognition. These findings not only revise the structural model of IGF2–receptor interactions but also open new avenues for engineering IGF2 analogues with precisely tuned receptor selectivity.

Structure-based metagenomic discovery and stabilization of functionally diverse enzymes for environmental sustainability

Olga Khersonsky, Yuval Fishman, Sarel Fleishman - Weizmann Institute of Science

Metagenomic screening is critical for identifying novel enzymes with environmental, industrial and therapeutic potential. Although public genomic databases typically contain thousands of unique homologs for key enzyme families, the search for stable, functional enzymes is often hindered by the low frequency of high-performing candidates and the technical and financial requirements of high-throughput screening. Sequence-based clustering is likely to be dominated by mutations outside the active-site pocket, wasting effort on candidates with similar activity profiles. To address these challenges, we developed a scalable strategy that enables the discovery of novel functional properties in natural enzymes. We used the recent accurate AI-based structure predictors to develop a general strategy that selects candidates based on structure-based clustering and diversity of the active sites. Subsequently, we stabilize each candidate using our PROSS stability design algorithm, which was recently shown to work on AlphaFold models. This structure-based strategy for prioritizing enzymes is applicable to any homologous family of enzymes and is likely to yield large differences in catalytic properties. We are currently validating this strategy using two enzyme families important for environmental sustainability: Phosphotriesterase-like lactonases (PLLs) and laccases. PLLs are versatile biocatalysts capable of hydrolyzing toxic organophosphate compounds and interfering with bacterial quorum-sensing pathways. Laccases are multicopper oxidases widely utilized for the decomposition of synthetic dyes, recalcitrant herbicides, and polyaromatic hydrocarbon (PAH) pollutants. This research aims to expand the accessible functional diversity of natural enzymes, providing automated and rational solutions for bioremediation and pharmaceutical discovery.

Design of New Metal-Binding Peptides from Protein Fragments

Ján Michael Kormaník, Erik Andris, Tadeáš Kalvoda, Robin Kryštůfek, Milan Kožíšek, Lubomír Rulíšek - Institute of Organic Chemistry and Biochemistry of the CAS

Small metallopeptides offer a big potential for various applications, including medicine, biosensors or catalysis, but this territory is still under explored. Unfortunately, ML-based approaches are not developed to reliably design short metal binding peptides. In our recent study, we have designed a small peptide by taking fragments from proteins with known structure and joining them together. This peptide, which harbors the tetrahedral Cys2His2 motif commonly appearing in zinc fingers, binds Zn2+ ions with Kd = 220 nM.[1] In the present study, we aim to expand the scope to improve this result by including Asp and Glu as possible binding amino acids and also employing a more thorough computational protocol to select the best designs. Using a fluorescene-assay based screening, we obtained several good candidates that show strong metal-binding properties. [1] Kormaník, J. M., Herman, D., Andris, E., Culka, M., Gutten, O., Kožíšek, M., Bednárová, L., Srb, P., Veverka, V., Rulíšek, L. (2025). Design of Zn-Binding Peptide (s) from Protein Fragments. ChemBioChem, 26(7), e202401014.

Engineering Thermostable Proteins Using Large Language Models

Lukas Kuhajda[1], Tomas Honzik[1], Jiri Vondrasek[2], Daniel Georgiev[1]
1 University of West Bohemia in Pilsen, 2 Institute of Organic Chemistry and Biochemistry of the CAS

Proteins are essential functional units of life, with their activity being highly dependent on temperature. Organisms sustain metabolism across a vast temperature range, from -20°C to +122°C, while human proteins keep us alive within a narrow 30°C–42°C window. At low temperatures, enzymes struggle with insufficient kinetic energy to overcome activation barriers, whereas high temperatures cause irreversible protein denaturation and aggregation. While amino acid mutations can significantly shift a protein's optimal working temperature, no universal pattern governs this adaptation. There is a huge effort globally to discover new sources of enzymes functioning under extreme conditions. To address this, artificial neural networks (ANNs) were applied to design proteins with tailored thermostability. First, an ANN classifier was trained to categorize protein sequences into four thermostability groups: psychrophilic (<15°C), mesophilic (30°C–35°C), thermophilic (50°C–70°C), and hyperthermophilic (>80°C). The model outperforms existing approaches in third-party evaluations, and is demonstrating sensitivity to single mutations. A genetic algorithm and zero-shot prediction by the trained model were then used to introduce a chain of single mutations into a template protein, adjusting its optimal temperature upward or downward. The generated proteins exhibit high-confidence classification into the desired thermostability category, maintain maximal evolutionary conservation, preserve high structural similarity to the template protein, and display significant melting temperature shifts in molecular physics simulations. Laboratory validation is currently underway at IOCB Prague.

ECLIPSE: Exploring the functional dark proteomes of ESKAPE pathogens through sequence similarity connectomes of Protein Universe Atlas

Surabhi Lata, Dirk W.Heinz - Department of Molecular Structural Biology, Helmholtz Centre for Infection Research (HZI)

The accelerating crisis of antimicrobial resistance among ESKAPE pathogens demands the urgent identification of novel molecular targets. However, a substantial fraction of bacterial proteomes remains functionally uncharacterized, with many genes annotated as encoding hypothetical proteins. These protein sequences often lack significant similarity to known protein families when using conventional homology-based annotation methods and thus remain “dark”. This limits our ability to explore their role in pathogenicity, and it is thus crucial to bridge this substantial gap in pathogen biology by developing novel strategies to illuminate these “dark” regions of the ESKAPE panproteomes. Here we introduce ECLIPSE (ESKAPE Connectome Linkage and Inference for Proteome Sequence Exploration), a network-based framework that systematically identifies poorly characterized regions of target bacterial panproteomes. ECLIPSE embeds target-pathogen proteomes within the global sequence-similarity network of the Protein Universe Atlas [1] and detects network components composed predominantly of functionally unannotated proteins (“dark nodes”). These components are prioritized using taxonomic diversity analysis, where Shannon entropy indices quantify the richness and evenness of dark components. This strategy highlights evolutionarily conserved yet poorly characterized protein groups that are enriched in multidrug-resistant pathogens. We have extracted a portion of proteins from these taxonomically stratified components which are enriched with ESKAPE pathogen and Pseudomonas specific components. To extract biologically relevant nodes from a list of dark components we have classified the protein according to its length and are subsequently explored through downstream analyses, including structure-guided functional inspection using AlphaFold-predicted models. In a case study of the opportunistic Gram-negative bacterial pathogen Pseudomonas aeruginosa panproteome, ECLIPSE identifies multiple evolutionarily coherent representatives of the “dark proteome” that are structurally defined yet functionally unannotated, several of which are also enriched across other ESKAPE pathogens. These candidates may contribute to virulence or multidrug resistance, thus representing promising targets for experimental characterization. Reference: Durairaj, J., Waterhouse, A. M., Mets, T., Brodiazhenko, T., Abdullah, M., Studer, G., Tauriello, G., Akdel, M., Andreeva, A., Bateman, A., Tenson, T., Hauryliuk, V., Schwede, T., & Pereira, J. (2023). Uncovering new families and folds in the natural protein universe. Nature, 622(7983), 646–653

Structure-Based Pharmacophore Modeling of DprE1 and Protein–Protein Complex Modeling of DprE1–DprE2 for Anti-Tuberculosis Drug Discovery

Csongor Mátyás, Blanka-Eszter NAGY, Krisztina BOROS, Levente Csaba NAGY, Jürgen BRÉM, László Csaba BENCZE - Babeș-Bolyai University, Enzymology and Applied Biocatalysis, Faculty of Chemistry and Chemical Engineering

The fight against tuberculosis continues to face critical hurdles, particularly due to the emergence of antibiotic-resistant strains of Mycobacterium tuberculosis. A promising strategy to overcome this challenge involves targeting the essential epimerases DprE1 and DprE2, which are crucial for cell wall biosynthesis. In this study, we focused on the stabilization and structural characterization of DprE1 and its interaction with DprE2 to support rational drug design. Using a curated set of high-resolution X-ray crystallographic structures of DprE1, we conducted a comprehensive structural analysis to identify conserved features within the active site. This enabled the development of a robust, structure-based consensus pharmacophore model that captures recurring ligand-protein interaction patterns across diverse chemotypes. The pharmacophore model was validated through molecular docking (AutoDock Vina with Vina and Vinardo scoring functions), which successfully reproduced known binding poses and informed the selection of optimal protein conformations for virtual screening. The combined use of pharmacophore filtering and docking enabled more targeted screening of the Enamine database, focusing on novel scaffolds and fragment-derived compounds. Although docking scores alone showed limited correlation with reported IC_{50} and MIC values for over 500 compounds, integration with our pharmacophore model enhanced the relevance of selected hits. In parallel, we constructed a structural model of the DprE1–DprE2 complex using protein-protein docking and homology modeling approaches, aiming to better understand the interface and allosteric effects that may influence DprE1 function and ligand binding. These models were further examined for possible interactions, using CAVER 3.0 and molecular dynamics simulations. Ongoing work includes experimental validation and further refinement of the DprE1-DprE2 complex model to explore its implications for protein stabilization and ligand binding dynamics. Acknowledgements: This work was supported by the BACPROBIO project, funded by European Union – NextGenerationEU and the Romanian Government, under National Recovery and Resilience Plan for Romania, contract no 760251./28.12.2023, cod PNRR-C9-I8-CF92/31.07.2023, through the Romanian Ministry of Research, Innovation and Digitalization.

Native-State Heterogeneity in Leptin Reveals Hidden Structural and Oligomeric States

Grace E. Orellana[1], Chad A. Brautigam[2], Yao Hu[3], Qian Wang[3], and Ellinor Haglund[1], Vysakh Ramachandran[4], Davit Potoyan[4]

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 Chemistry, Iowa State University, Ames, USA

Leptin is a pleiotropic hormone that regulates energy homeostasis, satiety, and food intake, and is transported out of the cell through secretory vesicles where it encounters environments that vary in pH, protein concentration, and macromolecular crowding. Liquid–liquid phase separation (LLPS), a process in which a homogeneous solution separates into two liquid phases, has been linked to protein aggregation in vivo. Interestingly, we observe that wild-type leptin can form irreversible aggregates, raising the question of whether transient intermolecular interactions serve as a nucleation point for aggregation in diet-induced obesity. Because leptin aggregates at high concentrations in vitro, the chemical environment within secretory vesicles may promote aggregation during the secretory pathway. We therefore investigated the role of pH in leptin folding and aggregation and found that lower pH stabilizes leptin, possibly reflecting the environment of the secretory pathway. We hypothesize that rapid changes in chemical environment may also promote phase separation of leptin, thereby increasing its aggregation propensity.

Exploring the physiological amylome and the molecular mechanisms of its regulation

Dorota Maria Pfizenmaier[1], Anastasiia Kovalenko[1,2], Caroline Wilson-Zbinden[1], Gintare Petraityte[1], Federico Uliana[1,3], Izabella Krystkowiak[4], Claudia C. Schmidt[1,2], Ino Karemaker[1], Alvar Gossert[1,5], Norman Davey[4], Matthias Peter[1,2]

1 Institute for Biochemistry, Department of Biology, ETH Zürich; 2 Bringing Materials to Life Initiative, ETH Zürich; 3 Johannes Gutenberg University Mainz, Biozentrum; 4 Division of Cancer Biology, The Institute of Cancer Research, London; 5 Biomolecular NMR Spectroscopy Platform (BNSP), Department of Biology, ETH Zürich

Amyloid aggregation was conventionally considered irreversible and pathological, especially in the context of neurodegenerative diseases. However, recent research revealed that stress-induced, reversible amyloids perform important cellular functions. Yet little is known about reversible amyloids and their common biochemical and regulatory features. Yeast pyruvate kinase Cdc19 and its human homologue, PKM2, are examples of such stress-inducible reversible amyloids. They are regulated via pH-sensing core motifs, distinguished by amyloid properties and protonatable/charged residues. Using these core motifs as paradigm, we designed a bioinformatic screen and an in vitro validation pipeline to search for pH-responsive, reversible amyloid cores in yeast and human proteomes. This approach allowed for discovery of novel pH-sensing amyloid sequences and identification of their common biophysical properties. Selected full-length candidates with conserved pH-sensing amyloidogenic regions were further analyzed in yeast using fluorescence microscopy and SDS-resistance assays, revealing multiple proteins aggregating in stationary phase. Among them, we focused on the asparagine synthetase Asn1 and characterized the regulation and physiological relevance of its pH-dependent aggregation. In another line of experimentation, we show that not only pH-dependent protonation, but also phosphorylation can modulate the charge of amyloid cores and thus regulate their assembly and disassembly. Altogether, these discoveries expand our understanding of stress-inducible, reversible protein aggregation and its function.

Ovo, an Open-Source Ecosystem for De Novo Protein Design

David Prihoda[1], Marco Ancona[1], Tereza Calounova[1], Adam Kral[1], Lukas Polak[1], Hugo Hrban[1], Nicholas J. Dickens[2], and Danny A. Bitton[1]

1 Applied Research and Innovation, MSD Czech Republic s.r.o.; 2 Amazon Web Services LLC

The protein design field is rapidly advancing, with frequent emergence of new models and pipelines for designing de novo proteins with tailored properties and functions not found in nature. However, the current tool landscape is fragmented, tools are hard to install and deploy, and require significant computational expertise to integrate into end-to-end, scalable pipelines. A particular challenge is managing many sequences, structures, and metrics for downstream testing and retrospective analysis of input parameters. To address this need, we introduce Ovo, an open-source de novo protein design ecosystem that consolidates models, workflows, data management, and interactive visualization into a scalable, infrastructure-agnostic platform. Ovo features Nextflow-based workflow orchestration, a storage layer, and both command-line and graphical interfaces that democratize scaffold design, binder design and diversification, and validation workflows. Ovo's novel ProteinQC module computes comprehensive sequence and structure descriptors, contextualizing designs against reference sets. Ovo plugins let the community add new workflows and user interfaces to accelerate adoption of emerging methods and facilitate community-driven benchmarking. Ovo lowers engineering barriers and demystifies the design process, allowing experts and non-technical users to design proteins at scale. With community-driven development, Ovo can accelerate de novo protein design and advance discovery in therapeutics and biotechnology.

Advancing Protein Engineering: Reliable Functional Evaluation and Selective Binder Activation

Jakub Ptacek, Katerina Jiraskova, Klara Postulkova, Zsofia Kutilova, Jiri Vondrasek - Institute of Organic Chemistry and Biochemistry of the CAS

Engineering proteins for real-world applications requires not only improved function, but also reliable ways to evaluate performance and control activity in complex biological environments. Our poster brings together two complementary projects addressing these challenges. In the area of enzymatic plastic degradation, we focus on PETases as promising tools for PET recycling, where progress is often slowed by poor reproducibility and difficult cross-study comparison. Building on our benchmarking work, we established a standardized framework for comparing major Ideonella sakaiensis PETase variants and developed the RAVEN platform for transparent visualization of enzyme performance across assay conditions. We further showed that differences in N-terminal processing contribute substantially to the variability of apparent PETase activity, highlighting sample quality as an important but often overlooked determinant of experimental outcome. These results provide a more reliable foundation for PETase research, while also supporting our broader aim to move beyond incremental optimization of existing scaffolds toward new PET-degrading enzymes designed by de novo and machine-learning-guided approaches, with the goal of identifying biocatalysts better suited for industrially relevant conditions. In parallel, we are developing a modular strategy for tumor-selective activation of protein binders. The concept uses tumor-associated mild extracellular acidosis as a spatial cue to activate an otherwise masked targeting protein. In our design, the HER2-binding affibody ABY-025 serves as the validation binder, while HIV-1 protease functions as an acidity-tuned masking and self-cleaving module: under near-physiological conditions the construct is intended to remain in an OFF state, whereas under mildly acidic tumor-like conditions it should unmask itself and restore target binding. The overall goal is to establish a transferable platform for tumor-acidity-gated protein prodrugs that can improve selectivity and reduce off-target activity.

Design of proteins by temperting of message passing graph neural network

Vojtěch Spiwok, Klára Kostelníková - University of Chemistry and Technology, Prague

We present a method for free (untemplated) or partially conditioned design of proteins. The sequence/structure is iteratively optimized by predicting its structure and optimization of its sequence. The process can be controlled by a variable temperature, which depends on the agreement of the structure or sequence with the predefined goals.

Large-scale experimental exploration of protein space

Elsa Stern[1], Thomas Müntener[1], Dylan Abramson[2], Ines Hertel[1], Gytis Jankevicius[1], Guillaume Mas[1], Gert E.Folkers[1], Basile I. M. Wicky[2], Sebastian Hiller[1]
1 University of Basel, 2 ETH Zürich

The emergence of deep learning and artificial intelligence has ushered in a new era of statistical structural biology, in which the hard-earned data of the past decades is employed to create new knowledge and enable, for example, the design of fully new proteins. However, the basic data on which their success is based are not ideal for this unforseen task due to their heterogeneity and the lack of negative examples. With the recent advances in protein design and large-scale automated protein production, it is possible to generate systematic datasets suitable for AI training. With such a pipeline, employing nuclear magnetic resonance as the analytical method of choice, 384 de novo designed proteins are produced and characterized.

Short Peptides as Model Systems for Mapping Sequence–Thermodynamics Relationships

Josef Šulc, Jiri Vondrasek - Institute of Organic Chemistry and Biochemistry of the CAS

Short peptides provide experimentally and computationally tractable model systems for studying how amino-acid sequence controls conformational behavior, solvation, entropy, and free-energy landscapes. Although they lack the full structural context of folded proteins, they capture fundamental physical principles that shape intrinsically disordered regions, protein linkers, terminal tails, loops, and flexible connectors in multidomain architectures. Their limited size makes them especially suitable for systematic molecular dynamics studies, where large sequence libraries can be compared across shared structural coordinates and thermodynamic descriptors.

We present the current state of a near-production computational framework for building a thermodynamic atlas of short peptide ensembles. The platform integrates molecular dynamics-derived feature tables, potentials of mean force, cumulative distributions, replica diagnostics, convergence metrics, Ramachandran observables, and optional time-series data into a unified analysis environment. Each peptide is represented not as a single structure, but as an ensemble of probability distributions over structural coordinates such as terminal distance, backbone torsions, compactness, contacts, and solvent exposure.

From these distributions, the framework derives standardized descriptors including free-energy minima, thermally accessible fractions within 1–5 kT, basin populations, free-energy widths, barriers, local curvature, entropy-like measures, and convergence-aware uncertainty estimates. This allows short peptides to be compared as thermodynamic objects, revealing how sequence motifs such as glycine-rich, serine-rich, proline-rich, charged, polar, or mixed compositions bias extension, collapse, flexibility, and structural degeneracy.

High-dimensional peptide relationships are explored using PCA, UMAP, PMF-based clustering, and region-conditioned PMF averaging. In this representation, peptides cluster according to similarities in their free-energy landscapes rather than sequence alone. UMAP axis-correlation and PMF-bin attribution connect embedding regions back to molecular states, such as torsional preferences, compactness regimes, or shifts in end-to-end distance distributions.

This project aims to establish short peptides as a systematic model space for linking sequence, ensemble structure, and thermodynamic behavior. By turning large simulation datasets into interpretable maps, the atlas provides a foundation for peptide classification, force-field comparison, convergence assessment, and rational design of flexible or structured peptide elements in larger biomolecular systems.

Drosophila as a Model for Structural and Functional Conservation of Leptin Signaling

Amanda Taitano, Joanne Yew, Ellinor Haglund - University of Hawaii at Manoa

Analogous proteins across species reveal how conserved signaling outcomes can arise from proteins with divergent sequences. In Drosophila melanogaster (Drosophila), the cytokine-like protein Unpaired 1 (Upd1) acts as a leptin analog through the conserved JAK/STAT pathway despite low sequence similarity. To explore this conservation, we generated a neuronal Upd1 knockdown line and examined metabolic phenotypes in vivo. Reduced Upd1 expression resulted in an increase in weight and triacylglycerol (TAG) levels under both standard and high-fat diet conditions. These results establish a clear genotype-phenotype link between reduced Upd1 and metabolic imbalance similar to ob/ob mice (leptin knockout). Fertility and fecundity assays are currently being performed to determine whether reduced Upd1 signaling also affects reproduction, as fertility is shown to be impaired in mammalian leptin-deficient models. In parallel, we are expressing and purifying the protein in vitro for structural studies. Both full-length and truncated constructs are being tested to improve solubility and allow downstream structural analysis. AlphaFold predictions suggest a compact helical core surrounded by flexible loops and regions of intrinsic disorder, which may contribute to the aggregation observed during purification. Together, these in vivo and in vitro results help establish this fly system as a model for probing leptin signaling.

CVFormer: Data-Driven Collective Variables via Transformer Autoencoders

Guglielmo Tedeschi, Vojtech Spiwok - Department of Biochemistry and Microbiology, University of Chemistry and Technology, Prague

Molecular Dynamics simulations provide insights on the biomolecular motion, however it is intrinsically limited by timescale separation. Rare events such as folding, conformational transitions, or allosteric rearrangements often occur on timescales that exceed those accessible to standard MD. Enhancing sampling methods, such as Metadynamics, addresses this limitation by enhancing sampling along a set of Collective Variables (CVs), which act as low-dimensional descriptors of the system's slow degrees of freedom. However, the effectiveness of metadynamics critically depends on the quality of the chosen CVs, suboptimal variables can lead to inefficient, or incomplete, exploration of the free-energy surface. The identification of physically meaningful and dynamically relevant CVs therefore remains a central challenge in enhanced sampling. CVFormer is a Transformer-based autoencoder for the extraction of CVs from Molecular Dynamics trajectories. The model compress conformational information into a low-dimensional latent space (typically 2D), suitable for driving enhanced sampling simulations such as metadynamics. By analyzing this reduced representation, we can not only characterize metastable states and transition pathways, but also trace back which structural degrees of freedom are responsible for shaping the learned collective variables. A key aspect of the approach is the fully data-driven selection of CVs. Instead of relying on handcrafted descriptors, the model learns representations directly from molecular dynamics trajectories. Attention weights provide residue-level importance scores, enabling the identification of structural regions that contribute most strongly to the learned collective variables. This interpretability is complemented by mutual information analysis within the latent coordinates, allowing a quantitative assessment of dependency and redundancy. The methodology is demonstrated on the Trp-cage mini-protein, a well-established benchmark for folding and conformational studies. CVFormer successfully learns a low-dimensional representation that captures the essential folding landscape and separate metastable states. The analysis highlights specific residues with dominant contributions to the learned CVs, in agreement with known structural determinants of Trp-cage stability. The selected CVs were subsequently used to bias metadynamics simulations, leading to efficient enhancement of conformational transitions and accelerated crossing of free-energy barriers.

Encoded and non-genetic alternative protein variants expand human functional proteome

Vjaceslav Tretjacenko[1], Tehila Leiman[1], David Morgenstern[2], Omer Asraf[1], Orna Dahan[1], Dvir Dahary[1], Yitzhak Pilpel[1]

1 Department of Molecular Genetics, Weizmann Institute of Science; 2 Nancy and Stephen Grand Israel National Center for Personalized Medicine G-INCPM, Weizmann Institute of Science

Each stage of the Central Dogma contributes to proteome diversity through mechanisms such as heterozygosity, somatic mutations, transcriptional errors, and translational errors. As a result, a diverse array of protein variants can coexist within a single proteome, such as that of humans. However, until now, methods to detect, quantify, and evaluate the functional consequences of these variants have been lacking. Here we examined a large-scale proteogenomic dataset from 29 healthy human tissues and uncovered 13,910 confidently localized variants representing 7,215 unique single amino acid substitutions co-existing alongside their corresponding reference proteoforms. We found that the abundance of both genetic (SNP's, somatic mutations) and mistranslated protein variants mirrors their allele frequencies in the human population. Moreover, we show that non-genetic substitutions may provide a distinct route for exploring protein sequence space, circumventing the mutational constraints imposed by the genetic code. In addition, we provide experimental validation of non-genetic substitution on selected purified proteins. We demonstrate specific and recurring non-genetic variation patterns upon amino acid starvation in proteome-wide analyses of cancer-derived cell lines and identify hundreds of substituted non-genetic proteoforms that recur consistently in multiple healthy individuals or map to annotated protein functional sites. We propose that these substitutions constitute a novel class of functional protein phenotypic variants. Finally, we highlight potential non-genetic routes of immunoglobulin diversification. Collectively, our findings indicate that non-genetic amino acid substitutions in human proteins provide an abundant source to expanding the functional proteome.

Discovering Distal Mutations that Shape Enzyme Dynamics Using Ultra-High-Throughput Droplet Screening

Pavel Vanacek[1,2,6], Hana Faldynova[1,2], David Kovar[1,2], Ankit Jain[3], Michaela Slanska[1], Marek Martinek[4], Andreas Jakob[4], Marie Sulova[1,2], Michal Vasina[1,2], Joan Planas-Iglesias[1,2], Sergio M. Marques[1,2], Naina Verma[6], David Damborsky[1,2], Christoffel P. S. Badenhorst[5], Tomas Buryska[3], Flora Wing Yin Chiu[3], Marika Majerova[1,2], Tana Kohutekova[1], Petr Kouba[7], Nela Sendlerova[1], Andrew deMello[3], Jiri Damborsky[1,2], Josef Sivic[7], Uwe T. Bornscheuer[5], David Bednar[1,2], Stanislav Mazurenko[1,2], Lenka Hernychova[6], Martin Marek[1,2], Petr Klan[4], Stavros Stavrakis[3], Zbynek Prokop[1,2]
1 Loschmidt Laboratories, Department of Experimental Biology and RECETOX, Masaryk University; 2 International Clinical Research Centre, St. Anne's University Hospital Brno; 3 Institute for Chemical and Bioengineering, ETH Zürich; 4 Department of Chemistry and RECETOX, Masaryk University; 5 Institute of Biochemistry, University of Greifswald,

Enzyme performance with non-native or bulky substrates is often limited not by chemistry in the active site, but by long-range dynamics and substrate-access pathways, which remain difficult to identify by rational design. Here, we applied ultra-high-throughput fluorescence-activated droplet sorting with a sterically demanding fluorogenic substrate to explore dynamic constraints shaping catalysis. Screening a focused variant library yielded multiple distal mutations, all positioned more than 11 Å from the catalytic centre. These substitutions consistently decreased apparent KM, indicating improved substrate acquisition or tunnel opening. Two variants exhibited contrasting mechanisms: one increased catalytic efficiency through enhanced cap-domain flexibility, while another reduced substrate inhibition and shifted halogen specificity through local restructuring of the access tunnel. Integrated analyses combining HDX-MS, molecular dynamics, QM/MM calculations, and ML-based flexibility predictions revealed that selection with bulky substrates favours mutations modulating conformational dynamics rather than active-site chemistry. These findings highlight the power of ultrahigh-throughput microfluidic screening to uncover hidden dynamic hotspots that govern substrate routing and enable enzyme adaptation. Acknowledgement: This work was supported by the European Union and the MEYS of the Czech Republic within ESIF-MEYS Johannes Amos Comenius Programme (OP JAC) under the project CLARA (CZ.02.01.01/00/23_029/0008437 and 101136607).

ML-Based Conformational Sampling and Ranking Do Not Improve Rigid Protein–Protein Docking

Semen Yesylevskyy[1,2], Roman Stratiichuk[2], Roman Kyrylenko[2], Ihor Koleiev[2], Illia Savchenko[2], Taras Voitsitskyi[2], Vladyslav Husak[2], Serhii Starosyla[2], Alan Nafiiev[2]
1 Institute of Organic Chemistry and Biochemistry of the CAS; 2 Receptor.AI Inc.

Rigid protein–protein docking remains the dominant computational approach for predicting protein complexes when experimental data are unavailable, but it critically depends on avilable unbound (Apo) conformations that often differ substantially from the bound (Holo) state. ML-based conformational sampling methods, such as AFSample2 and BioEmu, promise to close this gap by generating near-Holo conformational ensembles prior to docking. We rigorously tested this hypothesis on complexes from the PINDER-AF2-Apo benchmark, a curated dataset designed to minimize ML training data leakage. Conformational ensembles generated by AFSample2 and BioEmu were clustered and subjected to all-versus-all rigid docking with HDOCK, yielding over 43,000 predictions evaluated by CAPRI criteria. Performance was compared against standard Apo docking using physics-based (HADDOCK3), knowledge-based (HDOCK), and ML-based (dMaSIF, PIsToN) scoring functions. Our results reveal two compounding failures. First, ML sampling rarely produces conformations closer to the Holo binding interface than the original Apo structures. Second, even when near-Holo conformations are present in the ensemble, no tested scoring function can reliably prioritize them. ML-based scorers generalize poorly beyond their training data, while physics-based alternatives fall short of native HDOCK scoring. Increasing sampling depth to 200 conformations per subunit further worsened results by diluting ensembles with noise. These findings define a clear double bottleneck in ML-enhanced docking workflows - inadequate conformational coverage of binding-relevant states and insufficient scoring quality.

List of participants

Derrick Agwora
nyamwayd@natur.cuni.cz
Charles University in Prague

David Budean
dbudean@hawaii.edu
University of Hawai'i of Mānoa

Prokopis C. Andrikopoulos
prokopios.andrikopoulos@lf1.cuni.cz
First Faculty of Medicine, Charles University

Chi Fung Chow
chow@mpi-cbg.de
MPI-CBG

Francisco Javier Arriaza Gallardo
francisco.arriaza1@msd.com
MSD Czech Republic

Patricia Clark
pclark1@nd.edu
University of Notre Dame

David Baker
dabaker@uw.edu
University of Seattle

Valerie Daggett
daggett@uw.edu
University of Washington

Adolfo Bastida-Pascual
bastida@um.es
Universidad de Murcia

Jiří Damborský
jiri@chemi.muni.cz
Masaryk University, ICRC-FNUSA

Karel Berka
karel.berka@upol.cz
Palacky University Olomouc

Nikolay Dokholyan
axp2yt@virginia.edu
University of Virginia

Jakub Borbála
jakab.bori98@gmail.com
Pázmány Péter Catholic University

Michal Doležal
michal.dolezal@uochb.cas.cz
IOCB Prague

Alexandra Botková
botkova@natur.cuni.cz
Charles University

VU Duc Duy
duc-duy.vu@pasteur.fr
Institut Pasteur Paris

Filip Buchel
filip.buchel@gmail.com
Charles University

Gabor Erdos
gabor.erdos@ttk.elte.hu
ELU Budapest

Wenrui Fan
wenrui.fan@sheffield.ac.uk
Centre for Machine Intelligence, University of Sheffield

Samuel Fanči
samofanci@gmail.com
Charles University

Christos Feidakis
christos.feidakis@uochb.cas.cz
IOCB Prague

Stephen Fried
sdfried@jhu.edu
John Hopkins University

Eliska Fürst
eliska.fuerst@mdc-berlin.de
Max Delbrück Center Berlin

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

Claudia Leticia Gomez Flores
c.gomez@fkf.mpg.de
Max Planck Institute for Solid State Research

Ambrus Gordos
ambrus.gordos@vrgtherapeutics.com
Eötvös Loránd University/ VRG Therapeutics

Moushmi Goswami
moushmi.goswami@lf1.cuni.cz
First Faculty of Medicine, Charles University

Ellinor Haglund
ellinorh@hawaii.edu
University of Hawai'i of Mānoa

Mehrnoosh Hazrati
mehrnoosh.hazrati@jh-inst.cas.cz
J. Heyrovský Institute of Physical Chemistry of the CAS

Tamás Hegedűs
tamas@hegelab.org
Semmelweis University

Sebastian Hiller
sebastian.hiller@unibas.ch
University Basel

Tiffany J. Callahan
callahantiff@gmail.com
Manas AI

David Jakubec
david.jakubec@dianabiotech.com
DIANA Biotechnologies, a.s.

Lukasz Joachimiak
Lukasz.Joachimiak@UTSouthwestern.edu
University of Texas

Aleksandra Kalitnik
aleksandra.kalitnik@pwr.edu.pl
Wrocław University of Science and Technology

Jakub Kaminský
kaminsky@uochb.cas.cz
IOCB Prague

Olga Khersonsky
olga.khersonsky@weizmann.ac.il
Weizmann Institute of Science

Alena Khmelinskaia
akhmelin@cup.lmu.de
Ludwig-Maximilians-Universität München

Petro Khoroshyy
khoroshyy@uochb.cas.cz
IOCB Prague

Ján Michael Kormaník
kormanik@uochb.cas.cz
IOCB Prague

Nelia Korotushak
neliakorotushak@gmail.com
Charles University

Radoslav Krivák
rkrivak@gmail.com
IOCB Prague

Lukas Kuhajda
kuhajdal@ntis.zcu.cz
University of West Bohemia in Pilsen

Sandro Kuppel
sandro.kuppel@helmholtz-munich.de
Institute for Computational Biology

Zsófia Kutilová
zsofia.kutilova@uochb.cas.cz
IOCB Prague

Surabhi Lata
surabhilata94@gmail.com
Helmholtz Centre for Infection Research

Leonidas Lundell
leonidas.lundell@msd.com
MSD Czech Republic

George Makhatadze
makhag@rpi.edu
Rensealler Polytechnic Institute

Csongor Mátyás
csongor.matyas@ubbcluj.ro
Enzymology and Applied Biocatalysis
Research Center - BBU

Dirk Maurer
dirk.maurer@cytiva.com
Cytiva

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

Grace Orellana
georella@hawaii.edu
University of Hawai'i of Mānoa

Dorota Maria Pfizenmaier
dorota.pfizenmaier@bc.biol.ethz.ch
IBC, ETH Zürich

Ilaria Piazza
ilaria.piazza@scilifelab.se
Stockholm University

Antonio Pineda-Lucena
apinedal@unav.es
CIMA Universidad de Navarra

Tomas Pluskal
tomas.pluskal@uochb.cas.cz
IOCB Prague

Vojtech Poriz
vojtech.poriz@msd.com
MSD Czech Republic

Klara Postulkova
klara.postulkova@uochb.cas.cz
IOCB Prague

David Příhoda
david.prihoda@gmail.com
MSD Czech Republic

Iva Pritisanac
iva.pritisanac@helmholtz-munich.de
Institute of Computational Biology

Jakub Ptáček
jakub.ptacek@uochb.cas.cz
IOCB Prague

Bruno Ruiz Blanco
bruno.ruiz-blanco@curevac.com
CureVac SE

Gerd Schluckebier
schluckg@gmx.net
Novo Nordisk A/S

Eugene Shakhnovich
shakhnovich@chemistry.harvard.edu
University of Harvard

Ayala Shiber
ayalashiber@technion.ac.il
Israel Institute of Technology

Ahrum Son
ahson@scripps.edu
Scripps Research Institute

Andrea Soranno
soranno@wustl.edu
Washington University

Vojtech Spiwok
spiwokv@vscht.cz
University of Chemistry and Technology, Prague

Elsa Stern
e.stern@unibas.ch
University of Basel/Biozentrum

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

Amanda Taitano
ataitano@hawaii.edu
University of Hawai'i of Mānoa

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

Štěpán Timr
stepan.timr@jh-inst.cas.cz
Heyrovsky Institute Prague

Agnes Toth-Petroczy
tothpet@mpi-cbg.de
Max Planck Institute

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

Pavel Vaňáček
pavelvanacek@gmail.com
FNUSA ICRC

Sameer Velankar
sameer@ebi.ac.uk
EMBL-EBI

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

Gennady Verkhivker
verkhivk@chapman.edu
Chapman University

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

Tereza Votrubová
tereza.votrubova@uochb.cas.cz
IOCB Prague

Basile Wicky
basile.wicky@bsse.ethz.ch
ETH Zürich

Semen Yesylevskyy
semen.yesylevskyy@uochb.cas.cz
IOCB Prague

Notes

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

Editors: Jiří Vondrášek, Anna Strachotová, Tereza Votrubová
Print: VENICE Praha s.r.o.
Year: 2026
Pages: 84

Supported by

IOCB TECH

Peptone

BeneMeat

Schoeller INSTRUMENTS

Published by IOCB Prague
ISBN 978-80-86241-73-9