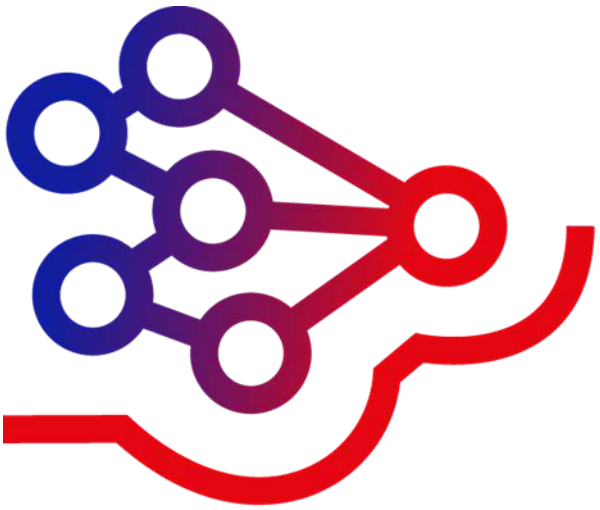

ML4NGP

MACHINE LEARNING FOR NON GLOBULAR PROTEINS

ABSTRACT BOOK

4TH MEETING on MACHINE LEARNING AND NON-GLOBULAR PROTEINS

19 – 22, May 2026

Targowa Creativity Centre, Warsaw, Poland

cost
EUROPEAN COOPERATION IN SCIENCE & TECHNOLOGY

Funded by the European Union

KEYNOTE SPEAKERS

Prof. Rohit Pappu
Washington Univ. in St. Louis

USA

Prof. Ben Schuler
University of Zurich

Switzerland

Prof. Andrea Sinz
University of Halle-Wittenberg

Germany

Prof. Tuomas Knowles
University of Cambridge

UK

INVITED TALKS

I01

Beyond static structures: how protein language models help understanding protein flexibility

Michael Heinzinger

Institute of Computational Biology, Helmholtz Center, Munich, 85764, Germany.
TUM School of Computation, Information and Technology, Technical University of Munich, Munich, 85748, Germany.

“no information provided”

I02

Peptide dynamics inside condensates on different scales

Michael Feig

Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, Michigan 48824, United States

"no information provided"

I03

Intrinsically Disordered Proteins Shape Organ-Specific Aging Trajectories

Michal Sharon

Department of Biomolecular Sciences, Weizmann Institute of Science, Rehovot 7610001, Israel

„no information provided“

I04

Intrinsic disorder and fibril formation by the Henipavirus W proteins: molecular grammar and functional impact

Sonia Longhi

AFMB - Architecture et fonction des macromolécules biologiques, Faculté des Sciences de Luminy - Case 932 163, Avenue de Luminy 13288 Marseille Cedex 09 - France

„no information provided“

I05

Molecular Models of Brain-Derived α-Synuclein Fibrils Reveal a Fuzzy-Coat-Mediated Mechanism for Selective Peptide Binding

Salvador Ventura

Institut de Biotecnologia i Biomedicina and Departament de Bioquímica i Biologia Molecular, Universitat Autònoma de Barcelona, Bellaterra, Barcelona, 08193, Spain
Hospital Universitari Parc Taulí, Institut d'Investigació i Innovació Parc Taulí (I3PT-CERCA), Universitat Autònoma de Barcelona, Sabadell, Spain

"no information provided"

I06

Chaperone-mediated regulation of tau phase separation

Daniela P. Freitas [1,2], Gea Cereghetti[3], Tanushree Agarwal[3], Giulio Tesei[4, 5], Fan Cao[4], Kresten Lindorff-Larsen[4)] Tuomas P. J. Knowles[3], Cláudio M. Gomes[1,2]

[1]Biosystems and Integrative Sciences Institute, Faculdade de Ciências, Universidade de Lisboa, Portugal
[2]Department of Chemistry and Biochemistry, Faculdade de Ciências, Universidade de Lisboa, Portugal;
[3]Yusuf Hamied Department of Chemistry, Centre for Misfolding Diseases, University of Cambridge, United Kingdom;
[4]Department of Biology, University of Copenhagen, Denmark;
[5]Department of Biomedical Science, Faculty of Health and Society, Malmö University, Sweden;
cmgomes@fc.ul.pt , https://folding.campus.ciencias.ulisboa.pt/

Tau is an intrinsically disordered protein implicated in several neurodegenerative disorders. It undergoes self-assembly through pathological aggregation into fibrillar amyloids, but can also accumulate in metastable membraneless condensates via liquid–liquid phase separation. Although the biological significance of tau condensates remains incompletely understood, growing evidence indicates that phase separation and aggregation are mechanistically linked, with condensates potentially promoting the formation of toxic species. This raises the possibility that endogenous modulators, including chaperones, regulate the phase separation-aggregation continuum.
We have previously shown that S100B, a small Ca^{2+}-binding protein highly expressed in the brain, acts as a chaperone that inhibits tau aggregation and modulates its phase separation [1,2]. Building on these findings, we carried out a detailed mechanistic analysis of this modulatory effect by combining experimental and computational approaches. We used CALVADOS, a coarse-grained model for intrinsically disordered and multidomain proteins, to predict the influence of S100B on tau phase separation under different crowding and ionic strength conditions. We then employed PhaseScan, a microfluidic platform that enables high-resolution phase diagram determination, together with complementary biochemical approaches, to characterise this behaviour in vitro.

Both computational and experimental analyses confirmed that tau undergoes phase separation and that this process is inhibited by S100B. At lower ionic strength, the calcium-bound form of S100B exerted stronger inhibition than the apo form, whereas the opposite trend was observed at higher ionic strength. The data further suggest that masking of the tau microtubule-binding region by S100B contributes to this effect, with S100B partitioning into tau condensates. In addition, in silico analyses indicated a non-uniform distribution of S100B within tau droplets, consistent with observations by confocal microscopy.
Overall, S100B counteracts tau demixing, highlighting a potential role for S100 chaperones in regulating tau condensate behaviour under physiologically relevant conditions. In this communication, I will discuss these findings within a broader perspective on chaperone-mediated modulation of phase separation in other disease-relevant proteins.

[1] Moreira & Gomes J Neurochem 2023, 166 (1), 76-86. DOI 10.1111/jnc.15756
[2] Moreira et al. Nat. Comm. 2021, 12 (1), 6292. DOI 10.1038/s41467-021-26584-2

Funding: Funded by the Fundação para a Ciência e Tecnologia (Portugal) through grant NeuroPHASE 2024.15005.PEX (to C.M.G), Ph.D. fellowship 2023.03769.BD (to D.P.F) and center grant UID/04046/2025 (to BioISI). Funded by the European Union through TWIN2PIPSA - Twinning for excellence in biophysics of protein interactions and self-assembly (GA 101079147)

I07

Beta-Arches as a Key Element in Improving Structure-Based and Machine-Learning Predictors of Amyloid Formation

Andrey Kajava

CRBM, CNRS, Montpellier, France

Amyloids are linked to serious human diseases, which include, but are not limited to, neurodegenerative diseases. Proteins also form “functional” amyloids that fulfill beneficial roles in organisms. Considering that the propensity of proteins to form amyloids is inherently determined by the amino acid sequence, a number of computational methods for predicting protein amyloidogenicity based on amino acid sequence analysis have been developed. Many of these methods relied on experimental data from short peptides studied in vitro, often under non-physiological conditions. However, the ultimate goal has always been to accurately predict amyloid formation in naturally occurring and disease-related proteins and peptides.
Analysis of known amyloid structures revealed that most naturally-occurring and disease-related amyloids invoke columnar structures produced by in-register stacking of “beta-arches.” This structural insight has enabled the development of structure-based methods for predicting amyloid formation and co-aggregation.
Today, with artificial intelligence revolutionizing many fields, researchers are exploring its potential for predicting amyloid formation. In this talk, I will show how knowledge of amyloid structures, particularly the concept of beta-arches, can enhance machine-learning predictors of amyloidogenicity. I will present benchmark results comparing our approach with existing methods,

demonstrating that the synergy between human expertise and AI is essential for building efficient ML predictors.

I08

AI-based protein analysis from multimers to metagenomes

Milot Mirdita

School of Biological Sciences, Seoul National University, Seoul, Republic of Korea

“no information provided”

I09

Order within intrinsically disordered regions: function, structure and evolution

Miguel Andrade
Institute of Organismic and Molecular Evolution, Faculty of Biology, Johannes Gutenberg University Mainz, Mainz, 55118, Germany

"no information provided"

I10

Intrinsic disorder in the age of protein design

Magnus Kjaergaard

Department of Molecular Biology and Genetics, Aarhus University, Aarhus 8000, Denmark
Danish Research Institute for Translational Neuroscience, Nordic European Molecular Biology Laboratory Partnership for Molecular Medicine, Departments of Biomedicine and Molecular Biology & Genetics, Aarhus University, Aarhus 8000, Denmark
Center for Proteins in Memory, Danish National Research Foundation, Department of Biomedicine, Aarhus University, Aarhus 8000, Denmark

“no information provided”

I11

Physics-Guided Diffusion Sampling for Conformational Variability

Ahmed Selim Üzüm, Turkan Haliloglu

Polymer Research Center, Bogazici University, 34342 Bebek Istanbul, Turkey

Capturing protein conformational variability remains a major challenge for structure-based modeling. Diffusion-based generative models can explore structural ensembles, but recovering

large-scale domain motions is difficult due to the uniform treatment of noise during denoising. Current approaches often rely on modifying multiple sequence alignments (MSAs) or incorporating molecular dynamics (MD) during training, which can fail when evolutionary information is limited or when slow collective motions are poorly sampled.

Here, we introduce GNM CARDS, a diffusion sampling strategy that can be applied to any protein diffusion model with pairwise conditioning, without additional training. GNMCADS enhances structural diversity by modulating noise along collective motions predicted by the Gaussian Network Model (GNM) using the Condition Annealed Diffusion Sampler (CADS), biasing denoising toward physically meaningful fluctuations. Implemented with AlphaFold3 and evaluated on 38 cases, GNMCADS improves the recovery of large-scale conformational changes, including multimeric systems and proteins with limited evolutionary information. The sampled motions agree with anisotropic network model–based Langevin Dynamics (ANM-LD) molecular simulations, supporting the physical realism of the predicted ensembles.

Together, these results demonstrate that integrating coarse-grained physical models with diffusion-based denoising enables efficient exploration of biologically relevant conformational landscapes.

SELECTED TALKS

T01

Generative design of intrinsically disordered proteins based on conditioned protein language models

Laure Carrière, Alexandre Huyghe, Matyas Pajkos, Pau Bernado, Juan Cortés

LAAS-CNRS, Toulouse, France
CBS, Montpellier, France

Intrinsically disordered proteins and regions (IDRs) are central to many biological processes, yet their rational design remains difficult because their function emerges from conformational ensembles rather than fixed structures. Here we present a generative framework for designing disordered protein sequences conditioned on target ensemble descriptors using protein language models (pLMs). We formulate IDR design as the task of generating amino acid sequences that realize specified biophysical properties and implement a transformer encoder–decoder architecture that maps numerical descriptors to sequences. By training models on datasets spanning two orders of magnitude in size, we show that the accurate control of sequence properties is achieved only at large data scale, whereas limited datasets lead to substantially reduced performance despite similar model architectures. These results demonstrate the feasibility of conditioning generative models on ensemble-level properties and identify data availability as the primary limiting factor for data-driven design of intrinsically disordered protein. Our findings support a data-centric paradigm for protein engineering in which expanding annotated datasets may be more critical than developing increasingly complex models.

T02

Variant Effect Predictors in Intrinsically Disordered Regions are Systematically Biased

Zsuzsanna Dosztányi

Eötvös Loránd University, Hungary

“no information provided”

T03

Disordered and not so disordered proteins in bacterial biofilms

Dirk Linke

University of Oslo, Norway

Intrinsically disordered proteins (IDPs) and proteins containing Intrinsically Disordered Regions (IDRs) are essential components of the extracellular matrix in bacterial biofilms, where they often form functional amyloid fibers. These proteins, which lack a fixed three-dimensional structure in their

native state, are critical for biofilm function: cell adhesion, structural stability, and resistance to environmental stress. They often contain high proportions of charged and polar residues, enabling them to remain unfolded until required, preventing premature aggregation within the cell. I will give some examples form literature and our own work, discussing curli and autotransporters and their ability to remain unfolded at critical time points during export, adhesion, and biofilm formation.

T04

IDR-Mediated Nucleosome Engagement by the Pioneer Factor Sox2

Sveinn Bjarnason[1], Jens Nicolai Vilstrup Decker[2], Ásdís H. V. Laufeyjardóttir[1], Eliška Koutná[3], Vaclav Veverka[3], Pétur O. Heidarsson[2]

[1]Department of Biochemistry, Science Institute, University of Iceland.
[2]Department of Biology, University of Copenhagen.
[3]Department of Cell Biology, Faculty of Science, Charles University.

Pioneer transcription factors initiate gene expression programs by engaging chromatin and remodeling nucleosomes. However, the molecular mechanism by which the pioneer factor Sox2 interacts with and remodels nucleosomes remains poorly understood, particularly the role of its intrinsically disordered regions (IDRs).
Here, we use single-molecule FRET (smFRET) and NMR spectroscopy to characterize the conformations of Sox2 in complex with nucleosomes and to examine its binding modes, position-dependent affinities, and the resulting structural effects on the nucleosome.
Our results show that Sox2's IDR undergoes substantial extension upon nucleosome binding, and we identify a specific interaction site within this disordered region. Surprisingly, Sox2 preferentially binds non-specifically to nucleosome linker DNA, even when specific recognition sites are positioned within the nucleosome core. Furthermore, the IDR modulates both Sox2 binding affinity to the core nucleosome and the extent of nucleosome remodeling.
Identification of an interaction site within the Sox2 IDR advances our understanding of pioneer transcription factor mechanisms. Building on this insight, we are now designing improved Sox2 variants with enhanced cellular reprogramming potential.

T05

Achieving predictive all-atom simulations of disordered proteins and their condensates

Miloš T. Ivanović, Nicola Galvanetto, Aritra Chowdhury, Andrea Holla, Valentin von Roten, Daniel Nettels, Robert B. Best, Benjamin Schuler

Department of Biochemistry, University of Zurich, Zurich, Switzerland.

The dynamics of IDPs plays a crucial role in their biological function, allowing them to rapidly respond to binding partners and changes in conditions. We recently showed that, by tightly combining single-molecule FRET experiments with explicit-solvent all-atom simulations [1], we can quantitatively probe IDP dynamics not only in isolation [2,3], but also in small complexes [4,5] and

biomolecular condensates [5,6,7]. Simulations enabled us to rationalize the sequence dependence of IDP ensembles and dynamics across dilute [3] and dense phases [5,6], and to explain how charge-driven interactions can sustain fast conformational fluctuations even in a crowded environment [4,5]—a mechanism cells may broadly exploit in charge-dense compartments such as the nucleus. Despite this success, the quantitative accuracy can be improved, particularly for arginine-rich sequences. Here, we demonstrate how targeted experiments on single amino acids, using vapor-pressure osmometry, can be used to refine residue–residue, residue–ion and ion-io interactions, thereby improving all-atom simulations of IDPs and their condensates and bringing us to a stage where such simulations have predictive power for all protein sequences, over timescales directly relevant to biological function. We are now using these experimentally validated all-atom simulations to train machine-learning–based coarse-grained models of intrinsically disordered proteins and biomolecular condensates. By learning effective, sequence-dependent interactions from large atomistic datasets, this approach opens the way to transferable CG simulations at the length and time scales relevant for non-globular protein function and assembly. Osmometry-guided force-field refinement is therefore not only improving all-atom accuracy, but also providing the high-quality training data needed for predictive ML-driven coarse-grained modeling.

[1] Nettels, Galvanetto, Ivanović, Nüesch, Yang, Schuler, Nat. Rev. Phys. 6, 587–605 (2024).
[2] Nüesch*, Ivanović*, Nettels, Best, Schuler, Biophys. J. 124, 3408–3427 (2025).
[3] Ivanović, Holla, Nüesch, von Roten, Schuler, Best, JACS Au (2026).
[4] Chowdhury et al., Proc. Natl. Acad. Sci. U.S.A. 120, e2304036120 (2023).
[5] Galvanetto*, Ivanović*, Chowdhury, Sottini, Nüesch, Nettels, Best, Schuler, Nature, 619, 876–883 (2023).
[6] Galvanetto, Ivanović, Del Grosso, Chowdhury, Sottini, Nettels, Best, Schuler, Proc. Natl. Acad. Sci. U.S.A. 122, e2424135122 (2025).
[7] Ivanović, Best, Curr. Opin. Struct. Biol. 93, 103101 (2025).

T06

Intrinsically disordered regions in mycobacterial transcription

Lukáš Žídek[1], Dávid Tužinčin[1], Martin Černý[1], Andrea Holla[1], Libor Krásný[2], Jarmila Hnilicová[3,4], Ben Schuler[2]

[1]NCBR, Faculty of Science and CEITEC, Masaryk University, Kamenice 5, CZ-62500 Brno, Czech Republic
[2]Department of Biochemistry, Universität Zürich, Winterthurerstr. 190, CH-8057 Zurich, Switzerland,
[3]Institute of Microbiology, of the Czech Academy of Sciences, Vídeňská 1083, CZ-14220 Prague, Czech Republic
[4]Faculty of Science, Charles University, Viničná 5, CZ-12844 Prague, Czech Republic

The transcription apparatus of bacteria involves several intrinsically disordered regions essential for the correct gene expression. We have investigated structural features and interactions of disordered regions of the Sigma A factor and a novel transcription regulator CrsL of Mycobacterium smegmatis using nuclear magnetic resonance (NMR) and single-molecule Förster resonance energy transfer (sm-FRET).

The N-terminal domain with an amino-acid composition indicating disorder distinguishes mycobacterial Sigma A from the corresponding sigma factors with a well-ordered N-terminal domain, typical for other bacteria. NMR studies of the free Sigma A, including chemical shift analysis and relaxation measurements, revealed a complex hierarchy of ordering ranging from well-ordered domains via regions of reduced flexibility to complete disorder. Studies of Sigma A interactions with RNA polymerase and a transcription factor MoaB2, complementing the cryo-electron microscopy (cryo-EM) data, showed different roles of the N-terminal disordered domain in these complexes. Whereas the N-terminal domain remains disordered in the complex with MoaB2, it exhibits partial ordering in the RNA polymerase holoenzyme, documented by NMR peak broadening and altered distance distribution of the fluorescent labels derived from sm-FRET. Yet, the interacting N-terminal domain remains invisible in cryo-EM electron densities, except for a short helix. These findings help to clarify the so-far open question of the involvement of the N-terminal domain of Sigma A in mycobacterial transcription initiation.
CrsL was identified as an interacting partner of an essential transcription factor CarD. Analysis of the chemical shifts of free CrsL revealed its disordered nature, with two regions of increased helical propensity. Comparison of experimental data with predictions provided by various on-line tools showed the best agreement with PsipreD and ESpritz. NMR analysis of CrsL bound to CarD confirmed formation of two helices in the complex (folding upon binding), identified the binding interface in accordance with the AlphaFold Multimer prediction, and indicated that CrsL binding interferes with the CarD dimerization. The results provided an insight into the molecular mechanism of the regulatory role of CrsL in mycobacterial stress response.
Both presented studies document that combination of bioinformatic tools with various biophysical methods gives us an opportunity to describe physiological roles of intrinsically disordered regions at the molecular level. This work was supported by the Czech Science Foundation grants 22-12023S and 25-16037S.

T07

Switch from a fuzzy to an ordered binding mode in the disordered regulator HigA2 mediates a ratio-sensing autoregulatory gene circuit

Zala Živič, Tadej Medved, Zala Urbič, Remy Loris, Jurij Lah, San Hadži

University of Ljubljana, Slovenia

For globular transcription factors (TFs), the relationship between structure, sequence, and the output of the gene circuit is well established, with the lac operon serving as a paradigm. In contrast, such mechanistic understanding remains limited for TFs containing intrinsically disordered regions (IDRs). Here, we investigate the higBA2 toxin-antitoxin module from the human pathogen Vibrio cholerae, where binding of the HigA2 repressor is strengthened by fuzzy interactions between the HigA2 IDR and its operator. Expression of the cognate HigB2 toxin causes transcriptional de-repression by removing these fuzzy interactions. Upon toxin addition, HigA2 IDR-DNA fuzzy interactions are substituted for the high-affinity, folding-upon-binding interaction with HigB2 toxin, leading to transcriptional de-repression. By encoding two competitive binding modes (ordered/fuzzy) within the same IDR the regulatory circuit gains a new functionality, which operates as a sensor of the cellular toxin:antitoxin ratio. This functionality enables more dynamic circuit response and

homeostatic control over concentration ratio of two proteins. Such anti-cooperative circuit does not require communicating binding sites on the operator, supporting an alternative evolutionary solution of the ratio-sensing regulatory layer in toxin-antitoxin modules.

T08

Ion mobility mass spectrometry unveils global protein conformations in response to conditions that promote and reverse protein phase separation

Christina Robb, Mxolisi Madoda, Thuy Dao, Jakub Ujma, Carlos Castañeda, Rebecca Beveridge

University of Strathclyde, United Kingdom

Phase separation is a process by which biomolecules, particularly proteins and nucleic acids, condense into a dense phase that resembles liquid droplets. Dysregulation of phase separation is implicated in disease, yet the relationship between protein conformational changes and phase separation remains difficult to discern. This is due to the high flexibility and disordered nature of many proteins that phase separate under physiological conditions and their tendency to oligomerize.
In published work, used ion mobility mass spectrometry (IM–MS) to investigate the conformational states of full-length ubiquilin-2 (UBQLN2) protein, phase separation of which is driven by high-salt concentration and reversed by noncovalent interactions with ubiquitin (Ub). We found that UBQLN2 exists as a mixture of monomers and dimers and that increasing salt concentration, which drives LLPS, causes the UBQLN2 dimers to undergo a subtle shift toward extended conformations. UBQLN2 binds to Ub in 2:1 and 2:2 UBQLN2/Ub complexes, which have compact geometries compared to free UBQLN2 dimers. Together, these results suggest that extended conformations of UBQLN2 are correlated with UBQLN2's ability to phase separate.
In more recent work, we strove to understand the mechanism in which differentially linked tetra-ubiquitin chains, specifically those that are linked via lysine-48 (K48) or K63, affect the propensity of UBQLN-2 to phase separate. These results are surprising and interesting, as they do not reflect the expected behaviour that has been observed in the liquid phase. Overall, delineating protein conformations that are implicit in phase separation will greatly increase understanding of the phase separation process, both in normal cell physiology and disease states.

T09

Structure-function studies of iron mineralization in extreme IDP confinement

Raz Zarivach

Department of Life Sciences, Ben-Gurion University of the Negev

The process of biomineralization, which involves the formation and regulation of minerals, relies on specialized proteins. One of the initial proteins involved in iron biomineralization systems is ferritin, which creates a confined space that enables the concentration, nucleation, and growth of iron. Other systems use proteins that often have regions lacking a specific structure, known as IDPs, but these play a vital role in their biomineralization function. To understand the complex process of

cation diffusion, confinement, nucleation, and mineral growth, we developed a new system by creating controlled extreme confinement of IDP inside a ferritin core through its attachment to the ferritin's C-terminal. Using cryo-electron microscopy, we unveiled the dynamic interaction between ferritin, magnetosome-associated magnetite-interacting IDPs, and the nucleation and growth of iron oxide minerals. With this approach, we demonstrated for the first time that extreme confinement can alter how minerals nucleate and influence their diffusion within open ferritin. Notably, we established a new connection between the engineered IDP condensate and its recognition and interaction with mineral particles. These findings highlight the inherent ability of ferritins and disordered proteins to change the chemical environment and undergo conformational shifts in response to environmental cues. Overall, these results provide valuable insights into the role of disordered proteins in controlling mineral nucleation and growth during biomineralization.

T10

Amyloid Proteins Through the Lens of AI Prediction Tools

Alicja W. Wojciechowska[1, 2], Jakub W. Wojciechowski[2], Gert Vriend[3], Malgorzata Kotulska[1,4]

[1]Wroclaw University of Science and Technology, Faculty of Fundamental Problems of Technology, Department of Biomedical Engineering, Wroclaw, Poland
[2]Sano Centre for Computational Medicine, Krakow, Poland
[3]Baco Institute of Protein Science, Mindoro, Philippines
[4]Stellenbosch Institute for Advanced Study, Stellenbosch University, South Africa

While three-dimensional structures of globular and transmembrane forms are available for many amyloid proteins, structures of their fibrillar forms are scarce. Amyloids pose major challenges for both experimental structure determination and computational modelling. Prediction of an unknown amyloid characteristics requires a two-step approach. First, the protein should be assessed regarding its sequence, including general characteristics and prediction of potentially amyloid propensity regions. There are bioinformatics tools developed for this purpose that are based on a range of different methods. The second stage of modelling should address the structure of a protein in its multimeric state. The field of structure modelling has experienced immense progress, resulting in general-purpose AI methods of outstanding accuracy, such as AlphaFold. Despite its success with other classes of proteins, the method is still not effective in modelling amyloids and some other classes of proteins. Subsequently, additional AI-based structure-modeling tools were released, some designed for more specialized applications. Only a few tools have attempted to address the problem of amyloids. None of the methods provide a general working solution. Furthermore, as they were calibrated on pathological amyloids, their applicability to functional amyloids is even further limited. We evaluated the amyloid-modelling performance of the current top modelling software, using different datasets and a few case-study proteins. The performance of AI tools in prediction of amyloid structures seems hindered by low abundance of amyloid structures in the PDB and high prevalence of structure data for their non-fibrillar forms. Most amyloids are predicted as globular multimeric proteins that resemble their homologs seen during training, which tend to be assigned with a higher quality score than fibrillar models that are closer to their true structures. Another underlying cause of the low modelling accuracy could be that amyloid sequence–structure

relationships may not conform very well to Anfinsen's thermodynamic hypothesis, as indicated by the pronounced polymorphism of many pathological amyloids and their very high sensitivity to environmental conditions, which introduces an additional challenge to solving the problem. Encouragingly, however, in a few cases of our study AI tools demonstrated an ability to generalize beyond memorising the PDB and provide a working solution. There remains a clear need for more accurate and reliable prediction approaches. Given the limited number of high-resolution amyloid structures currently available, such developments may benefit from a model system based on an amyloid-like functional protein with a diverse set of variants. An appropriate case study will be presented.

T11

Systematic mutational mapping reveals optimal amyloid formation for RIPK function

Mariano Martín, Benedetta Bolognesi

Institute for Bioengineering of Catalonia (IBEC), The Barcelona Institute of Science and Technology, Barcelona, Spain.

Amyloid formation, typically associated with neurodegeneration, can also serve essential biological functions. RIPK1 and RIPK3 kinases assemble functional amyloids via their RHIM domains to drive necrosome formation and necroptosis — a form of programmed cell death with broad implications in inflammation and disease. The RHIM domain is embedded within a large intrinsically disordered region of these kinases, making the precise sequence determinants of its controlled amyloid nucleation both challenging to predict and critical to understand for deciphering cell death mechanisms and for therapeutic intervention.
We applied deep mutagenesis combined with massively parallel functional assays to systematically map the sequence-function landscape of RIPK1 and RIPK3 RHIM domains. Profiling ~3,000 single-amino acid variants, we quantitatively linked mutational effects on amyloid nucleation to downstream necroptotic signaling outputs. Both kinases rely on a conserved aliphatic tetrad as the structural core for amyloid nucleation yet diverge mechanistically: RIPK3 nucleation is dominated by this core interface, while RIPK1 requires an additional aliphatic surface to achieve efficient nucleation. Cross-species mutational atlases from mouse RIPKs confirm evolutionary conservation of these principles.
A key finding is that functional signaling depends on finely balanced amyloid propensity — variants that either reduce or enhance nucleation equally impair necroptosis. Human population genetics corroborates this, showing such variants are extremely rare, consistent with evolutionary optimization of RHIM domains at a precise "sweet-spot" of amyloid propensity required for signaling. This reveals how evolution has tuned aggregation propensity within a disordered context to enable robust and specific biological function, rather than suppress it.
By linking quantitative mutational landscapes to functional amyloid assembly, this work advances our mechanistic understanding of how amyloid propensity can be calibrated for signaling. Beyond cell death biology, these insights into the sequence rules governing functional amyloids have direct implications for synthetic biology, where the ability to design and engineer amyloid assemblies with predictable, tunable activities remains a key challenge.

T12

TMcluster: Nuanced clustering of protein structures using similarity matrices and dimensionality reduction

Oriol Bárcenas[1,2], Salvador Ventura[1], Ramon Crehuet[2]

[1]Institut de Biotecnologia i de Biomedicina (IBB) and Departament de Bioquímica i Biologia Molecular, Universitat Autònoma de Barcelona, Bellaterra (Barcelona), 08193, Spain.
[2]Institute for Advanced Chemistry of Catalonia (IQAC) of the Spanish Council for Scientific Research (CSIC), Barcelona, Spain.

Intrinsically Disordered Proteins (IDPs) play important roles in biological processes. However, they challenge traditional structural analysis methods because of their vast and heterogeneous conformational landscapes. To address these challenges, we present a computational framework. It quantifies and visualizes conformational diversity across molecular dynamics trajectories and experimentally derived structures.
Using a normalized similarity metric that assesses the similarity between two protein conformations, we compute pairwise structural similarities across all conformations, yielding a square similarity matrix. This matrix effectively captures both intra- and inter-trajectory relationships, as well as comparisons among experimental or predicted structures. We then apply Uniform Manifold Approximation and Projection (UMAP) for dimensionality reduction, yielding a two-dimensional representation in which conformations with high structural similarity cluster together. The subsequent application of density-based clustering (DBSCAN) enables us to identify conformational ensembles and recurring structural motifs without prior assumptions about the number of clusters.
This combined approach offers interpretable visualizations of IDP dynamics and structural heterogeneity, facilitating a deeper understanding of IDP behavior. For example, it has enabled us to interrogate the conformational differences of helical peptides with varying helicity and study the highly disordered IDP luciferase, illustrating the framework's versatile applicability.
We build upon the methodology of Appadurai et al. [1]. Our framework provides a generalizable strategy for mapping and comparing the conformational landscapes of IDPs and other flexible biomolecular systems. It also sets the stage for further exploration of IDP functionalities in different biological contexts.

References:
1. Appadurai, R., Koneru, J. K., Bonomi, M., Robustelli, P. & Srivastava, A. Clustering Heterogeneous Conformational Ensembles of Intrinsically Disordered Proteins with t-Distributed Stochastic Neighbor Embedding. J. Chem. Theory Comput. 19, 4711–4727 (2023).

T13

Local frustration analysis reveals the energetic landscape of intrinsic disorder

Franco L. Simonetti, Edgar P. Chacón, Maria I. Freiberger, Alexander M. Monzon, Diego U. Ferreiro, R. Gonzalo Parra.

Life Sciences Department, Barcelona Supercomputing Center, Barcelona, Spain
Life Sciences Department, Barcelona Supercomputing Center, Barcelona, Spain
Laboratory of Computational and Quantitative Biology, Institut de Biologie Paris Seine, Sorbonne Universite, Paris, France
Department of Biomedical Sciences, University of Padova, Padova, Italy

Introduction. Intrinsically disordered proteins and regions (IDPs/IDRs) lack stable tertiary structure yet play central roles in cellular regulation, signaling, and molecular recognition. A prevailing hypothesis has associated disorder with widespread energetic conflict, suggesting that high frustration underlies structural heterogeneity. Is intrinsic disorder a manifestation of extreme frustration, or does it reflect a fundamentally different energetic regime?
In this work, we have analysed local frustration patterns across structurally heterogeneous ensembles to systematically characterize the energetic organization of IDPs/IDRs. We analyzed NMR ensembles of highly flexible proteins, proteins from the Protein Ensemble Database (PED) which are highly disordered, and ML Boltzman reweighted ensembles of fully disordered systems. We calculated frustration both at the contact and single residue levels, and used radial distribution functions to assess its spatial organization.
Results. While disordered regions show enrichment in highly frustrated interactions relative to ordered ones, they do not contain extreme high frustration. Instead, they exhibited a depletion of minimally frustrated contacts and are dominated by neutral interactions relative to folded cores. Spatial distribution analyses around ordered and disordered residues revealed that disordered regions display lower densities of minimally frustrated interactions across distances, and that high frustration interactions localize in 2nd and 3rd coordinating shells at 2A and 6A distance, respectively. Amino acid pair analysis further demonstrates that hydrophobic interactions strongly contribute to minimal frustration in order-order contacts but are underrepresented in disorder-disorder interactions. Instead, disorder-disorder interactions are dominated by neutral and are enriched in specific pairwise interactions that are highly frustrated between residues containing G, K, P, D, E identities.
Conclusion. We show that intrinsic disorder is not the result of excessive local energetic conflict but the absence of a strongly funneled interaction network. Disordered ensembles are energetically permissive rather than highly frustrated, which could enable conformational plasticity without requiring widespread destabilizing interactions. We propose that IDPs exist at the edge of foldability where a few stabilizing interactions with their partner or small molecules would permit them to fold into their functional structure We expect this work to contribute in further guiding the functional and dynamic characterization of disordered regions and improve our understanding of how disorder supports molecular recognition, versatility, and regulation.

T14

Exon-Informed Multiple Sequence Alignments for Enhanced Coevolution Signal Detection on Intrinsically Disordered Proteins

Diego Javier Zea, Kemal Demirtaş, Turkan Haliloğlu

Université Paris-Saclay, CEA, CNRS, Institute for Integrative Biology of the Cell (I2BC), 91198 Gif-sur-Yvette, France;
Department of Chemical Engineering, Boğaziçi University, 34342 Bebek, Istanbul, Turkey;
Polymer Research Center, Boğaziçi University, 34342 Bebek, Istanbul, Turkey

Background. Deep, diverse multiple sequence alignments (MSAs) are essential for reliable coevolutionary inference, yet proteins containing intrinsically disordered regions (IDRs) remain challenging to align [1]. Exon-informed approaches can improve residue-level homology assignment by leveraging gene structure, but they are only applicable for species with sufficient transcriptomic data. In previous work, we introduced a transcript-aware alignment pipeline based on ThorAxe [2] and a coherence-based quality score for constructing and evaluating IDR-containing MSAs. Here, we build on that foundation to propose Iduna (Intrinsically Disordered Unit Aligner), a workflow for deeper, more diverse transcript-aware MSAs.
Methods. Iduna builds on ThorAxe to generate exon-informed MSAs for proteins with reliable gene- and transcript-level annotations. It runs ThorAxe across a grid of exon-clustering percent identity cutoffs. The optimal threshold is chosen through repeated sequence subsampling. In that process, the alignments produced from random subsets are compared to the full alignment via profile-profile alignments to quantify stability. The most coherent alignment is retained as a seed. This seed is then expanded to increase alignment depth and sequence diversity. MMseqs2 [3] is used to search in UniRef30 for additional homologs. HMMER [4] is then used to train a profile HMM on the seed alignment and to align the newly retrieved homologs back to the seed profile, producing deeper MSAs while preserving the seed's structure.
Iduna's MSAs have been used to compute coevolution matrices using Gaussian DCA [5] through MIToS [6]. These scores have been used to assess downstream functional signal recovery with a Transfer Entropy (TE) integrated Gaussian Network Model (GNM) framework, GNM-TE, benchmarking against EV couplings [7].
Results & Conclusions. Adaptive, coherence-driven selection of the ThorAxe exon-clustering percent identity cutoff allows for the selection of a high-quality MSA. In this work, we have shown that Iduna's HMM-based expansion of those alignments effectively increases alignment depth and diversity. Gaussian DCA coevolution matrices computed from those MSAs have enabled more precise identification of functional sites by GNM-TE on IDRs in comparison to EVcouplings. In the future, we plan to evaluate whether those MSAs can improve structural modeling.

References
[1] Riley et al. PLoS ONE 2023 18, e0288388.
[2] Zea et al. Genome Res 2021 31, 1462–1473.
[3] Steinegger and Söding. Nat Biotechnol 2017 35, 1026–1028.
[4] Eddy. PLoS Comput Biol 2011 7, e1002195.
[5] Baldassi et al. PLoS ONE 2014 9, e92721.
[6] Zea et al. Bioinformatics 2017 33, 564–565.
[7] Hopf et al. Bioinformatics 2019 35, 1582–1584.

T15

Zero-Shot Prediction of Thermodynamic Properties of Proteins

Gábor Erdős, Zsuzsanna Dosztányi

ELTE Protein Dynamics Research Group, Hungary

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.

T16

Machine Learned Coarse-Graining of Non-Globular Proteins

Andrea Guljas, Aldo S. Pasos Trejo, Jacopo Venturin, Cecilia Clementi

Department of Physics, Freie Universität Berlin, Germany

Machine Learning (ML) has transformed the way in which we solve scientific problems in the modern era; in structural biology, ML approaches such as AlphaFold [1] allow predicting a protein's folded structure from its sequence alone, a task that had remained unsolved for more than 50 years. However, the folded structure encodes only part of the information required to understand biological function. For example, intrinsically disordered proteins (IDPs), a class of proteins often involved in signalling, do not have a well-defined native state but populate transient stable states in solution. Despite their importance, experimental and computational characterization of dynamical processes in protein systems is still extremely challenging.
In recent years, we have proposed a machine-learned, chemically transferable, coarse-grained (MLCG) model that can simulate proteins with very low similarity to the ones in the training set [2]. Notably, our model correctly reproduces folding dynamics of proteins starting from an extended

structure, quantitatively estimates changes of free energy upon mutation, and explores the dynamics of disordered proteins [2], all with a speed-up of several orders of magnitude with respect to the corresponding atomistic simulations. Since its initial release [2], we have significantly expanded the scope of our MLCG model by enlarging our training dataset, improving its treatment of long-range interactions, and optimizing the physically-informed correction terms incorporated into our training pipeline.
The goal of our enhanced MLCG model is now the study of challenging biological targets. As a starting point in the field of non-globular proteins, we have optimized our model to reproduce the conformational ensemble of the intrinsically disordered human islet amyloid polypeptide (hIAPP). In aberrant conditions, hIAPP monomers are known to pathologically aggregate and form fibrils that are implicated in the development of Type II diabetes [3]. The efficiency and scalability of MLCG allows us to characterize the disorder-to-order transition and fibril formation of hIAPP, and extrapolate to large systems, previously unattainable using conventional molecular dynamics techniques.

[1] Jumper et al., Highly accurate protein structure prediction with AlphaFold, Nature, 596, 583–589, 2021.
[2] N. E. Charron et al., Navigating protein landscapes with a machine-learned transferable coarse-grained model, Nat. Chem., 17, 1284–1292, 2025.
[3] D. C. Bhowmick, et al., Molecular Mechanisms of Amylin Turnover, Misfolding and Toxicity in the Pancreas, Molecules, 27, 2022

T17

Generating conformational ensembles of multidomain proteins with flexible regions from AlphaFold models using AFflecto

Mátyás Pajkos, Ilinka Clerc, Christophe Zanon, Pau Bernadó, Juan Cortés*

LAAS-CNRS, Université de Toulouse, CNRS, Toulouse, France
LAAS-CNRS, Université de Toulouse, CNRS, Toulouse, France
LAAS-CNRS, Université de Toulouse, CNRS, Toulouse, France
Centre de Biologie Structurale, Université de Montpellier, INSERM, CNRS, Montpellier, France
LAAS-CNRS, Université de Toulouse, CNRS, Toulouse, France

Unlike well-folded domains with stable conformations, proteins containing both folded domains and intrinsically disordered regions (IDRs) are challenging to characterize structurally because they exist as highly dynamic conformational ensembles. While AlphaFold (AF) and related prediction methods provide accurate models for folded domains, they typically represent disordered regions as a single static conformation. Several computational approaches have been developed to generate conformational ensembles of intrinsically disordered proteins, but most focus on isolated disordered segments rather than full proteins containing both ordered and flexible regions. Only a few methods address this broader problem, and these are often computationally demanding or difficult to use. Here, we present AFflecto, an easy-to-use computational framework for generating large

conformational ensembles of multidomain proteins with disordered regions, starting from AF structural models. AFflecto automatically identifies flexible segments and classifies them according to their structural context as tails, linkers, or loops. This region-classification strategy combines the TED domain prediction with AF-derived confidence metrics, including pLDDT scores and the PAE matrix, enabling more accurate identification and classification of flexible regions. These regions are then sampled using efficient stochastic algorithms derived from the FReSa algorithm, which reconstruct protein conformations from fragment-based structural statistics while preserving realistic local geometries. This approach enables efficient exploration of conformational space while maintaining the structural constraints imposed by rigid domains. By explicitly modeling flexible regions in the context of the full protein architecture, AFflecto provides a practical solution for generating physically plausible ensembles of multidomain proteins and other flexible protein systems. The resulting ensembles can be used to interpret experimental data such as SAXS and NMR, or as starting points for further computational studies of protein dynamics. AFflecto is available both as a web server and as a locally runnable version, making it accessible for interactive use as well as larger-scale computational studies.

T18

MAPPIE: Map of Protein-Protein Interaction Embeddings for Functional Discovery

Mert Cihan, Miguel A Andrade-Navarro

Institute of Organismic and Molecular Evolution, Faculty of Biology, Johannes Gutenberg University Mainz, Mainz, 55118, Germany

Protein-protein interactions (PPIs) underlie virtually all cellular functions, and their disruption is central to many diseases. Traditional PPI analysis is largely network-centric, modeling proteins as nodes and interactions as edges. Although effective for graph-level organization, this representation does not directly encode biochemical or functional context and can obscure interaction-specific roles of the same protein across pathways, cellular compartments, tissues, or dynamic cell states. In particular, topology alone may fail to reveal functionally related PPIs that are distant in the interaction graph.
We present MAPPIE (Map of Protein-Protein Interaction Embeddings), a deep learning framework that learns an interaction-level latent space from combined protein sequence embeddings. Each PPI is represented by combining protein language model embeddings of its two partners and projected through an autoencoder to obtain a compact latent representation. The model is trained on 212,927 experimentally supported PPIs, and its architecture and hyperparameters are tuned to maximize local proximity of interactions sharing domain-domain interaction (DDI) annotations, thereby promoting biologically coherent neighborhoods. Importantly, DDI information is used only for parameter optimization and evaluation, not as input features, enabling MAPPIE to generalize to interactions lacking domain annotation.
In the resulting map of PPIs, each interaction occupies a position whose local neighborhood captures shared functional properties independent of graph topology. Functional interpretation of a PPI is achieved by enrichment analysis of annotated terms among its neighboring interactions, including Gene Ontology categories, disease associations, and domain interaction annotations. Candidate or newly predicted PPIs can be projected into this space and assigned functional context

based on the enriched annotations of nearby interactions, enabling interaction-level function inference without reliance on direct network connectivity.
By decoupling functional organization from strict network structure and leveraging neighborhood enrichment in embedding space, MAPPIE provides an interaction-centric framework to understand PPI function. This spatial representation supports hypothesis generation, candidate prioritization, and mechanistic interpretation of known and predicted PPIs, offering a practical bridge between sequence-derived protein representations and biologically meaningful interaction landscapes.

T19

Large-scale proteomics and Deep learning for the generation of protein-protein interaction (PPI) maps at atomic resolution

Jinmei GAO[1,2,3], Caroline SANCHO[1,2], Chloé ANTOINE[3], Vincent COLLURA[3], Jessica ANDREANI[1,2], Jean-Christophe RAIN[3], Diego Javier ZEA[1,2,4] and Raphaël GUEROIS[1,2]

[1]Centre CEA Paris-Saclay, 91190, Gif-sur-Yvette, France
[2]I2BC, 1 Avenue de la Terrasse, 91190, Gif-sur-Yvette, France
[3]Hybrigenics Services, 1 rue Pierre Fontaine, 91000, Evry-Courcouronnes, France
[4]Université Paris-Saclay, 3 rue Joliot Curie, 91190 Gif-sur-Yvette, France

The rapid progress of artificial intelligence has made it possible to predict the three-dimensional struc-ture of most proteins at atomic resolution [1]. A fundamental challenge is to map as comprehensively as possible the structures of multiprotein assemblies that underlie cellular function [2]. This objective comes up against several constraints, such as limitation of AlphaFold in predicting interaction motifs involving disordered regions, whose performance has proven to be more limited than in other catego-ries of interactions [3], and the type of evolutionary information required to identify co-evolution sig-nals, which are much weaker at complex interfaces than in monomers. We explore protein interaction maps obtained from different proteomic approaches, such as Affinity Purification Mass Spectrometry [4] and Yeast Two-Hybrid screens from Hybrigenics Services company [5] using deep-learning methods. We tested sequence segmentation strategies and different ways of integrating evolutionary infor-mation to improve the sensitivity of AlphaFold predictions. While challenges remain for achieving high specificity binding prediction for disordered interaction motifs, we are evaluating potential remedial strategies, including the use of evolutionary information from closely related homologs.

References

1. Jumper, J., Evans, R., Pritzel, A. et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021; 596:583–589. https://doi.org/10.1038/s41586-021-03819-2
2. Evans, R., O'Neill, M., Pritzel, A., Antropova, N., Senior, A., Green, T., Žídek, A., Bates, R., Blackwell, S., Yim, J. et al. Protein complex prediction with AlphaFold-Multimer. BioRxiv. 2021; https://doi.org/10.1101/2021.10.04.463034
3. Bret, H., Gao, J., Zea, D.J. et al. From interaction networks to interfaces, scanning intrinsically disordered regions using AlphaFold2. Nat Commun. 2024; 15: 597. https://doi.org/10.1038/s41467-023-44288-7

4. Novikov, N.M., Gao, J., Fokin, A.I. et al. NHSL3 controls single and collective cell migration through two distinct mechanisms. Nat Commun. 2025; 16: 205. https://doi.org/10.1038/s41467-024-55647-3
5. Gao J. et al., in preparation

T20

WormLink: A Tool for Engineering Synthetic Spacer Regions

Laura K. Gherman[1,2], Michael J. Plevin[1,2]

[1]Department of Biology, University of York, York,
[2]York Structural Biology Laboratory, University of York, York

Intrinsically Disordered Proteins (IDPs) frequently drive Liquid-Liquid Phase Separation (LLPS) to form functional organelles. A prominent example is the pyrenoid, a phase-separated condensate found in hornworts and algae, responsible for 30% of the global CO2 fixation. Pyrenoids are formed via interactions between Rubisco and tandem repeat proteins called "Linkers", composed of short helical Rubisco Binding Motifs (RBMs) and intrinsically unstructured "spacer" regions. Engineering pyrenoid-like functionality into crops offers a promising strategy for enhancing carbon fixation and yield. Although the biophysical rules of linker architecture and Rubisco-binding are becoming better understood, and we can modulate linker protein properties via sequence engineering, natural algal proteins fail to effectively phase-separate crop-plant Rubisco. To overcome these limitations, we are employing de novo design strategies to create tailored synthetic linker proteins.
While tools exist for developing synthetic RBMs (e.g. phage display, computational design, etc), a major challenge lies in the design of IDRs that optimally link binding domains to generate a multivalent protein with the desired binding properties. Common approaches such as glycine-serine repeats, or IDRs from naturally occurring proteins have severe practical limitations ranging from reduced propensity for phase-separation to unfavourable ensemble conformations. We have developed WormLink, a new computational tool to support the design of IDRs in multidomain binding proteins. The tool takes as input a 3D structure with binding motifs docked at binding sites and calculates optimal sequence lengths for IDRs that connect different pairs of motifs. WormLink couples with tools that generate sequences with appropriate ensemble properties to generate synthetic IDRs for linking binding motifs and generating multidomain binders. We will show how WormLink can support multivalent binder design using examples from Rubisco-Linker complexes.

SELECTED FLASH POSTERS

F01

Hierarchical Deep Learning for Tandem Repeat Proteins Classification

Nevena Ćirić, Jovana Kovačević

Faculty of Mathematics, University of Belgrade, Serbia

Repeat proteins are a widespread class of mostly non-globular proteins containing repetitive subsequences, a so-called repeat units that often occur in tandem arrangements when observed in 3D structure of the protein. These tandem repeats are considerably diverse, ranging from the repetition of a single amino acid to domains of 100 or more residues.
Improvement of the methods for identification of protein tandem repeats and subsequently the increasing number of the known proteins containing repetitive elements necessitates their classification to facilitate further understanding of their sequence-structure-function relationships. According to Kajava's classification scheme based on the repeat unit's length, general structural arrangement and mode of interaction between the repeat units [1], tandem repeats are classified into five main classes and further divided into subclasses that reflect repeat unit topology, differing in secondary structure arrangement and/or overall structure within the repeat.
The classical approach to obtain the (sub)class assignment for a newly identified tandem repeat is by simply transferring this information from the "master" repeat unit, that is the repeat unit from database of predetermined tandem repeats with associated (sub)classes found to be most similar to the newly identified tandem repeat. This procedure usually implies some kind of structural search algorithm in order to assign master repeat unit, which further implies known tertiary structure of the protein with newly discovered tandem repeat. With intention to tackle the problem of analyzing proteins with unknown 3D structure and facilitate classification of tandem repeats by using only the sequence information, as well as to explore sequence-structure relationship between repeat units sequence and structural characteristics of their corresponding (sub)class, here we propose neural network based model for classification of tandem repeats based on the multiple sequence alignment of its units sequences. Additionally, this model can be further utilized to create an end-to-end pipeline for identification and classification of tandem repeats only from sequence information.

F02

Exploring Binding Signals in Intrinsically Disordered Regions Through AlphaFold Embeddings

Julie DANIEL, Diego Javier ZEA

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

Background. Identifying binding sites located within intrinsically disordered regions (IDRs) remains challenging. Recent benchmarking efforts such as CAID3 [1] have evaluated methods addressing this problem. Most of the highest-scoring methods were using embeddings from protein language

models. However, current performance remains limited, with reported AUC values around 0.6, suggesting that significant improvements are still possible.
Methods. In this work, we investigate whether protein embeddings from AlphaFold2 [2] can capture signals that distinguish binding residues from non-binding residues within IDRs. Such embeddings are interesting for this task, as they capture both evolutionary and structural information. To explore this question, we assembled a dataset of several thousand proteins derived from DisProt [3]. Using this dataset, we performed exploratory analyses on residue-level and pair-level embeddings generated by AlphaFold2.
Results & Conclusions. Preliminary analyses in reduced-dimensional embedding spaces reveal differences between disordered binding regions, non-binding disordered regions, and structured regions. Suggesting that these embeddings may encode non-linear signals associated with binding propensity within intrinsically disordered regions. Interestingly, similar dispersion patterns of the embedding space are observed across proteins with diverse characteristics. These results indicate that AlphaFold-derived embeddings may provide informative features for downstream tasks aimed at identifying binding regions in IDRs and motivate the development of predictive machine learning models.

References :
[1]Critical Assessment of Protein Intrinsic Disorder Round 3 - Predicting Disorder in the Era of Protein Language Models, M. Mehdiabadi et al., Proteins: Structure, Function, and Bioinformatics 2026
[2] Highly accurate protein structure prediction with AlphaFold — John Jumper et al., Nature, 2021
[3]DisProt — DisProt: intrinsic protein disorder annotation in 2020, Nucleic Acids Research, 2021

F03

Homology-based analysis of protein structures and sequences with GTalign-web, COMER and PPI3D web servers

Justas Dapkūnas, Mindaugas Margelevičius, Česlovas Venclovas

Institute of Biotechnology, Life Sciences Center, Vilnius University

In recent years, the progress in protein structure prediction has been impressive, and the structures are readily available for almost all known proteins. However, to answer all questions on the biological function of a protein of interest, knowing only its single structure is not enough. A lot of useful information can be gathered from homology-based inference. As a result, tools to identify homologous proteins are still indispensable in molecular biology research. Here, we present 3 user-friendly web servers for protein analysis. GTalign-web (https://bioinformatics.lt/gtalign) performs structure-based search for biomolecular structures and allows searching among both the experimental structures in the PDB database and the predicted structures in the AlphaFold Database. The newest version of GTalign-web accepts not only individual protein chains, but also structures of biomolecular complexes involving proteins, RNA and DNA. By enabling rapid and sensitive structure search GTalign-web provides a useful resource for structure analysis and functional annotations. Yet, sometimes structure is not available for a protein. In these cases, the COMER web server (https://bioinformatics.lt/comer) can be of use. COMER conducts a protein

sequence homology search based on profile-profile alignments. Given the sequences of the proteins, it enables their annotation at structure, family and function levels by identifying close and remote homologs in numerous protein databases, such as the Protein Data Bank, Pfam, SwissProt and others. The PPI3D web server (https://bioinformatics.lt/ppi3d) is dedicated to the homology-based analysis of protein-protein, protein-peptide and protein-nucleic acid interactions. These tools enable better interpretation of the protein structure models and the analysis of proteins in the context of the available data for their homologs. Therefore, we believe that our web servers might be useful for the research on both globular and non-globular proteins.

F04

Detecting Non-Canonical Coiled-Coil Domains

Stanisław Dunin-Horkawicz, Kamil Kaminski

Department of Protein Evolution, Max Planck Institute for Biology, Tübingen, Max-Planck-Ring 5, 72076 Tübingen, Germany;
Institute of Evolutionary Biology, Faculty of Biology, University of Warsaw, Zwirki i Wigury 101, 02-089, Warsaw, Poland

Coiled coils form a large and diverse class of non-globular fibrous proteins. They are formed by alpha helices assembling into typically regular bundles and other higher-order assemblies, such as sheets or tubes. Since the development of sequence-based (COILS) and structure-based (SOCKET) coiled-coil detection tools in the 1990s and early 2000s, the field saw relatively limited progress until the widespread application of deep learning in bioinformatics. This led to the development of sequence-based tools such as DeepCoil, CoCoNat, and CoCoPRED, offering substantially improved accuracy and speed. These tools, along with the classical ones, were developed mostly with typical, canonical coiled coils in mind. There is, however, a large group of non-canonical coiled coils that are frequently missed by current approaches, despite their often important functional roles. In this presentation, I will discuss our current work on developing new methods for coiled-coil detection beyond what is considered canonical.

F05

LCRPlatform Update: Similarity-Based Annotation Transfer and Extended Functional Analysis of Low-Complexity Regions in Proteins

Syliwa Morawiec[1], Patryk Jarnot[1], Joanna Ziemska-Legięcka[2], Urszula Ziemska[3], Marcin Grynberg[2], Aleksandra Gruca[1]

[1]Silesian University of Technology, Gliwice, Poland
[2]Institute of Biochemistry and Biophysics, PAS, Warsaw, Poland
[3]University of Warsaw, Warsaw, Poland

We present the latest update to LCRPlatform, a comprehensive metaserver for the precise identification, functional annotation, and evolutionary analysis of low-complexity regions (LCRs) in

protein sequences. To ensure robust detection, the platform integrates five benchmark algorithms — SEG, CATS, fLPS, SIMPLE, and GBSC — enabling researchers to perform multi-method LCR searches through a single, customizable interface.
The updated server streamlines the workflow by accepting protein IDs or FASTA-formatted sequences via direct input or file upload. A key new feature is the ability to upload custom sequences absent in public databases, enabling the investigation of functions of LCRs detected within the target sequences. Results are organized across five specialized analytical modules:
- Annotations: Integrates functional metadata from the LCRAnnotationsDB, which synthesizes curated data from 12 external repositories (including UniProtKB/Swiss-Prot and InterPro) via an interactive feature viewer.
- Domains: Leverages NCBI's CDSearch to map identified LCRs against CDD, Pfam, SMART, COG, TIGRFAM, and KOG databases, with user-defined coverage thresholds.
GBSC: Supports discovery of clusters of compositionally similar LCRs and facilitates the transfer of functional annotations between sequences.
- Tree View: Visualizes phylogenetic relationships among species represented in GBSC clusters, offering an overview of evolutionary context.
LCRPlatform distinguishes itself by shifting the analytical focus from full protein sequences to the specific functional context of individual LCRs. The new version enables annotation transfer from LCRAnnotationsDB to user-provided sequences by exploiting compositional similarity within GBSC clusters. Specifically, the platform detects LCRs in input sequences and applies GBSC to identify clusters sharing sequence patterns at user-defined coverage levels. The GBSC score then assesses the similarity between a query LCR and the cluster's STR pattern, supporting similarity-based hypothesis generation for previously unannotated regions
This granular approach facilitates the generation of high-confidence hypotheses regarding the roles of previously unannotated regions. With our recent server infrastructure update, all processing is conducted locally, providing significantly faster throughput and more reliable data synthesis than previous iterations. By consolidating these complex analytical steps, LCRPlatform remains a unique and essential resource for functional LCT protein characterization. The server is available under the following link: https://lcrplatform.lcr-lab.org/

F06

Changes in the Dynamics of Microtubule-Associated Protein MAP2c upon Binding to the Regulatory Subunit of cAMP-Dependent Protein Kinase

Alice Laníková[1], Viktor Bartošík[1], Vojtěch Vařečka[1], Séverine Jansen[2], Petr Padrta[2], Martin Blackledge[3], Lukáš Žídek[1], Fabien Ferrage[4], Pavel Kadeřávek[1]

[1]National Centre for Biomolecular Research, Faculty of Science and Central European Institute of Technology, Masaryk University, Kamenice 5, 625 00, Brno, Czech Republic
[2]Central European Institute of Technology, Masaryk University, Kamenice 5, 625 00, Brno, Czech Republic
[3]Institut de Biologie Structurale (IBS), CEA, CNRS, University Grenoble Alpes, Grenoble 38044, France
[4]Laboratoire des Biomolécules, LBM, Département de chimie, École normale supérieure, PSL University, Sorbonne Université, CNRS, 75005 Paris, France

The internal motions of proteins are crucial because they enable proteins to carry out their functions within living organisms. Conformational plasticity and dynamics are critical, particularly for intrinsically disordered proteins (IDPs). This feature enables IDPs to adopt different conformations and switch between them. As a result, IDPs can bind to multiple ligand targets and play their typical roles in signaling pathways, the regulation of cellular molecular machinery, etc. In this project, we investigated the dynamics of the microtubule-associated protein MAP2c. MAP2c belongs to the microtubule-associated protein (MAP) family of proteins. These proteins play an important role in regulating microtubule stability and dynamics. Along with Tau proteins, MAPs are brain-specific proteins; however, they differ in localization. Tau proteins are present in axons, while MAPs are found in the dendrites of neurons. MAP2c is the smallest MAP isoform and is characterized as an IDP. Its proper function is controlled by posttranslational modification. Abnormal MAP2c phosphorylation has been linked to serious psychiatric disorders such as depression and schizophrenia. Since phosphorylation regulates its function, we focused on characterizing changes in its dynamics associated with binding to the regulatory domain of the cAMP-dependent protein kinase in the N-terminal region of MAP2c. To obtain detailed information about the motions of bound and free MAP2c with atomistic resolution, we employed NMR methods that do not require modification of the molecule except for isotope labeling, which keeps the molecule in its native state. We investigated motions using an analysis of high-field NMR relaxation rates together with high-resolution relaxometry rates. This unconventional approach allowed us to reveal details about slow motions within the nanosecond-picosecond timescale. These motions are typically associated with intrinsically disordered proteins (IDPs) and are challenging to analyze using standard NMR methods developed to study biomolecular motions. Analyzing the experimental data in combination with molecular dynamics simulations provided insight into the motion distribution pattern along the MAP2c backbone and how this pattern changes when MAP2c binds to the regulatory domain of the cAMP-dependent protein kinase. Notably, significant changes were detected far from the binding site, suggesting a substantial redistribution of motions in MAP2c upon binding.

F07

Characteristic Positional Distributions of Functional Classes of Protein Regions with Compositional Bias, Structural Disorder, and Low Complexity

A Gruca[1], D P Kreil[2]

[1]Silesian University of Technology, Poland
[2]Boku University Vienna, Austria

The biological importance and the functional roles of compositionally biased, structurally disordered, and low-complexity sequence regions are increasingly recognized. Most reports, however, remain anecdotal or focus on the characterization of a particular method for delineating a region of interest due to its unusually low variance in amino acid composition, lack of structural order, or low information complexity.
We here present a recent study that, for the first time, systematically compares functional associations of proteins featuring regions identified by a range of complementary approaches to

delineating compositional bias, structural disorder, and low information complexity in amino acid sequences.
We extend these analyses comprising functional annotations of proteins from ontologies such as the GeneOntology, KEGG, etc., by exploiting the unique resource of LCRAnnotationsDB database, which combines twelve complementary sources for curated annotations of the functions of low-complexity regions themselves rather than the proteins hosting them. We can thus present, for the first time, a systematic analysis of functions of these low-complexity regions proper.
Finally, for the first time, we move beyond early studies reporting a tendency of some low-complexity regions towards protein terminal sequences by a full-scale systematic quantitative study of function specific positional distributions of the regions of interest relative to the overall protein sequence, covering and contrasting regions of interest due to their unusually low variances in amino acid composition, lack of structural order, or low information complexity.

F08

Modeling intrinsically disordered regions from AlphaFold2 to AlphaFold3

Mahta Mehdiabadi, Silvio C.E. Tosatto, Damiano Piovesan

University of Padova, University of Padova, University of Padova

AlphaFold2 has demonstrated a remarkable success in predicting the structures of globular proteins and folded domains with near-experimental accuracy. However, it typically represents intrinsically disordered regions (IDRs), protein segments that lack a stable 3D structure under physiological conditions, as long extended loops that appear to float around the structured core. While AlphaFold2's static prediction cannot capture the conformational heterogeneity and the dynamic nature of IDRs, it performs well in predicting IDRs from sequence. AlphaFold3 introduces significant architectural and training modifications over its predecessor, including the use of cross-distillation aimed at reducing structural hallucinations in disordered regions. In this study, we look into how these models differ in representing IDRs. We evaluate the performance of AlphaFold3 and AlphaFold2 on disorder prediction, using the CAID3 benchmark. Our analysis shows that AlphaFold3 does not outperform AlphaFold2 in this benchmark. We observe that solvent accessibility remains a robust and consistent proxy for predicting intrinsic disorder across both models. However, changes in the predicted secondary structure content and pLDDT scores lead to different interpretations of disorder. Overall, our findings suggest that AlphaFold2 remains the preferred choice for identifying intrinsically disordered regions, as it avoids structural hallucinations while providing predictions comparable to those of AlphaFold3.

F09

A comparison of de novo designed amyloids with naturally occurring ones

Anahit Khachatryan[1], Hrachya Aslanyan[1], Rafayel Petrosyan[1,2]

[1]Institute of Physics, Yerevan State University, 0025 Yerevan, Armenia_Author1

[2]Zaven P. and Sonia Akian College of Science and Engineering, American University of Armenia, 0019 Yerevan, Armenia

While almost any protein can be forced to form an amyloid fibril under specific laboratory conditions, naturally occurring amyloid-forming sequences have distinct features in their amino acid sequences that differentiate them from the sequences of globular proteins. Out of the hundreds of millions of known protein sequences, only about one thousand are known to form amyloids under natural conditions. To achieve a more robust statistical comparison, we need a larger number of amyloid-forming sequences. Here, we used two different approaches for producing de novo amyloids. First, we fine-tuned protein language models (pLMs) (such as ProGen2 and ProtGPT2) using the naturally occurring amyloid-forming sequences. Secondly, we utilized the diffusion-based model, Chroma, providing it with the backbone structures of known amyloids from the PDB to generate novel amino acid sequences capable of forming similar amyloid fibril structures. Next, we generated thousands of amyloid-forming sequences with these two distinct methods and implemented a quality control module to evaluate the amyloidogenic potential of each generated sequence. Sequences that passed the quality control filters were included in a comparative analysis between those generated by the diffusion-based model, the pLM, and those that occur naturally.

F10

Sequence-Based Discovery of Modulators for Disordered Proteins

Milan Senćanski

Laboratory for Bioinformatics and Computational Chemistry, Institute of Nuclear Sciences VINCA, National Institute of the Republic of Serbia, Mike Petrovica Alasa 12-14, 11351 Belgrade, Serbia

We developed a computational framework that analyses protein and small molecule sequences by converting them into frequency space. Unlike traditional methods that depend on the 3D structure of the binding domains and sites, our approach works for proteins lacking a defined structure. This makes it applicable for studying IDPs, which are often considered difficult to target with drugs. To test the method, we applied it to the Tau protein, an IDP linked to neurodegenerative diseases. The analysis revealed several approved drugs from DrugBank and natural compounds whose sequences matched frequencies associated with Alzheimer's disease-related molecular recognition. Importantly, these compounds have been reported to be active in vivo, suggesting that our approach can help prioritise modulators with real therapeutic potential, even though they are not directly related to Tau protein modulation. Overall, our findings show that analysing the long-range interaction frequency space of proteins and small molecules is applicable to ligand recognition in IDPs. This strategy complements standard structure-based methods, offering a promising approach for early discovery of IDP modulators.

F11

The LLPS–Amyloid landscape of S100A and 14-3-3 proteins families

Saulė Rapalytė, Viktorija Karalkevičiūtė, Ieva Baronaitė, Orestas Antipas Pelivanov, Arijus Jaloveckas, Agnė Janonienė, Vytautas Smirnovas, Mantas Žiaunys, Darius Šulskis

Life Sciences Center, Vilnius University, Lithuania

S100A and 14-3-3 are two large protein families that perform a plethora of functions in the human body. S100A proteins are primarily associated with inflammation, acting as extracellular signalling molecules that trigger cascading inflammatory responses [1]. In contrast, 14-3-3 proteins regulate the phosphorylation of thousands of proteins by scaffolding kinases and their substrates, and have been implicated in neurodegeneration through modulation of protein–protein interactions and aggregation [2].
Protein aggregation is one of the first hallmarks of Alzheimer's and Parkinson's, which manifests through the formation of amyloid fibrils. This can happen due to impaired proteostasis, mutations, or during liquid-liquid phase separation (LLPS), in which proteins condense into membraneless organelles. In this work, we present a combination of biophysical and biochemical methods to decipher the phase separation and amyloid-forming properties of the 14-3-3 and S100A families.
We selected all isoforms of 14-3-3 family and 5 members (A1, A4, A8, A9, A12) of S100A family for the experiments. Despite being folded, 14-3-3 and S100 proteins were capable of LLPS or fibril formation. In the S100A family, the most amyloidegenic proteins were A8 and A9, whereas A1 accelerated other protein aggregation, and A12 suppressed in the presence of calcium ions. Within the 14-3-3 family, the ε isoform formed worm-like fibrils. Members of both families were capable of undergoing LLPS under molecular crowding conditions, albeit with varying propensity.
Collectively, this work expands the understanding of LLPS and amyloid formation of both families and identifies novel targets for therapeutic intervention in protein aggregation disorders.

[1] G. Sreejit, M. C. Flynn, M. Patil, P. Krishnamurthy, A. J. Murphy, and P. R. Nagareddy, 'S100 family proteins in inflammation and beyond', in Advances in Clinical Chemistry, vol. 98, Elsevier, 2020, pp. 173–231. doi: 10.1016/bs.acc.2020.02.006.
[2] X. Huang, Z. Zheng, Y. Wu, M. Gao, Z. Su, and Y. Huang, '14-3-3 Proteins are Potential Regulators of Liquid–Liquid Phase Separation', Cell Biochem. Biophys., vol. 80, no. 2, pp. 277–293, Jun. 2022, doi: 10.1007/s12013-022-01067-3.

F12

High-Resolution NMR Characterization of the Molecular Interplay between α-Synuclein, Fasudil, and Calcium Ions

Filippo Turchi[1], Taylan Turan[2], Marco Shiavina[1], Giuseppe Brancato[2], Isabella Caterina Felli[1], Roberta Pierattelli[1]

[1]Department of Chemistry "Ugo Schiff" and CERM, University of Florence, Sesto Fiorentino, Italy
[2]Scuola Normale Superiore, Pisa, Italy

Human α-synuclein, an intrinsically disordered protein (IDP) implicated in several neurodegenerative diseases including Parkinson's disease, consists of 140 amino acids organized into three main

regions: a positively charged N-terminal tail, a hydrophobic central region, and a negatively charged C-terminal region. Its dynamic C-terminal region mediates interactions with small molecules and metal ions. Nuclear Magnetic Resonance (NMR) spectroscopy is a powerful technique for characterizing the interactions between IDPs and various binding partners at atomic resolution under physiological conditions.

In this study, we investigate the interaction between α-syn and fasudil, a compound known to bind to the C-terminal part of α-syn and delay the formation of its toxic aggregates. We also examine the modulatory role of calcium ions in this interaction. We employed high-resolution NMR spectroscopy, utilizing both 13C and 1H detected experiments and molecular dynamics (MD) simulations, to characterize the interactions between a C-terminal α-syn construct, the small molecule fasudil, and calcium ions.

Our study focuses on the side chains of aspartic acid, glutamic acid, and tyrosine residues, which play a key role in the binding of fasudil and calcium ions. We analyse their behaviours in the presence and absence of calcium ions to elucidate the influence of calcium on this process. Side-chain-resolved spectra indicate distinct driving forces for fasudil and calcium ion binding. MD simulations reveal that Ca2+ modifies the local electrostatic environment, decreasing the frequency of fasudil interactions through electrostatic screening and steric effects. Despite this, fasudil retains dynamic, reversible contacts with key tyrosine residues. Overall, the exposed α-synuclein conformations allow for simultaneous, ligand-specific interactions, highlighting side-chain hotspots governing binding in Ca2+rich conditions.

POSTERS

P01

Prediction of Antimicrobial Peptide–Amyloid-β Interactions Using Transformer Language Models and Graph Attention Networks

Hanan Abdul Kareem, Zuzana Gazova

Department of Biophysics, Institute of Experimental Physics, Slovak Academy of Sciences, Kosice, Slovakia

The aggregation of amyloid-β (Aβ) peptides is a critical molecular hallmark of Alzheimer's disease, leading to the formation of toxic amyloid oligomers and fibrillar plaques that contribute to neuronal degeneration. Antimicrobial peptides (AMPs) are characterized by amphipathic structures, allowing their interference with amyloidogenic proteins. Experimental studies have shown that certain AMPs, including lactoferrin, α-defensin, and cystatins, can bind Aβ peptides and inhibit oligomerization, fibrillization, and plaque formation while reducing amyloid-induced toxicity and inflammation (Bruno et al., 2022).
In our study, we used a computational framework to explore the interactions between AMPs and Aβ peptide, highlighting the potential of AMPs as modulators of amyloid aggregation. A dataset of approximately 300 AMPs, each with sequence lengths up to 25 amino acids, was docked against Aβ fibril structures using AutoDock, generating potential binding conformations, and PepATTRACT was used for reference comparisons.
We encoded Aβ aggregate and AMP sequences with pretrained transformer-based protein language models, such as ProtBERT and ESM, to generate contextual sequence embeddings that capture evolutionary and biochemical information. These embeddings were combined with structural and residue-level descriptors derived from the docked complexes, including one-hot amino acid encodings, residue positional indices, B-factor values, solvent-accessible surface areas, and Shannon entropies. Graph representations of the Aβ structure were constructed, with residues as nodes and edges representing spatial proximity. A Graph Attention Network (GAT) was then employed to model residue-level structural relationships and identify interaction patterns between AMPs and Aβ. At the same time, an Equivariant Graph Neural Network (EGNN) incorporated three-dimensional spatial coordinates from docked complexes to enhance the structural interpretation of peptide binding. Additional biophysical features, including hydrophobicity, intrinsic disorder, amyloid propensity, and peptide toxicity, were analyzed to characterize AMP behavior in relation to amyloid binding. Residue-level interaction maps facilitated the identification of potential binding hotspots within aggregation-prone regions of Aβ. The modelling framework successfully identified critical binding residues implicated in modulating Aβ aggregation. Attention maps highlighted potential binding hotspots within aggregation-prone regions of Aβ, providing valuable mechanistic insights. Overall, this framework provides a scalable computational strategy for screening AMPs, aiming to identify candidates that can interact with Aβ and potentially modulate amyloid aggregation.

Keywords: Antimicrobial peptides (AMPs), Amyloid-β, Transformer protein language models; Graph Attention Networks (GAT), Equivariant Graph Neural Networks (EGNN), residue-level interaction prediction, computational peptide screening

P02

An Integrative structure-based approach to link diseases associated mutations to pathogenicity

Matteo Arnaudi[1], Mattia Utichi[1], Karolina Krzesińska[1], Matteo Tiberti[2], Elena Papaleo[2]

[1]Technical University of Denmark,
[2]Danish Cancer Institute, Danish Cancer Institute

Pathogenicity annotations of germline mutations obtained through NGS techniques represent a significant advancement for healthcare and personalized medicine, in particular for disease treatment, genetic counseling, prevention, and monitoring. However, most part of them are Variants of Unknown Significance (VUS), which hinder patient care and therapeutic development due to challenges in determining their pathogenicity. Advances in Artificial Intelligence have led to the development of highly accurate pathogenicity classifiers for VUS classification. Nevertheless, these classifiers provide only probability scores, lacking the molecular mechanisms underlying mutation pathogenicity and limiting their clinical utility. This emphasizes the need for comprehensive interpretations essential for clinicians and molecular biologists.
To address this gap, we are developing a structure-based framework that integrates computational and experimental methodologies to predict and validate the effect of missense germline mutations associated with various diseases on protein structural features. We retrieve target protein structures through AlphaFold and/or the Protein-DataBank (PDB) database. We trim predicted structures using disorder propensity information obtained through a combined approach involving AF pLDDT score and CAID. We employ experimental databases of manually curated complexes for the selection of potential interactors and protein sequence analysis for the Short Linear Motifs (SLiMs) annotations. We combine the information to generate the complexes through AF-multimer in the absence of experimental structures.
We obtain structural ensembles through Molecular Dynamics simulations, to include protein-flexibility and improve prediction accuracy. We integrate free and binding energy calculations, sequence analysis, and Structure-Based-Statistical-Mechanical-Models to predict the effects of mutation collected from three different databases on protein stability, protein-protein interactions, changes in post-translational modification sites, and allosteric changes on protein pockets. These predictions guide experimental validation through a platform currently under development that integrates biochemical and microscopy techniques.
In our preliminary experiments we validated the destabilizing effect of the P263H variant on ARID3A, a transcriptional factor highly mutated in leukemia, and its interaction with the DNA damage sensor TP3BP1 through CHX assay and microscopy techniques. Additionally, we validated sequence analysis predictions identifying a SH3-mediated interaction between Lyn tyrosine kinase and P14arf, suggesting further research avenues in B-cell cellular pathways. These results, along with the computational workflow, not only provide a robust method for investigating mutation effects on protein structural features but also demonstrate potential for exploring intrinsically disordered proteins (IDPs), particularly SLiM interactions.
Overall, this framework supports molecular biologists in both basic and translational research, aiding in the understanding of disease mechanisms and the development of new drugs and treatments.

P03

Interpreting missense variation in intrinsically disordered proteins

Nazareth D J Robles, Silvio C E Tosatto, Maria Cristina Aspromonte

Department of Biomedical Sciences, University of Padova, Italy

Interpreting missense variants in intrinsically disordered proteins (IDPs) remains a major challenge in human genetics. Unlike structured proteins, IDPs lack a stable three-dimensional conformation and instead exist as dynamic ensembles of conformations. This structural flexibility complicates the functional interpretation of amino acid substitutions and limits the applicability of conventional structure-based prediction approaches. Consequently, variants located within intrinsically disordered regions (IDRs) are frequently classified as variants of uncertain significance (VUS), reflecting the limited availability of suitable analytical frameworks rather than the absence of functional impact. Accurate annotation of protein disorder is therefore essential for understanding how genetic variation affects these proteins. In this work, we leverage two major resources dedicated to the characterization of intrinsic disorder: DisProt, which provides experimentally validated and manually curated annotations of disordered regions, and MobiDB, which integrates curated evidence, structural data, and multiple disorder prediction methods. By combining these complementary resources, we constructed a comprehensive structural annotation framework capturing ordered regions, intrinsically disordered segments, and structural transition regions across a set of proteins associated with human disease. To quantify regional tolerance to amino acid substitutions, we recalculated the Missense Tolerance Ratio (MTR) using an established framework adapted to the latest release of population variation data (gnomAD v4.1.0). This metric estimates local genetic constraint by comparing the observed number of missense variants with the expected number of substitutions along protein sequences. We then integrated these regional constraint profiles with 33,124 missense variants reported in ClinVar to investigate how clinically observed variation relates to protein structural features. While average constraint levels differ only modestly between ordered and disordered regions, we observe strong depletion of missense variation within specific subregions. Ordered segments and structural transition regions show the highest levels of constraint. Importantly, localized low-tolerance subregions are also detected within intrinsically disordered regions, indicating that IDRs are not uniformly permissive but can contain functionally important elements. Overall, our results highlight the value of integrating high-quality disorder annotations with regional constraint metrics to better interpret missense variation across proteins with complex structural organization.

P04

Oligomeric-State-Dependent Inhibition of Aβ40 Aggregation by the 14-3-3ζ proteins

Bednarikova Zuzana[1], Kozelekova Aneta[2,3], Novotná Jana[2,3], Kralova Katerina[2], Hritz Jozef[2,4], Gazova Zuzana[1]

[1]Institute of Experimental Physics, Slovak Academy of Sciences, Watsonova 47, 040 01 Kosice, Slovakia, [2]CEITEC MU, Kamenice 5, 62500 Brno, Czech Republic,

[3]National Centre for Biomolecular Research, Masaryk University, Kamenice 5, Brno, 625 00, Czech Republic,
[4]Faculty of Science, Masaryk University, Kamenice 5, 625 00 Brno, Czech Republic

Aggregation of the Aβ peptide into amyloid fibrils is a key pathological feature of Alzheimer's disease and an important target for therapeutic intervention. Identifying endogenous proteins that modulate this process may provide valuable insight into mechanisms of amyloid regulation. The 14-3-3ζ proteins are known for their diverse regulatory roles in cellular processes, and their function depends on their oligomeric state, which can alternate between monomer, homodimer, and heterodimer (1).
In this study, we examined the effects of monomeric and dimeric variants of 14-3-3ζ on Aβ40 aggregation using ThT kinetic assays, global kinetic analysis (2), and atomic force microscopy. Both variants inhibited fibril formation in a concentration-dependent manner, with the monomeric form exhibiting substantially greater inhibitory potency. Kinetic analysis indicated that 14-3-3ζ primarily suppresses early stages of aggregation. Complementary nuclear magnetic resonance spectroscopy (NMR) and fluorescence quenching experiments revealed only weak, transient, and non-specific interactions between monomeric 14-3-3ζ and Aβ40 monomers.
These findings demonstrate that the anti-amyloid activity of 14-3-3ζ is strongly modulated by its oligomeric state and support a mechanism in which dynamic chaperone interactions redirect the Aβ40 aggregation pathway.
Acknowledgement. This work was supported by the Slovak Research and Development Agency under Contract no. APVV-22-0598, APVV-18-0284, Slovak Research Agency VEGA 2/0141/25 and MVTS COST ML4NGP. This publication was also funded by the EU NextGenerationEU through the Recovery and Resilience Plan for Slovakia under the project No. 09-I02-03-V01-00021. This project was also funded by the European Union's Horizon Europe program under grant agreement No 101087124 (ADDIT-CE). Views and opinions expressed are, however, those of the author(s) only and do not necessarily reflect those of the European Union or European Research Executive Agency (REA). Neither the European Union nor REA can be held responsible for them. We acknowledge Josef Dadok National NMR Centre of CIISB, CEITEC CF Biomolecular Interactions and Crystallography, and CEITEC Proteomics CF, Instruct-CZ Centre, supported by MEYS CR (LM2023042) and European Regional Development Fund-Project „Innovation of Czech Infrastructure for Integrative Structural Biology" (No. CZ.02.01.01/00/23_015/0008175). Computational resources for Proteomics CF were provided by the e-INFRA CZ project (ID: 90254), supported by MEYS CR.

References:
1. Z. Trošanová, et al. Journal of Molecular Biology 434 (2022) 167479
2. G. Meisl et al, Nat. Protoc., 11 (2016) 252–272

P05

Monte Carlo simulations of miniprotein folding sampled with an autoencoder

Jan Beránek, Guglielmo Tedeschi, Vojtěch Spiwok

University of Chemistry and Technology, Prague, Technická 5, Prague 6, 166 28, Czech Republic

Autoencoders are artificial neural networks used for non-linear dimensionality reduction and have many diverse applications. The architecture consists of an encoder transforming high dimensional inputs onto a low-dimensional embedding called latent space, and a decoder, which reconstructs these latent space values into the original high dimensional data.
In the work presented here, we use an autoencoder pretrained on diverse conformations of miniproteins for generating protein structures based on input latent space values. We sample from the structures generated by the network using Metropolis Criterion and we show that running such Monte-Carlo simulation of the miniprotein system provides a very computationally inexpensive way to visualize the structural behaviour of the studied miniprotein, including folding, unfolding and other events.
We show the results of this method applied to four model systems: Tryptophan Cage miniprotein, its non-folding variant, Villin headpiece and PDZ domain. We observe that although the distribution of structures sampled this way is slightly different from one sampled from sufficiently long classical molecular dynamics (MD) simulation, it is also provided at a tiny fraction of the computational cost of the MD simulation. We present the method as a potential alternative to classical simulations methods (like MD) that provide more accurate results, but require orders of magnitude longer wall times for the calculations.

P06

Homorepeat Overlap in Orthologs: Evolutionary Insights into Swappable and Non-Swappable PolyX Regions

Nerea Blanco, Pablo Mier

Andalusian Centre for Developmental Biology (CABD, UPO-CSIC-JA), Faculty of Experimental Sciences (Genetics Area), Universidad Pablo de Olavide, 41013 Seville, Spain

Repetitive stretches of identical amino acids, known as homorepeats or polyX, shape protein form, function and evolution. While their presence across orthologous proteins vouches at their functional importance, less is understood about whether different types of homorepeats can occupy the same position in related proteins: a concept we term "swappability". Here we study homorepeat overlaps in orthologs from 695 species, tracking swappable and non-swappable cases to reveal their evolutionary reach and functional implications. We obtained 17,415,904 orthologous protein pairs from OrthoMCL and kept the top 50% most similar pairs. We used polyX2 to detect homorepeats of at least eight identical residues in a ten amino acid window, and aligned each orthologous pair to detect overlaps between homorepeats. We focused on homorepeats that overlapped by at least half their length. Overlapping homorepeats proved rare: 99,386 non-swappable (same amino acid type) and 5,018 swappable (different amino acid type) among 8.7 million orthologous pairs. Swappable repeats were longer and enriched in acidic residues (D/E), while non-swappable varied by lineage. Based on their prevalence in our data, we focused on D/E and A/S overlapping homorepeat pairs. PolyD and polyE were frequently found as swappable, consistently maintaining their functional roles primarily linked to ribosomal biogenesis. In contrast, while polyA and polyS sometimes overlapped, their non-swappable forms displayed distinct functional enrichments, suggesting specialized, non-interchangeable functions. This work highlights the balance between evolutionary flexibility and constraint in homorepeat evolution.

P07

Substitution Matrix for Improved Analysis of Low Complexity Protein Sequences

Abdeljalil Bouziane[1], Maciej Dzikowski, Joanna Ziemska-Legięcka, Patryk Jarnot[2], Aleksandra Gruca[2], Marcin Grynberg[1]

[1]Institute of Biochemistry and Biophysics PAS, Faculty of Mathematics Informatics and Mechanics University of Warsaw, Faculty of Mathematics Informatics and Mechanics University of Warsaw
[2]Institute of Informatics Silesian University of Technology, Institute of Informatics Silesian University of Technology, Institute of Biochemistry and Biophysics Polish Academy of Sciences

Low-complexity regions (LCRs) pose a unique challenge in sequence analysis. Their biased amino acid compositions, low entropy, and distinct mutation mechanisms violate the statistical assumptions behind standard substitution matrices, which are optimized for globular, high complexity protein sequences. To address this challenge, we present the development of an LCR-specific substitution matrix designed to accurately capture evolutionary and functional relationships within low complexity sequences.
Our pipeline adapts a BLOSUM-style log-odds framework specifically for LCRs. Our methodology extracts LCRs from curated protein databases, performing context-aware clustering to group structurally equivalent fragments, and generating specialized multiple sequence alignments. From these alignments, we calculate residue frequencies, observed expected probabilities, and final scaled log-odds scores. We benchmark the resulting candidate matrices against standard (e.g., BLOSUM, PAM) and specialized matrices (e.g., EDSSMat, PHAT). Preliminary evaluations demonstrate that our custom specialized matrix succeeds in improving the retrieval of low complexity sequences, providing a reliable tool for existing sequence analysis programs to better handle similar cases.

P08

Recombinant expression and purification of pVHL isoforms in E. coli for in vitro screening of small molecules

Marianna Bussolino[1], Alessia Strappazzon[1], Lara Abad[1], Frederico Ferreira-da-Silva[2], Fàtima Fonseca[2], Sandra de Macedo Ribeiro[2], Silvio CE Tosatto[1], Emanuela Leonardi[1]

[1]Dept. of Biomedical Sciences, University of Padova, Italy
[2]Instituto de Biologia Molecular e Celular (IBMC) and Instituto para a Investigação e Inovação em Saúde (i3S), Universidade do Porto, Portugal

The two major gene products of the VHL gene, pVHL19 and pVHL30, function as tumor suppressors. They are involved in distinct cellular pathways, with a central role in oxygen homeostasis through their participation in an E3 ubiquitin-ligase complex that directs hypoxia-inducible factor (HIF) to degradation. Mutations in the VHL gene lead to the development of highly angiogenic tumors, a condition known as VHL syndrome. Under native conditions, pVHL exhibits a molten globule configuration with a high tendency to aggregate, which is further enhanced in pVHL19 due to the

absence of the N-terminal disordered domain. Consequently, despite several previous studies, the expression and purification of pVHL remain challenging.
Based on prior evidence supporting arginine as an effective agent for preventing aggregation and promoting correct folding [1], we optimized the expression of the two pVHL isoforms in E. coli and purified them using two different strategies. For pVHL30, we employed direct purification from the soluble fraction, whereas pVHL19 was recovered through on-column refolding from inclusion bodies (IBs). The purified proteins were further characterized by size-exclusion chromatography (SEC) to assess their state, followed by transmission electron microscopy (TEM). Additionally, circular dichroism (CD) and thermal shift assay (TSA) were performed to provide an analysis of the secondary structure and stability.
This optimization, supported by the experimental results, demonstrates that the purified proteins are suitable for functional and aggregation studies. These findings represent a first step toward the development of an in vitro system to screen small molecules that may prevent pVHL aggregation and explore their potential as therapeutic candidates for pVHL-associated tumors, including clear cell renal cell carcinoma (ccRCC).

[1] Shmueli MD et al. (2019), Oncogene, 38 (7), 1038-1049.

P09

Disordered oligomers formed by truncated tau proteins

Ondrej Cehlar[1,2], Stefana Njemoga[1], Andrej Kovac[1], Petra Majerova[1], Tomas Sadecky[2], Vladimir Garaj[2], Exequiel E. Barrera[3]

[1]Institute of Neuroimmunology, Laboratory of Structural Biology of Neurodegeneration, Slovak Academy of Sciences, Bratislava, Slovakia
[2]Department of Pharmaceutical Chemistry, Faculty of Pharmacy, Comenius University in Bratislava, Bratislava, Slovakia
[3]Instituto de Histología y Embriología (IHEM), Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), CC56, Universidad Nacional de Cuyo, Mendoza M5502JMA, Argentina

Soluble oligomers are thought to be the toxic agents of several neurodegenerative diseases and not the relatively inert mature filaments. Oligomers can propagate between neurons and could be isolated by in vivo microdialysis from brain interstitial fluid. Tau oligomers can be induced in vitro by seeding with amyloid-β (Aβ) oligomers or using tau with modified cysteines [1,2]. However, intermediate filaments with shorter amyloid interface compared to mature filaments were observed early in the course of tau aggregation [3].
Previously we have summarized structural data about oligomers formed by amyloid-β, α-synuclein, tau a prion protein, that show substantial lack of structural information about tau protein oligomers [4].
Isolated in vitro tau oligomers prepared from truncated tau were characterized by ion mobility mass spectrometry, size exclusion chromatography, dot blot and western blot using oligomer specific antibodies T22, A11 and TOMA-1. The structures of tau oligomers will be further characterized by SAXS, solid-state and solution state NMR and hydrogen-deuterium exchange techniques.

Moreover, representative structures of tau dimers and trimers were obtained after clustering of coarse-grained MD simulations. Set of previously reported small molecules with potential to inhibit tau aggregation was docked in the presumed binding pockets identified in these structures. The inhibition was probed by measuring the ThT signal of the aggregation reaction in the presence of small molecules.

1. L. Lasagna-Reeves, C., Castillo-Carranza, D., Sengupta, U. et al. Alzheimer brain-derived tau oligomers propagate pathology from endogenous tau. Sci Rep 2, 700 (2012).
2. Gerson, J.E., Sengupta, U., Kayed, R. (2017). Tau Oligomers as Pathogenic Seeds: Preparation and Propagation In Vitro and In Vivo. In: Smet-Nocca, C. (eds) Tau Protein. Methods in Molecular Biology, vol 1523. Humana Press, New York, NY..
3. Lövestam, S., Li, D., Wagstaff, J.L. et al. Disease-specific tau filaments assemble via polymorphic intermediates. Nature 625, 119–125 (2024)..
4. Cehlar, O.; Njemoga, S.; Horvath, M.; Cizmazia, E.; Bednarikova, Z.; Barrera, E.E. Structures of Oligomeric States of Tau Protein, Amyloid-β, α-Synuclein and Prion Protein Implicated in Alzheimer's Disease, Parkinson's Disease and Prionopathies. Int. J. Mol. Sci. 2024, 25, 13049.

This research was funded by the EU NextGenerationEU through the Recovery and Resilience Plan for Slovakia under the project No. 09I03-03-V04-00623. This work was supported by research grants VEGA 2/0125/23 and 1/0825/25. This research was further funded by the European Union's Horizon Europe program under the grant agreement No. 101087124.

P10

Mapping the landscape of antidiabetic peptides: toward a curated literature-derived database

Jarosław Chilimoniuk[1,2], Michał Burdukiewicz[2]

[1]Vilnius University, Lithuania
[2]Medical University of Białystok, Poland

Peptide-based therapeutics are increasingly explored as treatment options for diabetes because of their high specificity, diverse mechanisms of action and the relative ease with which their properties can be tuned through chemical modification. Many bioactive peptides are structurally flexible and do not adopt stable globular folds. In type 2 diabetes, peptide biology is also closely linked to amyloid formation. Pancreatic islets frequently contain deposits formed by aggregation of islet amyloid polypeptide (IAPP). These features highlight the importance of peptide sequence composition, structural plasticity and chemical modification in shaping both biological activity and aggregation-related behavior.
To facilitate a more systematic view of antidiabetic peptides reported in the literature, we are assembling a curated database that integrates experimentally described sequences together with contextual biological information. The current curation effort includes 329 publications, from which peptide sequences and associated annotations are being collected. For each entry, information is recorded on peptide origin, sequence composition, reported modifications and the experimental context in which antidiabetic activity was observed. When available, additional details such as

validation methods, in vitro or in vivo testing, and other reported biological activities are also included.
Bringing these dispersed data together into a single structured resource enables comparative examination of antidiabetic peptide properties and may help identify recurring patterns linking sequence features, modification strategies and biological function. Such analyses could provide insight into the structural and physicochemical factors that influence peptide stability and activity. The resource may also support broader studies of how conformationally dynamic peptides participate in metabolic regulation and how their properties relate to aggregation phenomena observed in diabetes.
Once completed, the database is intended to serve as a platform for exploring therapeutic peptide diversity and for guiding the discovery and rational design of antidiabetic peptides with improved functional profiles.

P11

AmyloDeep: pLM-based ensemble model for predicting amyloid propensity from the amino acid sequence

Alisa Davtyan, Anahit Khachatryan, Rafayel Petrosyan

Institute of Physics, Yerevan State University, Institute of Physics, Yerevan State University, Zaven & Sonia Akian College of Science and Engineering, American University of Armenia,

Amyloids are predominantly β-sheet-rich, stable protein structures that can maintain their presence in the human body for multiple years. Amyloid protein aggregates contribute to the development of multiple neurodegenerative diseases, such as Alzheimer's, Parkinson's, and Huntington's, and are involved in different vital functions, such as memory formation and immune system function. We would like to present Amylodeep which predicts amyloid propensity from the amino acid sequence. To build Amylodeep, first, we aggregated labeled amino acid sequence data from multiple sources, obtaining a roughly balanced dataset of sequences for binary classification. For a comprehensive evaluation, we fine-tuned four variants—ESM2-T6-8M, ESM2-T12-35M, ESM2-T30-150M, and ESM2-T33-650M—using the same training protocol and dataset. All models were initialized with pre-trained weights and fine-tuned with a frozen backbone, updating only the classification head parameters. Among the evaluated variants, the ESM2-T30-150M model consistently demonstrated superior performance for our task, achieving the best trade-off between model capacity and generalization and later we trained new models based on protein embeddings from ESM2 and UniRep(Unified representation of proteins). We trained SVM, XGBoost, the fine-tuned ESM2 150M, FNN with ESM2 embeddings and another FNN with UniRep embeddings. The predictions from these models were then unified into a single soft voting ensemble model, yielding highly robust and accurate results.
We further made a tool where users can provide the amino acid sequence and get the amyloid formation probabilities of different segments of the input sequence. Users can access the light version of AmyloDeep through the web server at https://amylodeep.com/, and the full model is available as a Python package at https://pypi.org/project/amylodeep/.

P12

Disorder-based profiling of plant stress proteins

Dragana Dudic

Faculty of Mathematics and Computer Science, Alfa BK University, Belgrade, Serbia

Non-globular proteins, particularly intrinsically disordered proteins (IDPs), play an important role in plant responses to environmental stress. While IDPs have been extensively studied in animal systems, plant stress-associated non-globular proteins remain comparatively underexplored, despite their abundance and functional relevance.
We performed a systematic analysis of stress-related plant proteins using exclusively publicly available datasets. Proteins annotated as involved in abiotic stress responses were collected from curated databases and compared to a length-matched set of non-stress plant proteins. Intrinsic disorder, low-complexity regions, and amino acid composition were analysed using established prediction tools.
Stress-associated proteins displayed a higher proportion of intrinsically disordered regions and low-complexity segments compared to the control set, together with characteristic amino acid enrichments consistent with flexible and dynamic protein behaviour. Dimensionality-reduction analyses further revealed distinct clustering of stress-related proteins based on disorder-derived features.
These results highlight plant stress proteins as a suitable and accessible testbed for machine-learning-based approaches to non-globular protein annotation. The study underscores the value of plant systems and open data resources for advancing comparative and methodological research on intrinsically disordered proteins.

P13

From SPuRs to Retrieval and Classification of Protein Fragments Defined by Sequence–Structure–Interaction Constraints

Radovan Dvorsky[1,2], Katarina Martonova[2], Adam Polak[2], Stefana Njemoga[2], Ondrej Cehlar[2], Rostislav Skrabana[2]

[1]Institute of Biochemistry and Molecular Biology II, University of Düsseldorf, Germany;
[2]Institute of Neuroimmunology, Bratislava, Slovakia

Consider the following question: how many protein fragments in the Protein Data Bank (PDB) begin with serine or threonine and form a hydrogen bond, via their side-chain hydroxyl group, to the backbone of a downstream residue at position i+3 or i+4? This is not merely a technical or self-contained query, but reflects a central problem in the analysis of structural motifs. In the present case, such fragments are related to SPuRs, which represent a structurally and functionally meaningful class of local motifs. We found that retrieving fragments defined by such specific sequence–structure–interaction constraints is not straightforward and requires a dedicated computational pipeline. We therefore set out to implement such a procedure, beginning with

systematic fragment retrieval. It soon became clear that this challenge could be approached in a much more general way, enabling the development of a broad and systematic framework for the comprehensive classification of local structural fragments across the entire PDB.
We found that such categorization provides a number of interesting insights. Beyond the well-defined patterns associated with classical secondary-structure elements, it reveals which sequence motifs adopt unique and well-defined conformations and which are more promiscuous, being compatible with multiple structural states. The results also show which fragments remain only weakly structured even within otherwise folded proteins, thereby providing a basis for functionally important local flexibility in some cases. In a complementary sense, one may also ask whether motifs that are only weakly structured, or not observed in the PDB at all, may represent precursors of intrinsically disordered fragments.
We will present results from our recent efforts in this area and discuss their potential and future perspectives in the light of advanced approaches, including those coupled with artificial intelligence and deep learning.

P14

Experimentally Refined Conformational Ensembles of Intrinsically Disordered Proteins from SAXS Data

Hamidreza Ghafouri, Silvio C.E Tosatto, Alexander Miguel Monzon

University of Padova, Department of Biomedical Sciences, Italy

Small Angle Scattering Biological Data Bank (SASBDB) contains a large collection of experimentally measured small-angle scattering profiles that provide valuable information on the global structural properties of biomolecules in solution. Among these techniques, Small-Angle X-ray Scattering (SAXS) has become a widely used approach to characterize the conformational ensembles of intrinsically disordered proteins and other flexible systems, as it reports on overall size, shape, and structural heterogeneity without requiring crystallization.
In this project, we performed a systematic analysis of SAXS data to investigate how well computationally generated conformational ensembles reproduce experimental scattering profiles. High-quality SAXS datasets corresponding to monomeric proteins were first selected from SASBDB using a series of filtering criteria designed to ensure reliable experimental measurements and appropriate sample conditions. These curated datasets served as benchmarks for subsequent ensemble generation and validation.
For each selected protein, conformational ensembles were generated using AFFlecto, a computational framework designed to sample structural variability and produce physically plausible ensembles compatible with the intrinsic flexibility of disordered or partially ordered proteins. The resulting ensembles were then evaluated against their corresponding experimental SAXS profiles.
To perform this analysis, we developed a dedicated validation workflow implemented in Python. This software calculates theoretical SAXS profiles for the generated conformations and compares them with the experimental data to quantify the agreement between simulation and experiment. The goodness of fit between calculated and experimental scattering curves was used as the primary metric to assess ensemble quality.

Importantly, the analysis was conducted using two complementary approaches. First, we evaluated the unbiased ensembles, in which all conformations contribute equally to the calculated scattering profile. Second, we applied a reweighting procedure to generate weighted ensembles that optimize agreement with the experimental SAXS data. Comparing these two scenarios allows us to assess the recoverability of the experimental signal from the initially generated ensembles and to evaluate whether the sampled conformational space is sufficiently representative.
Overall, this framework provides a systematic strategy to benchmark ensemble generation methods against experimental SAXS data and to investigate how effectively computational ensembles capture the structural heterogeneity of proteins in solution. The results contribute to improving the reliability of ensemble modeling approaches and their integration with experimental scattering data.

P15

Mapping conformation landscapes hidden in ESMFold protein language models

Vaclav Hanzl, Vojtech Spiwok

University of Chemistry and Technology, Prague, Technická 5, Prague 6, 166 28, Czech Republic

Introducing noise into the ESMFold attention mechanism can uncover plausible protein conformations at lower cost than MSA subsampling or molecular dynamics, as reported in our previous work. We now map and organize the reachable conformation landscape to cover important parts of the function cycle for proteins with substantial movement during their function. To reveal more about the nature of the latent and normally hidden information present in the statically trained model we further specialised the disruptions to affect individual layers and attention heads, clustered the resulting conformations and tried to increase the diversity by metadynamics-like mechanism steering the structures towards underexplored regions by shaping the noise. The whole landscape of diverse structures was then organized according to similarities of structures and their potential energies to elucidate meaningful movements.
We demonstrate potential application on LmrS efflux antiporter of Staphylococcus aureus (UniProt entry A0A0U1MT24_STAAU) which plays a role in multidrug resistance. Computational analysis of potential antibiotics adjuvants blocking this pump requires ensemble docking against structures covering the function cycle. There is no experimentally determined structure deposited in the Protein Data Bank but there is a wide coverage of other proteins in this major facilitator superfamily (MFS) giving a good chance of the required information being captured in the ESMFold protein language model. We compare our results with diffusion based models like Boltz 2 or AlphaFold 3 and models explicitly oriented to structure diversity like AlphaFlow, ESMFlow and BioEmu.

P16

Benchmarking Computational Methods to Characterize the Conformational Ensembles of Intrinsically Disordered Proteins

Richárd Hornyák, Gábor Erdős, Zsuzsanna Dosztányi

The function of intrinsically disordered proteins (IDPs) rely on a dynamic conformational ensemble, which are particularly hard to study experimentally, creating a need for computational methods that can accurately recover ensemble properties from sequence. Despite rapid progress in coarse-grained molecular dynamics (MD) force fields and neural-network predictors the field still lacks a common, experimentally anchored benchmark for evaluating sequence-to-ensemble methods. Here, we constructed an independent small-angle X-ray scattering (SAXS)-based benchmark from the Small Angle Scattering Biological Data Bank (SASBDB) to compare different computational methods for IDP ensemble characterization.
Starting from the SASBDB archive, we applied successive filters for strong disorder, experimental quality, monomeric behavior, near-native conditions, and training-set independence, yielding a final benchmark of 42 SAXS measurements for radius of gyration (Rg) analysis, with a subset containing usable pair-distance distributions, P(r). We benchmarked seven approaches spanning four method classes: the random-coil baseline RANCH, AlphaFold-Metainference (AF-MI), three coarse-grained MD force fields (CALVADOS2, Mpipi-GG, and COCOMO), and two neural-network methods (ALBATROSS and IdpGAN). Performance was evaluated using polymer-scaling behavior, direct agreement with experimental Guinier Rg, sequence-length-normalized Rg, P(r) agreement and computational cost. The experimental data followed polymer-like scaling with sequence length, with a fitted Flory exponent of 0.544. In raw Rg comparisons, CALVADOS2 and Mpipi-GG showed the strongest agreement with SAXS, whereas COCOMO and IdpGAN displayed systematic compaction bias, particularly for longer sequences. However, the random-coil baseline achieved Rg agreement comparable to the best MD models, demonstrating that sequence length is a major confounder and that raw Rg alone can overestimate apparent performance. After removing length dependence, agreement dropped sharply for all methods, indicating that capturing sequence-specific heterogeneity remains a major unresolved challenge. For P(r), Mpipi-GG showed the best overall agreement, closely followed by CALVADOS2, while AF-MI behaved heterogeneously across targets. In parallel, ALBATROSS and IdpGAN were orders of magnitude faster than MD, with ALBATROSS providing near-instant ensemble-level predictions.
Overall, this study establishes an independent SAXS-based benchmark for IDP ensemble methods and shows that robust evaluation must go beyond global size metrics to include normalized observables, real-space distributions, and efficiency. Among the tested approaches, CALVADOS2 and Mpipi-GG showed the best overall agreement with the experimental data, while the results also highlight clear opportunities for improving both simulation-based and learned predictors.

P17

A4D Antibody Aggregation Database (A4D-ABDB): identifying and optimizing aggregation propensity for therapeutic antibodies

Valentín Iglesias[1], Oriol Bárcenas[2], Aleksander Kuriata[3], Carlos Pintado-Grima[2], Julia Zalewska[3], Mateusz Zalewski[3], Sebastian Kmiecik[3], Michał Burdukiewicz[1], Salvador Ventura[2]

[1]Clinical Research Centre, Medical University of Białystok, Kilińskiego 1, 15-369 Białystok, Poland
[2]Institut de Biotecnologia i de Biomedicina and Departament de Bioquímica i Biologia Molecular, Universitat Autònoma de Barcelona, 08193 Bellaterra, Barcelona, Spain
[3]Biological and Chemical Research Center, Faculty of Chemistry, University of Warsaw, Pasteura 1, 02-093 Warsaw, Poland.

Protein aggregation has important implications in different pathologies and represents a bottleneck in the industrial production of proteins. Aggregated biotherapeutics can cause immunogenic responses upon administration, posing a health hazard to patients. The development of monoclonal antibodies (mAbs) has been on the rise over the last decades, thanks to their pharmacological advantages, with thousands advancing to clinical stages. Clinical mAbs have become the standard of care for multiple conditions, including cancer, autoimmune diseases, and also the first disease-modifying therapy for Alzheimer's, extending and improving the quality of life for millions of patients. Additionally, clinical mAbs are steadily entering the lists of best-selling prescription drugs. Aggregation of mAbs in production or storage entails economic losses, altogether making this aspect a major concern in developability of therapeutic mAbs. Aggrescan3D (A3D) has helped develop less aggregation-prone protein molecules, including antibodies. Since its last update, named A4D, we have included quality-of-life improvements such as a pH-dependent aggregation prediction to help design the optimal environmental conditions for specific molecular entities. Here, we present the A4D Antibody Aggregation Database (A4D-ABDB), the most extensive resource for aggregation prediction of therapeutic antibodies. A4D-ABDB includes A4D predictions for in silico protein engineering and, for the first time, allows estimation of the optimal pH for product formulation to enhance solubility of already patented molecular entities. In addition to A4D predictions, and in compliance with FAIR principles, this resource incorporates useful information for understanding each mAb, including its stage of development, target protein, and therapeutic indication, and links to gold-standard therapeutic databases. Importantly, A4D-ABDB is also designed to function as a hub for harmonising monoclonal antibody identifiers across resources, facilitating cross-database interoperability and improving the findability and reusability of therapeutic antibody data. Additionally, A4D-ABDB is designed to enable users to easily perform custom aggregation consultations on the A4D prediction server. The database is publicly accessible without registration. We foresee A4D-ABDB becoming a core resource for understanding and extracting features from successful mAbs, engineering novel mAbs, improving mAb developability, and as a cost-efficient method for optimising formulation conditions for newly developed therapeutic antibodies.

P18

Detecting non-canonical coiled coils beyond state-of-the-art methods

Kamil Kaminski[1], Stanislaw Dunin-Horkawicz [1,2]

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

Coiled-coil (CC) domains are common protein structural motifs formed by two or more α-helices wound around each other into supercoiled bundles. These motifs can be identified in protein structures using the SOCKET tool, which applies geometric criteria to detect them. However, because of the high diversity of coiled-coil structures, many non-canonical cases are missed by this rule-based approach.

To explore this issue, we used supervised geometric deep learning to learn representations of α-helices together with their structural surroundings. Clustering of these learned representations revealed many potential motifs that were not annotated as CCs, but resembled those identified as CCs by SOCKET (structure-based) and/or DeepCoil2 (sequence-based). To classify these candidates and to address label imbalance, while accounting for the uncertainty of negative examples, we subsequently applied Positive-Unlabeled learning. Investigation of the newly identified cases revealed a variety of non-canonical CC arrangements, demonstrating the applicability of the method.

P19

Atlas of amino acid roles in human proteins

Filip Kapral, Vojtěch Spiwok

Department of Biochemistry and Microbiology, University of Chemistry and Technology, Prague, Technická 5, Prague 6, 166 28, Czech Republic

While protein language models allow us to decode the contextual meaning of amino acids and predict structures, a major challenge remains: much of the human proteome is still poorly understood. Although the human genome encodes approximately 20,000 proteins, research efforts have concentrated on only a fraction of them. A protein is considered well studied if its Full Publication Equivalent (FPE) is ≥100; by this definition, only 4,817 proteins - about 24% of the proteome - meet this threshold. In contrast, nearly 44% have an FPE below 10 or have never been described in the literature, leaving a vast “dark space” in human proteomics.
If amino acids derive their functional roles from sequence context, residue-level embeddings provide a powerful lens to study protein structure and function. By extracting contextual representations for each residue and projecting them into a reduced latent space, we can systematically explore patterns of structural and functional similarity across proteins.
As a first step, we focus on well-characterized proteins, where existing structural and functional annotations provide a reliable reference framework. This approach allows us to investigate the biological signals captured in the embedding space and to assess whether residues with similar biochemical properties, structural environments, or functional roles cluster together.
To facilitate this exploration, we developed an interactive web-based application linking residue embeddings directly to their corresponding three-dimensional structures. This environment enables users to navigate the latent space, examine clustering behavior, and connect computational representations with structural context. As expected, residues with similar biochemical or structural roles tend to group together, supporting the idea that the embedding space captures functionally relevant signals.
Building on this foundation, our current work systematically characterizes residue-level embedding patterns within well-annotated proteins. By leveraging known structural and functional information, we aim to better understand which features are encoded in the latent space and how they relate to biochemical roles and structural environments. This rigorous interpretation is a necessary step toward future applications, where the same framework may be extended to underexplored regions of the proteome to help illuminate the largely unknown “dark space” of human proteomics.

This work was supported by the Ministry of Education, Youth and Sports (reg. No. LUC24136, LM2023055), ELIXIR CZ and COST (ML4NGP, CA21160) and by grant of Specific university research – grant No A2_FPBT_2026_030.

P20

Energy Landscape Consequences of Disulfide Bonds in Proteins

Arnau Francesca Llaberia, Miguel Fernandez-Martin, Gonzalo Parra

Evolutionary Systems Biophysics

Introduction. Disulfide bonds are a common evolutionary strategy that contributes to protein structure stability and has additional roles beyond structural support, including facilitating the folding of small peptides that lack a large hydrophobic core, structuring regions within otherwise disordered proteins and regulating protein function through redox-dependent conformational changes. However, the energetic consequences of disulfide bond formation across proteins remain poorly characterized. Here we investigate how disulfide bonds influence protein energy landscapes using local energetic frustration analysis and disorder predictions (SETH). We processed all Protein Data Bank (PDB) structures containing at least one disulfide bond (full dataset n = 6997) and a subset of proteins with structures available in both oxidized and reduced states (paired dataset n = 699). We calculated local frustration patterns both with FrustratometeR (structure-based) and FrustrAI-seq (sequence-based with evolutionary information)
Results. Overall, we found that cysteines that participate in disulfide bonds (SS-Cys) are more stable (enriched in minimally frustrated states) compared to non-bonded cysteines. Low dimensional projections of the protein sequences using ProtT5 reveal extensive heterogeneity in terms of disulfide bonds number, protein length and frustration patterns. Short peptides that lack a large hydrophobic core cluster in regions enriched in disulfide bonds that facilitate their folding. Intrinsically Disordered Proteins (IDPs), as predicted by SETH, occupy a region characterized by long sequences and a low proportion of highly frustrated positions and increased neutral frustration. In this case, regions surrounding the disulfide-bonded cysteines display significantly higher local structural order, forming locally ordered segments within generally unordered structures. Finally, in cases where the disulfide bond is not critical for structural stability (SSbond in the reduced and oxidized states), we observe that the disulfide bond significantly alters the local energy. In several cases, Cys-Cys contacts transition from a minimally frustrated state to neutral or highly frustrated upon oxidation with functional consequences (e.g. GolB metalloprotein binds Au(I) when reduced).
Concluding Remarks. Our results reveal a clear energetic distinction between disulfide-bonded and non-bonded cysteines and suggest that proteins containing disulfide bonds can be categorized into distinct structural and energetic scenarios. The biophysical consequences of disulfide bonds are varied. In short peptides disulfide bonds compensate for their lack of hydrophobic cores whereas in intrinsically disordered proteins, disulfide bonds bring together regions that otherwise would be in strong energetic conflict. Interestingly, in systems where cysteine contacts persist in both redox states, disulfide formation can redistribute local energetic strain acting as local energetic regulators of function.

P21

NMR-Validated Molecular Dynamics Reveals Phosphate-Induced Ordering in Recombinant Spider Silk Protein

Cajsa Malm, Samuli Ollila, Mykhailo Girych, Thomas Scheibel

Spider silk-based materials provide high-performing materials with unique combinations of properties such as strength, toughness, and flexibility. The main challenge is producing these materials at large scale, since spiders are carnivores and cannot be farmed together. Recombinant production of spider silk-like proteins fiber spinning is possible, but there is a lot of room for optimization of production, processing and fiber properties.
Spider silk proteins consist of highly repetitive domains that form nanocrystals and flexible amorphous regions, making up the foundation of the material's properties. However, the relationship between amino acid sequence, processing conditions and resulting fiber properties are poorly understood. This is partially due to intrinsically disordered and repetitive nature of spider silk proteins. Nevertheless, phosphate ions are known to facilitate liquid–liquid phase separation and pre-assembly of spider silk proteins, which is an important step in producing highly compact fibers.
Here we apply the automized quality evaluation-based simulation selection (QEBSS) [1], combining molecular dynamics (MD) simulations and NMR spin relaxation times, to characterize phosphate ion interactions with spider silk proteins. We have already demonstrated that QEBSS can be used to characterize differences between disordered protein ensembles and their interactions with ions [1,2]. Here we apply QEBSS to interpret previously published experimental NMR data for phosphate ion (H_2PO_4) interactions with recombinant spider silk proteins from Nephila clavipes [3]. Our QEBSS ensembles reproduce experimental differences measured at the presence of 10 mM and 300 mM phospate for a recombinant spider silk protein monomer. Our simulations quantify the protein rigidification upon addition of phosphate manifested by increasing hetNOE, and reveal detailed intrachain contacts and peptide–ion interactions. Our results pave the way for realistic simulations of spider silk protein multimers and aggregates, and provide insights on performance of different MD simulation for disordered protein-ion interactions.

[1] Cajsa's manuscript in chemRxiv
[2] Otaviani et al.
[3] Efsthathias chemRxiv manuscript (Biophysical Journal, in press)

P22

AmyloComp-3D: A Structure-Based Predictor of Amyloid Co-Aggregation

Simona Manasra, Elio Torquet, Limin Tian, Valentin Gonay, Gudrun Aldrian and Andrey V Kajava

CNRS-CRBM, France

Traditionally, amyloid fibrils have been viewed as homogeneous structures composed of identical protein copies stacked in a parallel, in-register manner. However, an increasing number of experimental studies has shown that different amyloid-forming proteins can co-aggregate, forming heterogeneous fibrils containing multiple proteins. Such co-aggregation events are biologically

significant, linked to infectivity, functional amyloid formation, and potentially to cross-seeding between proteins in amyloidoses. Despite growing evidence for co-aggregation, the understanding of its molecular determinants and the development of computational tools to predict it remain at an early stage.
Recently, to address this gap, our team developed the AmyloComp program (1), which predicts the co-aggregation potential between pairs of proteins. The method models the axial stacking of β-arch motifs—core structural elements of amyloid fibrils (2). β-arches are first predicted using ArchCandy, and their compatibility between proteins is then evaluated using compatibility matrices.
With the rapidly increasing number of experimentally determined amyloid structures, we extended this approach to incorporate structural data. We developed a computational pipeline for automatic detection of β-arches in complex amyloid structures, enabling the construction of a library of experimentally observed β-arches. This resource was used to improve co-aggregation prediction, leading to the development of AmyloComp-3D, which we applied to identify candidate protein pairs for subsequent experimental testing.

References.
1. Bondarev, Stanislav A., et al. "AmyloComp: a bioinformatic tool for prediction of amyloid co-aggregation." Journal of Molecular Biology 436.17 (2024): 168437.
2. Kajava, Andrey V., Ulrich Baxa, and Alasdair C. Steven. "β arcades: recurring motifs in naturally occurring and disease-related amyloid fibrils." The FASEB journal 24.5 (2010): 1311.

P23

Experimental and AlphaFold modelling of short regions in the disordered tau fragment dGAE engaged in complexes

Katarína Martonová[1], Adam Polák[1], Stefana Njemoga[1], Ondrej Cehlár[1], Michaela Škrabanová[1], Katarína Tomková[1], Radovan Dvorský[2], Rostislav Škrabana[1]

[1]Institute of Neuroimmunology, Slovak Academy of Sciences, Bratislava, Slovakia,
[2]Institute of Biochemistry and Molecular Biology II, Medical Faculty, Heinrich Heine University Düsseldorf, Düsseldorf, Germany

AlphaFold-Multimer is able to correctly identify interacting Intrinsically Disordered Regions (IDRs) in proteins and predict their mode of binding with a given partner [1,2]. This ability is particularly relevant for the intrinsically disordered protein tau, a central driver of neurodegenerative tauopathies, which pathologically undergoes conformational transitions that promote fibril assembly. The aggregation-prone dGAE fragment (tau 297–391) is particularly prone to form fibrils and seed aggregation of full-length tau [4,5]. Several short linear motifs (SLiMs) were found within dGAE contributing to tau-tau interactions or defining the epitopes recognized by monoclonal antibodies. The complexes of tau peptides and Fabs of monoclonal antibodies DC8E8, MN423 have been solved by X-ray crystallography (PDB ID: 2V17, 5MO3), showing an internal hydrogen bond stabilizing tau SLiM [3].
To investigate how these local structures influence both aggregation and antibody recognition, targeted mutations were introduced to disrupt specific sidechain or main chain interactions within

the SLiMs. Wild-type and mutant dGAE variants were expressed in E. coli, purified by cation exchange and size-exclusion chromatography, and their binding to monoclonal antibodies targeting relevant tau epitopes was evaluated by surface plasmon resonance. The impact of mutations on tau aggregation was assessed in vitro.
To complement these experimental studies, AlphaFold-Multimer modelling was employed to predict the structural impact of mutations on tau–Fab recognition. Wild-type and mutant dGAE sequences were modelled in complex with each Fab to evaluate potential changes in interface geometry. Multiple predictions per complex were generated to account for conformational variability. Complementarity-determining region (CDR) loop RMSDs and interface geometries were analyzed to investigate mutation-induced changes. In addition, AlphaFold modelling was extended to explore how short IDRs may influence tau-tau interactions and aggregation propensity.

1. A. Omidi et al., Proc. Natl. Acad. Sci. U.S.A., 121, (2024), e2406407121.
2. J. Abramson et al., Nature, 630, (2024), 493–500.
3. R. Skrabana et al., Mater. Struct. Chem. Biol. Phys. Technol., 29, (2023), 144
4. M. Goedert et al., Cytoskeleton, 81, (2024), 95–102.
5. K. Kitoka et al., Angew. Chem. Int. Ed., 63, (2024), e202407821.

This work was supported by the grants APVV 21-0479, VEGA 2/0125/23 and 2/0141/23, by NextGenerationEU through the Recovery and Resilience Plan for Slovakia under project No. 09-I02-03-V01-00021, and by the SAS Grant Programme for PhD students APP0774.

P24

Impact of Anionic Aggregation Inducers on Disease-Associated Tau Mutants

Joana Paixão[1], Ana Coutinho[1,2], Carlota Silva[1], Ana Azevedo[1], Manuel Prieto[1], Martin Hof[3], Ana M. Melo[1]

[1]Institute for Bioengineering and Biosciences (iBB), Instituto Superior Técnico, Lisbon, Portugal
[2]Centro de Química Estrutural, Dep. Química e Bioquímica, Faculdade de Ciências, Lisbon, Portugal
[3]J. Heyrovský Institute of Physical Chemistry of the CAS, Prague, Czech Republic

The abnormal deposition of the microtubule-associated protein tau in the brain is a defining hallmark of several neurodegenerative disorders (tauopathies), including Alzheimer's disease and frontotemporal dementia. Disease-associated mutations in the MAPT gene, such as P301L, impair tau physiological function and enhance its aggregation propensity. Increasing evidence suggests that negatively charged molecules, including lipid membranes and polyanions, can promote tau misfolding and self-assembly. Here, we investigated how distinct anionic aggregation inducers—heparin and anionic lipid membranes—modulate the aggregation behavior of wild-type (WT) and P301L tau. Tau aggregation kinetics were monitored using Thioflavin T (ThT) fluorescence assays in solution and in the presence of heparin and phosphatidylserine (POPS)-containing lipid vesicles. In parallel, fluorescence correlation spectroscopy (FCS) was employed to characterize the initial recruitment of WT and P301L tau to lipid membranes with distinct compositions, including anionic membranes (POPS or $PI(4,5)P_2$) and phase-separated membranes (POPC:Chol:SM). Our results show that both WT and P301L tau exhibit negligible interaction with zwitterionic membranes

but bind preferentially to negatively charged liposomes, highlighting the dominant role of electrostatic interactions in tau–membrane association. ThT fluorescence data further indicate that POPS-containing lipid vesicles do not promote extensive amyloid fibril formation of WT and P301L tau. In contrast, heparin acts as a potent aggregation inducer for both tau proteins, with a more pronounced effect observed for the P301L mutant. Together, these findings demonstrate that distinct anionic environments differentially modulate tau aggregation.

P25

Amino acid homorepeats in tape measure proteins correlate with bacteriophage tail length

Antonio Moreno-Rodríguez, Antonio J. Pérez-Pulido, Pablo Mier

Andalusian Centre for Developmental Biology (CABD, UPO-CSIC-JA), Faculty of Experimental Sciences (Genetics Area), Universidad Pablo de Olavide, 41013 Seville, Spain

The tape measure protein (TMP) dictates the tail length of tailed bacteriophages. However, the sequence features that drive the process of length fine-tuning have not been described yet. Tandem repeats (TRs), stretches of amino acids organized as multiple adjacent copies of the same or very similar sequence motif, are contributors to TMP length variation. Homorepeats (polyX regions) are a specific type of TRs composed of stretches of identical amino acids, forming low complexity regions that contribute to the protein structural flexibility and dynamics. Here, we performed a large-scale analysis of polyX regions across 12 million phage proteins to test whether these homorepeats help explain the variable length of flexible phage tails. PolyX tracts were detected in 41% of all phage proteins and in 95.5% of TMPs, with polyA, polyG, polyS, and polyT dominating the composition. Furthermore, the number of polyX tracts in TMPs scaled directly with protein length. About half of the TMPs have both polyX and TRs, but nearly half of the polyX-containing TMPs lacked any TRs, showing that polyX are independent features rather than byproducts of larger repeats. Morphological comparisons showed that phages with long and flexible tails (Siphoviridae) combine polyX and TRs, whereas shorter or contractile-tailed phages rely mostly on polyX. Last, the total amino acid content in polyX regions correlated significantly with experimentally measured tail lengths. Our results identify polyX tracts as an important and tunable component of TMPs that enable precise control of phage tail length.

P26

Characterization of Alpha-synuclein Fibrils with Highly Divergent Thioflavin-T Binding Properties

Kamile Mikalauskaite, Mantas Ziaunys, Vytautas Smirnovas

Institute of Biotechnology, Life Sciences Center, Vilnius University, Vilnius, Lithuania

Alpha-synuclein (α-syn) is a 140-amino acid presynaptic protein associated with neurodegenerative disorders, including Parkinson's disease, dementia with Lewy bodies and multiple system atrophy. The aggregation of this protein into amyloid fibrils is highly dependent on environmental factors, resulting in structurally and morphologically distinct fibrils. Mostly during aggregation studies to

detect and quantify these aggregates the fluorescent dye Thioflavin-T (ThT) is used. However its binding affinity and fluorescence intensity vary across different α-syn fibril strains. This divergence suggests that ThT fluorescence intensity may not consistently correlate with the total quantity of amyloid aggregates.
In this study, a large library of α-syn fibril samples was generated under identical conditions which promote structural variability. Samples exhibiting the highest and lowest ThT fluorescence intensities were selected and replicated during seeded aggregation. These samples were then characterized by their self-replication propensities, secondary structures using Fourier-transform infrared spectroscopy (FTIR) and morphologies employing atomic force microscopy (AFM).
The results of these studies revealed that fibrils with high ThT intensity have a significantly lower rate of self-replication compared to samples with low ThT intensity. Furthermore, these two groups showed differences in their secondary structure compositions and morphological features. These results demonstrate that ThT binding properties are linked to specific structural characteristics that differ from those facilitating rapid self-replication. The findings support the conclusion that ThT fluorescence intensity is a property of the specific fibril strain rather than a direct indicator of aggregate amount. This work highlights the limitations of using ThT as a universal quantitative tool in the study of alpha-synuclein strains.

P27

Multivalent Tau–Microtubule Interactions as a Case Study in Non-Globular Protein Function

Segev Naveh-Tassa, Prof. Koby Levy

Weizmann institute of science, Weizmann institute of science, Israel

Non-globular proteins often function through highly dynamic and heterogeneous interactions that defy static structural descriptions. A central challenge is therefore not only understanding how such proteins bind their partners, but how they encode robust biological function in the absence of a single dominant binding mode.
Tau is a dynamic microtubule-associated protein essential for maintaining microtubule (MT) stability and neuronal function. MTs are polymeric assemblies composed of repeating α/β-tubulin dimers, whose formation and regulation rely on transient yet coordinated protein interactions. Tau performs this regulation through its intrinsically disordered architecture and multivalent binding capacity, but these same properties make its mode of action difficult to characterize structurally. Disruption of Tau–MT interactions, as occurs in Alzheimer's disease and related tauopathies, leads to impaired MT regulation and neuronal dysfunction.
Here, we apply coarse-grained molecular dynamics simulations, calibrated against experimental FRET and FCS measurements, to investigate Tau–MT interactions and to uncover how multivalent binding emerges from Tau's intrinsically disordered architecture. Our simulations capture Tau interactions with both the structured tubulin surface and the flexible, disordered C-terminal tails of tubulin, revealing how electrostatic complementarity enables Tau to dynamically engage multiple tubulin dimers while remaining conformationally heterogeneous.
We show that Tau multivalency is not a single fixed property but a collective outcome of sub-regional contributions, electrostatic tuning, and sequence composition. Distinct Tau sub-regions contribute differentially to MT engagement, while isoform variation modulates the

balance between extended, highly multivalent binding and more compact, transient association. In this framework, Alzheimer's disease–associated phosphorylation weakens multivalent engagement, thereby compromising MT stabilization.

P28

N-terminal truncation of tau dGAE (297–391) highlights the importance of residues 321–325 in early fibril formation

Stefana Njemoga[1], Zuzana Bednarikova[2], Exequiel Barrera[3], Zuzana Gazova[2], Ondrej Cehlar[1]

[1]Institute of Neuroimmunology, Slovak Academy of Sciences, Bratislava, Slovakia
[2]Institute of Experimental Physics, Slovak Academy of Sciences, Kosice, Slovakia
[3]IHEM, Universidad Nacional de Cuyo (CONICET), Mendoza, Argentina

Advances in cryo-electron microscopy have established the structures and sequence composition of tau amyloid fibrils across different diseases. However, the mechanisms underlying disease-specific fibril morphology remain unclear. Short amyloidogenic motifs within the tau sequence are considered key determinants of aggregation due to their intrinsic β-sheet propensity and ability to form stable fibril cores. Truncation, a common post-translational modification of tau, can modulate aggregation by exposing or altering these motifs, motivating the present study. We employed an integrative approach combining experimental techniques and molecular dynamics (MD) simulations to investigate aggregation of five N-terminally truncated tau variants. AFM and ThT fluorescence assays revealed distinct amyloidogenic propensities and identified the tau variant 321-391 (numbered according to the longest CNS tau isoform 1-441) as the minimal E391-truncated construct capable of forming amyloid fibrils in vitro, whereas the shorter tau326-391 variant did not form fibrils. All-atom and coarse-grained simulations showed that structural transitions were driven by known amyloid-nucleating motifs, including the G-motif, PHF6**, and PAM4. However, several structural features potentially explaining the different aggregation propensities of tau321-391 and tau326-391 were observed, including a higher prevalence of hairpin-like conformations in tau321-391 in all-atom simulations and a pronounced α-helical propensity in its N-terminal region. Dimeric structures of tau321-391 obtained from simulations were further used for small molecule docking. Four candidate compounds were selected and subsequently tested for their potential aggregation-inhibiting activity using ThT fluorescence assays.

This work was supported by the Scientific Grant Agency of the Ministry of Education, Research, Development and Youth of the Slovak Republic and the Slovak Academy of Sciences (VEGA) grant nos. 02/0141/25, 2/0125/23 and 1/0825/25, Slovak Research and Development Agency under the Contract nos. APVV-18-0284 and APVV-22-0598, and MVTS COST grant ML4NGP. This research was further funded by the EU NextGenerationEU through the Recovery and Resilience Plan for Slovakia under the projects No. 09I03-03-V04-00623 and 09I02-03-V01-00021. This research was further funded by the European Union's Horizon Europe program under the grant agreement No. 101087124 - ADDIT-CE and by the European Union's Horizon 2020 research and Innovation programme under the Marie Skłodowska-Curie grant agreement No 873127 - InterTAU. Part of the research results were obtained using the computational resources procured in the national project National competence centre for high-performance computing (project code: 311070AKF2) funded by the European

Regional Development Fund, EU Structural Funds Informatization of society, Operational Program Integrated Infrastructure.

P29

Tessellation-based comparison of conformations, interfaces, and ensembles of non-globular proteins

Kliment Olechnovic

Vilnius University Life Sciences Center

Non-globular proteins (including intrinsically disordered proteins, amyloid fibrils, repeat proteins, and other extended assemblies) pose fundamental challenges for structural comparison. Unlike compact globular domains, these systems are often elongated, conformationally heterogeneous, and stabilized by repetitive or multivalent interfaces rather than well-defined cores. Conventional comparison approaches rely primarily on coordinate superposition and distance-based metrics such as RMSD, TM-score, or binary contact overlap. However, superposition becomes ambiguous for flexible or non-compact structures, while simple distance thresholds fail to account for the geometric context of neighboring atoms. In amyloids and other extended assemblies, subtle rearrangements of inter-chain packing can define distinct polymorphs yet remain poorly captured by global coordinate deviations - each polymorph identity is often interface-defined rather than backbone-defined.
Robust structural similarity measures are therefore essential - not only for analytical comparison and clustering of conformational ensembles, but also as a foundation for machine learning workflows, where distance metrics underpin embedding, benchmarking, and model evaluation.
Voronoi tessellation provides a natural geometric framework for describing molecular packing. In a solvent-accessible surface–constrained tessellation, atoms are represented as balls partitioning space into Voronoi cells, and interatomic contacts are defined by shared Voronoi faces with well-defined areas. Unlike distance-based contacts, tessellation-derived contact areas incorporate the influence of all surrounding neighbors and can be directly summed and consistently aggregated at atom, residue, and chain levels. This representation is particularly well suited for analyzing interfaces, as contact areas provide an additive and geometrically meaningful measure of packing interactions.
The Voronoi tessellation-based Contact Area Difference score (CAD-score) was previously established as a superposition-free similarity measure for evaluating structural models of globular proteins by directly comparing matched contact areas. However, its applicability to non-globular proteins, where extended interfaces and repetitive packing dominate structural organization, has not yet been systematically explored.
Here, we present CAD-score-LT, a methodologically extended new implementation that substantially improves computational efficiency, scalability, and usability. By leveraging optimized tessellation-based contact area computations, CAD-score-LT enables rapid evaluation of large structural datasets, efficient all-to-all comparisons, and construction of similarity matrices suitable for fast clustering of structural ensembles. It naturally supports both global assessment of entire structures and highly flexible localized evaluation of specific regions - for example, inter-subunit

interfaces or binding sites. CAD-score-LT provides a rigorous, scalable, and ML-compatible similarity framework that is especially suitable for analyzing non-globular protein architectures.

P30

Long-range cross talk between intrinsically disordered regions modulates FOXA1 interactions with DNA

Magnús A. Rúnarsson, Vanessa Ung, Davide Mercadante, Pétur O. Heidarsson, Kristinn R. Óskarsson

University of Iceland

FOXA1 is a pioneer transcription factor (pTF) that is essential for multiple embryonic organ developmental pathways and acts as a key postnatal regulator of hormone-responsive nuclear receptors across diverse tissues. Several mutations in the protein have in addition been linked to worse outcomes of breast and prostate cancer therapies. Predictions indicate that approximately 80% of FOXA1 residues reside in intrinsically disordered regions (IDRs), forming two large IDRs that flank the forkhead DNA-binding domain (DBD). These two IDRs differ considerably in their chemical composition. The N-terminal IDR contains a negatively charged patch at the very end of the N-terminus followed by an uncharged region, whereas the C-terminal IDR exhibits evenly distributed positive and negative charges. Here, we investigate the roles of these two IDRs using single-molecule Förster resonance energy transfer (smFRET) and coarse-grained molecular dynamics (CGMD). Our experiments reveal potential long-range crosstalk between the N-terminal and C-terminal IDRs, as well as interactions between the C-terminal IDR and the DBD. Using FOXA1 variants truncated at either the C-terminal or N-terminal IDR, we observe that removal of the C-terminal IDR reduces DNA-binding affinity, whereas removal of the N-terminal IDR increases DNA binding affinity. These results suggest that the N-terminal IDR plays an autoinhibitory role in regulating FOXA1 DNA binding. Together, our results highlight how features of intrinsically disordered regions modulate long-range intramolecular interactions and regulate DNA binding in pioneer transcription factors.

P31

Towards a Structural Understanding of the Intrinsically Disordered Actin-Binding Protein Synaptopodin

Max Pawlyta, Frans Mulder

JKU Linz, JKU Linz

The axon initial segment (AIS) is a specialized region of neurons that separates the axonal and somatodendritic compartments and plays a key role in the initiation of action potentials. Structural plasticity of the AIS directly affects neuronal excitability. The AIS contains a specialized smooth endoplasmic reticulum sub-compartment known as the cisternal organelle, a putative regulator of Ca^{2+} trafficking. Formation of this structure depends on the actin-binding protein synaptopodin. The brain isoform of synaptopodin (~74 kDa) is predicted to be fully intrinsically disordered. Here, we aim to characterize the structural ensemble, dynamics, and interaction network of synaptopodin using a

combination of biophysical methods. As the interaction of synaptopodin with actin is of primary importance in this investigation, we first focused on expressing and purifying its known actin-binding region. Preliminary circular dichroism and NMR spectroscopy experiments confirm that this region is largely disordered with only minor residual secondary structure. Ongoing work focuses on backbone resonance assignment and on characterizing interactions between synaptopodin and its binding partners. We aim to integrate these results with single-molecule FRET and molecular modeling approaches to obtain a more complete picture of synaptopodin's structure and function.

P32

Disordered Tails of Myc:Max Regulate DNA-Binding Kinetics and Specificity via Transient Competition

Shira Peer, Prof. Yaakov Levy

Weizmann Institute of Science, Israel

Intrinsically disordered regions (IDRs), flexible and dynamic protein segments, critically regulate protein function. Surprisingly, IDRs within DNA-binding proteins (DBPs) often feature negatively charged residues, counterintuitive given DNA's negative charge. At the same time, the dynamic and heterogeneous nature of IDRs poses challenges for experimental characterization of their molecular mechanisms. Here, we use coarse-grained molecular dynamics simulations to dissect how negatively charged IDRs modulate DNA recognition kinetics and specificity in the Myc:Max transcription factor complex. We find that the Max N-terminal tail accelerates target site recognition through transient autoinhibitory interactions with the DNA-binding domain (DBD), which bias one-dimensional diffusion toward hopping rather than continuous sliding. This mechanism accelerates the search process and thus the overall binding kinetics. Moreover, the Max N-tail enhances sequence specificity. With the tail present, binding to low-affinity sites is selectively suppressed by reduced association rates, whereas binding to target sites remains largely unaffected. We further show that attractive interactions between the Myc and Max tails partially attenuate these effects, revealing an additional layer of regulation mediated by IDR-IDR coupling within the heterodimer. Together, these results demonstrate how negatively charged IDRs encode regulatory function through disorder and dynamics, fine-tuning the balance between rapid genomic search and accurate target discrimination. Our findings provide a mechanistic framework for understanding how non-globular regions contribute to molecular recognition in DNA-binding proteins.

P33

SAXS–Guided Modeling of the Aryl Hydrocarbon Receptor–Hsp90 Complex and Its Assembly with Importins

Santiago Rodriguez, Axel Sanfilippo, Pau Bernadó, Jakub Gruszczyk,

Centre de Biologie Structurale, Université de Montpellier, INSERM, CNRS, Montpellier, France

The aryl hydrocarbon receptor (AHR) is a ligand-activated transcription factor involved in xenobiotic sensing and cellular homeostasis. In its cytosolic state, AHR forms a multiprotein complex with HSP90 and XAP2 (p81 complex), and upon activation it translocates to the nucleus thanks to the interaction with importins α1 and β1. Despite the availability of partial high-resolution cryo-EM structures, the conformational organization p81 and its complex of importins complex remain poorly understood.
In this work, we combined integrative molecular modeling and small-angle X-ray scattering (SAXS) to characterize the conformational landscape of p81 alone (255.55 KDa) and in complex with importins (407.32 KDa). Starting from available experimental structures of p81 (PDB:7ZUB) and the hetero-dimer of importins α1,β1 (PDB:8GCN), missing regions were reconstructed with AlphaFold3, and large conformational ensembles for the two complexes were generated with MoMA-FReSa. Finally, structures compatible with the SAXS data were derived through a sub-ensemble selection procedure.
Our results show that the N-terminal region of AHR (1-270), encompassing the NLS, the DNA-binding and the PAS-A domains, remains highly flexible in solution in the context of p81 and its complex with importins, which explains its absence in the cryo-EM model. However, a detailed analysis of the SAXS-driven sub-ensemble suggests that this region experiences some degree of compactness, indicating a conformational equilibrium involving partially folded states of the PAS-A domain.
Together, these findings reveal that the structural plasticity of AHR is strongly modulated by its biomolecular context and demonstrate the power of SAXS-driven ensemble molecular modeling to describe large, highly dynamic transcriptional regulatory complexes.

P34

Using Single-Molecule FRET to Decipher the Effects of Oncogenic Mutations on the Pioneer Factor FOXA1

Magnús A. Rúnarsson, Magnea S. Sigmarsdóttir, Kristinn R. Óskarsson, and Pétur O. Heiðarsson

Department of Biochemistry, Science Institute, University of Iceland.

Pioneer transcription factors (pTFs) play a key role in maintaining cell function and identity through chromatin remodeling and gene regulation. Their unique ability to bind nucleosomal DNA and access transcriptionally silent genes enables them to manipulate the chromatin landscape, thereby providing other transcription factors with access to their target sequences. Among the most well-known pTFs is FOXA1, a protein critical for development and hormonal gene regulation. FOXA1 is strongly linked to steroid-hormone-regulated cancers through its interactions with nuclear receptors (NRs), such as estrogen receptor alpha (ERα), that are often upregulated in hormone-dependent tumors. The relationship between FOXA1 and NR regulation is apparent in estrogen receptor-positive (ER+) breast cancers. FOXA1's chromatin remodeling abilities are key for activating ERα, which is essential for the expression of estrogen-dependent tissues, with over 90% of estrogen-regulated genes relying on FOXA1 for estrogen response. FOXA1 mutations have been associated with resistance to endocrine therapies such as estrogen receptor antagonists, making it an important subject of research. A mutational hotspot in FOXA1 is the wing 2 (W2) region of its DNA-binding domain (DBD). Mutations within this region are thought to alter DNA-binding specificity, but their mechanistic effects remain poorly understood due to the dynamic nature of FOXA1. While the DBD is largely structured, the W2 region is predicted to be intrinsically disordered, which

complicates conventional structural studies. Single-molecule FRET (smFRET) provides a powerful approach for studying the disordered character of FOXA1 by measuring intramolecular distances between fluorophores attached to specific positions within the protein. Preliminary data on W2 mutants indicate altered DNA-binding behaviour, frequently reducing binding affinity for canonical FOXA1 recognition sites. Additionally, the S250F W2 mutation shows detectable changes in dye distances, suggesting it alters the structural ensemble of FOXA1. These findings show that smFRET can uncover how cancer-associated mutations reshape FOXA1's conformational dynamics and DNA recognition. This mechanistic insight may help explain how these mutations contribute to hormone-dependent cancer progression and resistance to endocrine therapies.

P35

AmyloGraph Ab-DB: A Resource for Amyloid–Antibody Interactions

Mariia Rutkowska[1], Valentín Iglesias[1], Jarosław Chilimoniuk[1], Damian Wilary[1], Julia Szkóp[1,2], Ronja Tittel[3], Martyna Podlasiak[1,4], Michał Burdukiewicz[1]

[1]Medical University of Białystok, Poland
[2]University of Warsaw; Poland
[3]Brandenburg University of Technology, Germany
[3]University of Porto, Portugal

Antibodies targeting aggregation-prone proteins such as alpha-synuclein (α-syn), amyloid β protein (Aβ), and tau are increasingly investigated as therapeutic agents, diagnostic tools, and molecular probes of amyloid formation. These proteins are associated with multiple conformational states and follow diverse aggregation pathways, which can differentially affect antibody recognition. Although a growing body of literature describes antibody binding to amyloid species and their effects on aggregation, the available evidence remains fragmented across publications, and no dedicated resource currently enables systematic comparison of amyloid-antibody interactions.
To address this gap, we are developing AmyloGraph Ab-DB, a curated database of experimentally supported interactions between antibodies and amyloids. The resource is designed to integrate three complementary layers of information: the antibody, the amyloid target, and their interaction. Accordingly, the database captures antibody identity and type, available sequence and structural information, amyloid identity, species and sequence modifications, as well as the experimental context in which the interaction was reported. Moreover, AmyloGraph Ab-DB records the reported effects of antibodies on amyloid aggregation, the experimental methods used to assess the interaction or its functional effect, amyloid species recognized by the antibody, and, where available, paratope and epitope information.
To date, 2,802 publications have been screened and 194 studies have been selected for detailed curation. From the 175 studies that passed first-stage validation, we extracted 426 experimentally supported amyloid-antibody interactions into standardized annotation forms. So far, 43 of these studies have also passed second-stage validation. During validation, 10 studies were excluded at the first stage, 1 retracted study was identified, and 1 additional study was excluded at the second stage. These findings highlight the importance of a multi-stage validation workflow for ensuring the reliability and consistency of curated amyloid-antibody interaction data. Furthermore, we developed and integrated a machine-learning model for automated literature screening, ensuring that the

database will remain scalable and synchronized with future updates. AmyloGraph Ab-DB aims to support comparative analyses of amyloid-antibody interactions and provide a useful reference for studies of aggregation mechanisms and amyloid-related diseases.

P36

Ligand docking into the truncated tau dimers as targets for design of small molecular therapeutics

Tomas Sadecky[1], Stefana Njemoga[2], Exequiel E. Barrera,Vladimir Garaj[1], Ondrej Cehlar[1,2]

[1]Department of Pharmaceutical Chemistry, Faculty of Pharmacy, Comenius University, Odbojarov 10, 832 32 Bratislava, Slovakia,
[2]Institute of Neuroimmunology, Slovak Academy of Sciences, Dubravska cesta 9, 845 10 Bratislava, Slovakia
[3]Instituto de Histología y Embriología (IHEM), Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), CC56, Universidad Nacional de Cuyo, Mendoza M5502JMA, Argentina_Author3

Alzheimer´s disease pathology is described mainly by formation of amyloid deposits composed of physiologically intrinsically disordered proteins (IDPs), such as extraneuronal amyloid plaques β and intraneuronal neurofibrillary tangles (NFTs), assembled by aggregated Tau protein. Current research trends in this field suggest, that the aggregation of tau protein can be stopped or filaments can be disaggregated by a small molecules derived from methylene blue (MB), rhodanine, benztiazole, idarubicine and also molecules found in nature like epigalocatechinegallate (EGCG), homoprejadomycine, cyanidine and others.
The main focus of our work was to successfully dock the above – mentioned small molecules into dimers of tau protein. Tau dimers were chosen to intervene with the earliest stages of tau aggregation. Conformations of tau protein were obtained via coarse grained molecular dynamics simulations. For docking of 27 ligands into dimers of tau proteins (41 conformations) we used program Glide and for energy computation MM-GBSA method in program Prime, both were implemented in the Small-Molecule Discovery Suite software package (Schrödinger Inc., USA). Post docking evaluation included visually inspecting key interactions between ligands and amino acid residues in binding pocket, such as hydrogen bonds, hydrophobic interactions or π-π interactions. Ligands were ranked based on the best binding energies expressed through GlideScore, and the more precise MM-GBSA Score.
MM – GBSA results suggest, that molecules which have the lowest binding energies (the highest affinity towards tau dimers) are benztiazoles, idarubicin, daunorubicin, rhodanine derivatives, and natural compounds, such as EGCG, cyanidine and homoprejadomycine.
The project was funded by the grant agency of Ministry of education SR, VEGA 1/0825/25.

P37

Extracting disorder information from ordered structures deposed in PDB

Rostislav Skrabana[1], Katarina Martonova[1], Adam Polak[1], Stefana Njemoga[1], Ondrej Cehlar[1], Radovan Dvorsky[2]

[1]Institute of Neuroimmunology, Slovak Academy of Sciences, Slovakia,
[2]Institute of Biochemistry and Molecular Biology II, Medical Faculty of the Heinrich Heine University Düsseldorf, Germany

Appropriately filtered and validated information from experimentally determined structures of proteins in PDB served for development of extremely successful artificial intelligence (AI) tools for prediction of three-dimensional structure of a protein directly from the sequence [1]. Currently, with AlphaFold-Multimer or AlphaFold 3, the complexes of proteins with their interaction partners are amenable to successful prediction [2]. In the recent period, AI tools were shown to be able to approach the next level challenges, consisting in prediction of the conformational ensembles (CE) of intrinsically disordered proteins (IDPs) and their interactions, using AlphaFold3 combined with the molecular dynamics and interaction potentials [3-5].
We wondered how strongly might be the information about IDPs' CEs already present in the structures deposited in PDB. We will present our strategy and first results to dissect this question.

1. J. Jumper, R. Evans, A. Pritzel et al., Nature, 596, (2021), 583–589
2. J. Abramson, J. Adler, J. Dunger, et al., Nature, 630, (2024), 493–500.
3.B. Novak, J.M. Lotthammer, R.J. Emenecker et al., Nature (2026). https://doi.org/10.1038/s41586-026-10141-2
4. V. Schnapka, T.I. Morozova, S. Sen et al., Nat Commun (2026). https://doi.org/10.1038/s41467-026-69172-y
5. A. Omidi, M. H. Møller, N. Malhis, J. M. Bui, J. Gsponer, Proc. Natl. Acad. Sci. U.S.A., 121, (2024), e2406407121.

This work was supported by the grants APVV 21-0479, VEGA 2/0125/23 and 2/0141/23, by NextGenerationEU through the Recovery and Resilience Plan for Slovakia under project No. 09-I02-03-V01-00021, and by the SAS Grant Programme for PhD students APP0774.

P38

Long-Range Interaction–Driven Identification of Carbonic Anhydrase Inhibitors as Direct Tau Binders

Kristina Stevanović, Milan Senćanski

Vinča Institute of Nuclear Sciences, University of Belgrade / National Institute of Chemistry, Ljubljana, Vinča Institute of Nuclear Sciences, University of Belgrade

Alzheimer's disease (AD) is strongly associated with the pathological aggregation of the intrinsically disordered Tau protein. Owing to its structural heterogeneity and highly dynamic conformational landscape, Tau presents a significant challenge for structure-based drug discovery. Although several carbonic anhydrase (CA) inhibitors, including Methocarbamol, have been reported to exert

beneficial effects in the treatment of neurodegenerative diseases – particularly in AD through indirect mechanisms such as enhanced Tau clearance – the possibility of direct interactions with Tau remains insufficiently explored.
Here, we present a multistep computational workflow for the identification of direct Tau binders among CA inhibitors. Contrary to well-established structure-based approaches, we choose long-range electrostatic interactions analysis, as intrinsically disordered proteins lack stable tertiary structure. Therefore, this workflow is primarily oriented toward protein-ligand recognition.
Our protocol integrates long-range interactions analysis, ligand druggability evaluation, coarse-grained molecular dynamics simulations, and ensemble-based molecular docking. This approach provides an evaluation of ligand binding across representative structural states, emphasizing the protein domains that play key roles in protein-ligand recognition. In this manner, reliance on a single static structure is avoided, and ligand binding is instead assessed across representative conformational states.
Using this approach, we identified several CA inhibitors as potential direct Tau modulators. More broadly, our work provides multiscale computational workflows for targeting IDPs and a transferable framework for studying ligand recognition in complex neurodegenerative disease targets.

P39

Differential Regulation of p27 Stability by VHL Isoforms and the Cancer-Associated p27E40K Mutation

Alessia Strappazzon, Lara Abad, Marianna Bussolino, Silvio C.E. Tosatto, Emanuela Leonardi

Department of Biomedical Sciences, University of Padova, Padova, Italy

VHL tumour suppressor protein (pVHL) is an E3 ubiquitin ligase whose best studied function is to bind to and down-regulate the hypoxia-inducible factor (HIF) family of oxygen-dependent transcription factors. Recent studies reported that pVHL also exerts numerous non canonical, HIF independent functions that are crucial for cellular homeostasis and, when lost, may contribute to drive cancer development.
Previous evidence indicated that all CDKN1 proteins (p21, p27, p57) can associate in vitro with VHL and that a cancer associated mutation in p27, namely p27E40K, alters this interaction. However, the biological significance and the impact of these associations on cancer development remain unclear.
To investigate if this mutation affects the interaction of p27 and VHL isoforms within mammalian cells, we used HEK293T and CAKI-1 cell lines along with co-immunoprecipitation. p27 protein turnover and cell proliferation were assessed using Cycloheximide (CHX) assay and MTT assay respectively.
Results revealed that both wild-type p27 and p27E40K are able to interact with VHL in vivo, although they display distinct protein turnover rates. Transfection of VHL30 isoform did not affect p27 stability, while the transfection of VHL19 increased it. However, co-expression of p27E40K with either VHL30 or VHL19 resulted in unchanged or significantly increased stability, respectively. While co-overexpression of p27 with VHL isoforms did not alter cell proliferation, the simultaneous transfection of p27E40K and VHL19 isoform appeared to increase it.

Together, these findings reveal that p27 and pVHL interact within cells, suggesting a novel mechanism by which VHL isoforms differentially regulate p27 stability, with potential implications for cancer-associated mutations.

P40

Ensemble structural analysis reveals how flanking regions modulate LIR motifs in intrinsically disordered autophagy receptors

Mattia Utichi[1,2], Henri-Baptiste Marjault[1], Martina Bellini[1,2], Jordi Hintzen[3,4], Laura Mattioli[1,5], Matteo Arnaudi[1,2], Alesia A. Tietze[3,4], Matteo Tiberti[1], Emiliano Maiani[1,6,7], Oana N. Antonescu[1,8], Elena Papaleo[1,2]

[1]Cancer Structural Biology, Danish Cancer Institute, Strandboulevarden 49, 2100, Copenhagen, Denmark
[2]Cancer System Biology, Section of Bioinformatics, Health and Technology Department, Technical University of Denmark, Lyngby, Denmark
[3]Department of Chemistry and Molecular Biology, Wallenberg Centre for Molecular and Translational Medicine, University of Gothenburg, Medicinaregatan 7B, 413 90 Gothenburg, Sweden,
[4]Center for Antibiotic Resistance Research in Gothenburg, The University of Gothenburg, Box 100, 405 30 Gothenburg, Sweden
[5]Current workplace: Section of Medical Biotechnology, Department of Bioengineering, Technical University of Denmark, Lyngby, Denmark
[6]Current workplace: Departmental Faculty of Medicine, UniCamillus—Saint Camillus International University of Health and Medical Sciences, Rome, Italy,
[7]Current workplace: IRCCS San Camillo Hospital, Venice, Italy
[8]Current workplace: ExpreS2ion Biotechnologies, Agern Allé 1, Hørsholm, Denmark

Short linear motifs embedded in intrinsically disordered regions (IDRs) mediate many regulatory protein-protein interactions. In selective autophagy, LC3-interacting regions (LIRs) enable cargo receptors to bind ATG8 family proteins and recruit autophagosomes. While the structural determinants of the canonical LIR core are well characterized, the contribution of the surrounding disordered sequence remains less understood. We investigated how sequence context and local structural elements within IDRs influence LIR-mediated interactions. Using the FYCO1-LC3B complex as a model system, we combined molecular dynamics simulations with ensemble-based computational mutagenesis and experimental validation. The simulations reveal a C-terminal α-helical element adjacent to the LIR core that remains stable in the bound ensemble and extends the interaction interface. Contact analysis identifies a network of hydrophobic and electrostatic interactions involving residues within this region, including a persistent interaction between FYCO1 E1287 and LC3B R70. These predictions were tested experimentally through mutagenesis, cellular co-immunoprecipitation assays, and quantitative binding measurements using microfluidic diffusional sizing. Disruption of the canonical LIR core abolishes binding, whereas mutations in the flanking region modulate interaction strength, indicating that these residues contribute to stabilizing the complex without replacing the canonical interaction. To examine whether similar features occur in other systems, we screened experimentally validated LIR-containing proteins for predicted secondary-structure propensity adjacent to the motif. This analysis identified several candidates

with potential helical flanking regions. As a representative case, we analyzed the selective autophagy receptor NDP52 and found that residues within its C-terminal helical extension contribute to LC3C binding. Overall, these results indicate that while the canonical LIR core provides the primary anchoring interaction, structural elements in the surrounding intrinsically disordered regions can modulate ATG8 recognition. Ongoing work combines computational predictions with cellular perturbation assays to investigate how these interactions influence autophagy-related cellular phenotypes.

P41

CVFormer: Data-Driven Collective Variables via Transformer Autoencoders

Guglielmo Tedeschi, Vojtěch Spiwok

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.

P42

Inhibitory Potential of LEA Proteins from Ramonda serbica on the Aggregation of A53T α-Synuclein

Marija Vidović[1], AnaPantelić[1], Tatiana Ilina[1], Mila Ljujić[1], Miroslav Gancar[2], Zuzana Bednarikova[2]

[1]Institute of Molecular Genetics and Genetic Engineering, University of Belgrade, Vojvode Stepe 444a, 11042 Belgrade, Serbia
[2]2 Department of Biophysics, Institute of Experimental Physics of the Slovak Academy of Sciences, Watsonova 47, 040 01 Košice, Slovakia

Extreme loss of cellular water, or desiccation (5–10% relative water content), typically leads to protein denaturation, aggregation, and the loss of membrane integrity. The ancient resurrection plant Ramonda serbica survives long periods of desiccation and fully recovers metabolic functions upon rehydration. A key role in this desiccation tolerance is played by Late Embryogenesis Abundant (LEA) proteins, a heterogeneous group of mostly intrinsically disordered proteins. These LEA proteins exhibit high structural plasticity, remaining disordered when fully hydrated but undergoing a structural transition into α-helical conformations during water loss.
This study explored the potential of LEAPs to interact with and stabilize aggregation-prone proteins, such as the human A53T α-synuclein mutant. This specific mutation is of high clinical significance as it is linked to early-onset familial Parkinson's disease and significantly accelerates the rate of α-synuclein aggregation into toxic fibrils. Utilizing Thioflavin T (ThT) fluorescence assays, we demonstrated that RsLEA30 effectively inhibits fibril formation. This was evidenced by significantly lower ThT fluorescence intensity in the presence of RsLEA30 at molar ratios of 1:1 and 1:3 compared to A53T α-synuclein alone. Atomic force microscopy further confirmed that RsLEA30 completely prevents fibril formation.
Furthermore, immunofluorescence staining in HEK 293 cells showed that the presence of RsLEA30 during fibril formation reduced the cellular uptake of these aggregates. These findings suggest that the structural flexibility of plant LEA proteins allows them to interact with the dynamic aggregation-prone conformations of A53T α-synuclein. This research provides a foundation for developing a new class of therapeutic stabilizers for protein-folding disorders.
Keywords: Ramonda serbica, LEA proteins, A53T α-synuclein, aggregation inhibition, protein structural plasticity, ThT assay.

Acknowledgement:
This research was funded by the Science Fund of the Republic of Serbia–RS (PROMIS project LEAPSyn-SCI, grant no. 6039663) and by the Ministry of Education, Science and Technological Development, the Republic of Serbia (Contract No. 451-03-136/2026-03/200042) and multilateral projects with Republic of Slovakia (337-00-3/2024-05/17) and Czech Republic, NEPTUNE (DS2; APVV-DS-R-24-0060). A.P. wishes to acknowledge the support of COST Action CA21160 (ML4NGP) for approving Short-Term Scientific Missions in Košice.

P43

On the menu today: NMR studies of the intrinsically disordered actin-binding region of Drebrin

Soma Varga, Julie Maibøll Kaasen, Zoltán Gáspári, Bálint Ferenc Péterfia, Frans A. A. Mulder

Drebrin (developmentally regulated brain protein) is a key component of the postsynaptic density (PSD), where it contributes to the structural organization of synaptic protein networks. Through its interaction with the postsynaptic scaffold protein Homer and its association with actin filaments, Drebrin helps anchor the PSD protein complex to the cytoskeleton and plays an important role in synaptic architecture and function. Structurally, Drebrin contains several unusual elements, including an evolutionarily divergent actin-depolymerizing factor homology (ADFH) domain that has lost strong actin-binding capability, as well as a Single Alpha-Helix (SAH) motif embedded within long intrinsically disordered regions.
Previous in vivo studies suggest that a specific disordered segment of Drebrin is crucial for binding filamentous actin (F-actin). However, despite its functional importance, the molecular and atomic-level details of the Drebrin–F-actin interaction remain unresolved. A detailed structural characterization of this interaction is therefore essential for understanding how Drebrin contributes to cytoskeletal organization within the PSD.
To address this challenge, we designed an intrinsically disordered Drebrin construct, D233, encompassing the region implicated in actin binding. Using multidimensional nuclear magnetic resonance (NMR) spectroscopy, we performed a near-complete backbone resonance assignment of the construct. In particular, 3D (HN)CO(CO)NH experiments were employed to facilitate the assignment of residues within this highly flexible region, which is typically difficult to characterize using conventional structural methods.
The resulting resonance assignments provide an important foundation for subsequent structural and interaction studies aimed at defining the binding interface between Drebrin and F-actin. This work establishes the experimental framework necessary for investigating the conformational properties of this disordered Drebrin segment and for mapping its interaction with actin at atomic resolution.
Overall, our study represents a critical step toward elucidating the molecular mechanisms by which Drebrin associates with the actin cytoskeleton and contributes to the organization of postsynaptic structures. These insights may ultimately advance our understanding of synaptic protein network assembly and the regulation of cytoskeletal dynamics in neuronal cells.

P44

Transparent AI-Assisted Discovery of Antibody–Amyloid Interaction Studies from Abstract-Level Evidence

Damian Michal Wilary, Dawid Sonor Kubkowski, Valentín Iglesias, Michal Burdukiewicz

Medical University of Bialystok, Poland

Background / Motivation. Scientific publications remain the primary vehicle for communicating research findings, yet key information is embedded in narrative text that is difficult to systematically search, compare, or reuse. Literature discovery therefore often relies on keyword-based retrieval followed by manual screening of large candidate sets of abstracts. This process is time-consuming and susceptible to biases introduced by opaque search tools or reputation-based heuristics. At the same time, many articles remain behind paywalls, making titles, abstracts, and metadata the most consistently accessible signals for literature discovery.

Methods. We present DeepSkim, an interpretable AI-assisted system designed to help researchers prioritize candidate publications. DeepSkim operates solely on pre–paywall publication elements (title, abstract, author list, and references) and learns a representation of relevance from a small set of user-labeled examples. We evaluate the approach on a curated dataset of publications describing antibody-mediated effects on amyloid aggregation, developed for the AmyloGraphAB database. Inclusion in this dataset requires strict experimental criteria, making relevance assessment challenging using abstract-level information alone.
Results. Despite these constraints, DeepSkim effectively prioritizes relevant studies within keyword-retrieved candidate sets. In this screening scenario, the system reduces the number of abstracts requiring manual inspection by up to 70% (measured with a target 95% recall) while maintaining coverage of studies meeting the strict inclusion criteria. Current work focuses on developing interpretable explanations that allow researchers to inspect the signals underlying the model's relevance predictions.
Conclusions / Significance. DeepSkim demonstrates that interpretable AI models can move beyond simple similarity search toward knowledge discovery workflows that substantially reduce manual literature curation effort while preserving transparency and reproducibility.

P45

Structural Roles of IDPs in Dynamic PPI Networks

Nenad Vilendečić[1], Milana Grbić[1], Miloš Radovanović[2], Dragan Matić[1]

[1]Faculty of Natural Sciences and Mathematics, University of Banja Luka,
[2]Faculty of Sciences, University of Novi Sad

Intrinsically disordered proteins (IDPs) play important roles in numerous biological processes due to their structural flexibility and ability to interact with multiple binding partners. Their properties are primarily investigated using molecular-level experimental and computational approaches. At the systems level, however, IDPs can be studied within the framework of protein-protein interaction (PPI) networks, where they have been reported to preferentially occupy central and highly connected positions.
Traditional PPI networks are typically represented as static graphs in which interactions are aggregated into a single, time-independent structure. Although such representations are useful for many analytical purposes, they cannot capture the temporal variability of protein interactions observed in living cells. To overcome this limitation, dynamic PPI networks have been introduced by incorporating temporal information into network models, allowing interactions to appear, disappear, or change over time.
In this study, we extend the analysis of IDPs to the dynamic setting and investigate their behavior in dynamic PPI networks. In particular, we examine the activity patterns of IDPs across different time points and compare them with those of non-IDP proteins. Furthermore, several network-based measures are employed to assess the centrality of IDPs and to evaluate their structural roles over time. The temporal evolution of IDP neighborhoods is also analyzed in order to determine whether IDPs tend to preserve common interaction partners throughout the observed time period.

The behavior of IDPs is analyzed using dynamic PPI networks obtained from five different data sources. The experimental results reveal several distinctive characteristics of IDPs within dynamic PPI networks. On average, IDPs exhibit higher activity rates than non-IDP proteins across time points. Furthermore, IDPs attain centrality values that exceed network averages, and their interaction neighborhoods demonstrate a certain degree of temporal stability.

P46

Cryo-EM Analysis of Alpha-Synuclein Fibril Structures Formed under Different Temperature Conditions

Mantas Ziaunys, Darius Sulskis, Kamile Mikalauskaite, Vytautas Smirnovas

Life Sciences Center, Vilnius University, Vilnius, Lithuania

Protein aggregation into amyloid fibrils is linked to various neurodegenerative disorders, including Alzheimer's and Parkinson's diseases. These conditions affect over 50 million people worldwide, creating a significant burden on healthcare systems. Despite decades of research, effective treatments are still available for only a small fraction of disorders, likely due to an incomplete understanding of the amyloid aggregation process and the environmental factors that influence it. Amyloid formation is influenced by various environmental parameters such as pH, ionic strength, and temperature. Changes in these factors significantly impact the mechanism of aggregation and the resulting fibril characteristics, including morphology, stability, and cytotoxicity.
This study investigates the influence of temperature on the aggregation of alpha-synuclein (α-syn), a 140-amino acid protein whose accumulation is a hallmark of Parkinson's disease. While temperature is frequently modulated to accelerate aggregation assays, its role in determining specific structural outcomes is often overlooked. We examined α-syn fibril formation across a wide temperature range and characterized the resulting aggregates using atomic force microscopy, infrared spectroscopy and cryogenic electron microscopy (Cryo-EM).
Our results reveal that α-syn fibril structures differ significantly when prepared below versus above physiological temperatures. Cryo-EM analysis identified a temperature-related transition between the dominant fibril species, alongside several unique aggregate types. These findings demonstrate that temperature is a critical parameter in determining the α-syn aggregation pathway and highlight the caution that should be taken when performing assays under different thermal conditions.

P47

From Repeated Peptide Motifs to Functional Biomaterials: Recombinant Elastin-Like Polypeptides as Tunable Protein Polymers

Bilge Kalyoncu, Sedef Kırdar, Esra Yuca Yilmaz

Yildiz Technical University Department of Molecular Biology and Genetics, Istanbul, Turkiye

Elastin-Like Polypeptides (ELPs) are stimuli-responsive recombinant protein polymers derived from the intrinsically disordered regions of the natural extracellular matrix protein Tropoelastin. They consist of repetitive (VPGXG)n pentapeptide motifs, where X represents a guest residue that can modulate the physicochemical properties of the polymer. Owing to their tunable sequence composition and physicochemical properties, ELPs have been extensively investigated for biomedical applications such as drug delivery, wound healing, and tissue engineering.
A defining feature of ELPs is their reversible inverse temperature transition. Above a characteristic transition temperature, ELPs undergo phase separation, aggregation, or self-assembly. This phase behavior can be systematically tuned by altering the identity of the guest residue, the number of pentapeptide repeats, and the overall molecular architecture of the polymer. Consequently, the physical and functional properties of ELPs can be precisely adjusted to meet the requirements of specific biomedical applications. Genes encoding ELPs can be designed and expressed in Escherichia coli using standard recombinant DNA techniques. The resulting protein can be efficiently purified without chromatography by inverse transition cycling (ITC), a purification method that exploits the reversible phase transition behavior of ELPs. In this work, we used recombinant ELPs as modular building blocks for the development of stimuli-responsive biomaterials. A series of ELP variants with different repeat numbers and guest residues (V, I) were recombinantly produced and purified. These ELPs were subsequently used to construct functional nano- and micro-scale assemblies. Depending on sequence design and environmental conditions, the polypeptides formed nanoparticles and other supramolecular structures suitable for biomedical applications.
The ELP-mediated nanoparticles were further integrated into three-dimensional (3D) biomaterial structures. Antimicrobial agents were incorporated into the nanoparticle-functionalized matrices to enhance their biological functionality. The resulting 3D constructs exhibited antibacterial activity against both Gram-positive and Gram-negative bacteria. Due to their structural properties and antimicrobial functionality, these materials show strong potential for use as wound patches and wound dressing systems. In our ongoing studies, we are further expanding this platform toward the development of recombinant ELP-based three-dimensional cell culture systems to support advanced biomaterial and tissue engineering applications.

P48

AlphaFold 3 Modeling of the KfrA–KfrB–KfrC R751 System Suggests a Shared KfrB Interface and Ion-Sensitive KfrA Stability

Diego Javier Zea, Małgorzata Adamczyk

Université Paris-Saclay, CEA, CNRS, Institute for Integrative Biology of the Cell (I2BC), 91198 Gif-sur-Yvette, France
Warsaw University of Technology, Faculty of Chemistry, Chair of Drug and Cosmetics Biotechnology, Laboratory of Systems and Synthetic Biology (LSSB), 00-664 Warsaw, Poland

Background. The kfrA–kfrB–kfrC operon of the IncP-1β plasmid R751 has been implicated in plasmid maintenance and spreading antibiotic resistance [1,2]. Previous work showed that these proteins interact pairwise, primarily through KfrA–KfrB and KfrB–KfrC interactions, suggesting that KfrB could be connecting those proteins. KfrA-type proteins are DNA-binding proteins with long alpha-helical coiled-coil regions. Here, we use AlphaFold 3 (AF3) ensemble modeling to compare KfrA–KfrB,

KfrA–KfrC and KfrB–KfrC interactions, quantify interface variability, and prioritize residues for mutational testing.
Methods. We modeled pairwise interactions with AF3, generating 100 models per interaction using 20 different random seeds. KfrA R751 was modeled as a homodimer to capture its coiled-coil architecture. To compare interface agreement across models, we computed interaction contact maps using heavy-atom proximity and quantified model-to-model differences with Jaccard distances, visualized by multidimensional scaling. We also explored ionic conditions and the presence of small molecules and ions commonly observed in crystallographic structures, and analyzed residue-level contact accumulation in large AF3 ensembles of the full KfrA–KfrB–KfrC complex.
Results & Conclusions. Unexpectedly, given the experimental measures, all modeled interactions show relatively low ipTM values. We have observed an interaction between KfrA dimer and KfrC, showing low ipTM values and strong interface dispersion, consistent with a weak or heterogeneous binding mode. In contrast, the best-scoring KfrA dimer–KfrB models fall within a cluster of similarly scoring structures, suggesting a more reproducible interface. Interface inspection highlights KfrB residues R4, M8, N9 and Q15 as candidate determinants of binding, with positions 8–9 repeatedly implicated in contacts with KfrA in large full-complex ensembles. In pairwise KfrB–KfrC models, KfrB M8 and N9 contact KfrC, therefore using the same surface that engages KfrA. Separately, homodimer modeling is sensitive to ionic conditions: both the highly alpha-helical KfrA and KfrC dimers show improved pTM and ipTM values when 9 Na+ and 9 Cl- ions are included relative to the no-ion condition. This effect is particularly strong for the KfrC dimer: median values across 25 models are ipTM 0.40 and pTM 0.48 in the NaCl condition, compared with 0.09 and 0.24 without ions. Together, these results support targeted mutagenesis of KfrB to test whether a shared KfrB surface governs the interaction with KfrA and KfrC. In future work, we will evaluate whether similar ionic conditions improve AlphaFold 3 predictions for other coiled-coil and highly alphahelical assemblies.

References.
[1] Adamczyk M. et al. Microbiology 2006.
[2] Adamczyk M. et al. BMC Microbiology 2021.

P49

Analysis of bacterial amyloids potentially involved in neurodegeneration using updated AmyLoad database

Julia Cieszko[1], Michał Burdukiewicz[2], Jarosław Chilimoniuk[2], Valentín Iglesias[2], Author5_Ronja Tittel[3], Małgorzata Kotulska[1]

[1]Wroclaw University of Science and Technology,
[2]Medical University of Białystok, Clinical Research Centre,
[3]Brandenburg University of Technology Cottbus-Senftenberg,

The gut-brain axis is increasingly recognized as a factor in neurodegenerative diseases, with bacterial amyloids possibly modulating protein aggregation [1]. As amyloidogenesis understanding evolves, bioinformatic tools need updates for accurate analysis.
The AmyLoad database [2] was updated in a new version [3], reflecting current knowledge on peptides and their fibrillation. Sequences once labeled amyloidogenic solely due to β-structure were excluded. Classification now distinguishes aggregating, non-aggregating, and fibril-forming peptides, and the type of evidence (direct, e.g., microscopy, or indirect,e.g., ThT assay).
Peptides with confirmed fibril-forming ability were extracted from the updated dataset and searched for in bacterial taxa enriched in the microbiota of neurodegenerative patients.Their proteomes revealed proteins with sequences identical to those in AmyLoad. Some matched amyloid-forming segments and were linked to neurodegeneration through fibril formation or cross-seeding.
These findings show that updated bioinformatics tools combined with experimental data help explore the role of bacterial amyloids in neurodegeneration. This is a first step toward understanding their involvement, which may aid in developing therapies for currently incurable diseases.

[1] Wojciechowska, A. W., Wojciechowski, J. W., Zielinska, K., Soeding, J., Kosciolek, T., & Kotulska, M. (2024). Aggregating gut: on the link between neurodegeneration and bacterial functional amyloids. bioRxiv, 2024.11.26.624671.
[2] Wozniak, P. P., & Kotulska, M. (2015). AmyLoad: website dedicated to amyloidogenicprotein fragments. Bioinformatics, 31(20), 3395–3397.
[3] The provisional website can be accessed at: https://burdukiewicz.com/ramyload/

P50

IDP-Specific Coevolution and Dynamics-Aware Modeling Enable Prediction of Allosterically Competent Binding Sites

Kemal Demirtaş, Diego Javier Zea, Türkan Haliloğlu

Department of Chemical Engineering, Boğaziçi University, 34342 Bebek, Istanbul, Turkey;
Université Paris-Saclay, CEA, CNRS, Institute for Integrative Biology of the Cell (I2BC), 91198 Gif-sur-Yvette, France

Predicting functional binding sites on intrinsically disordered proteins (IDPs) remains challenging due to their conformational heterogeneity and lack of stable structure and pre-formed binding pockets. Here we present an IDP-specific coevolution- and dynamics-informed framework to identify potential binding residues by quantifying the collective information transfer capacity of residues within the IDP. We introduce SeqGNM-TE, a method that integrates Transfer Entropy (TE), an information theoretic measure of dynamic influence, with a Gaussian Network Model (GNM) constructed entirely from IDP-specific coevolutionary coupling scores.

SeqGNM-TE is distinguished from classical GNM by its use of coevolutionary coupling scores to construct the connectivity matrix. Obtaining reliable coevolution data has been a critical challenge in IDPs. Initial efforts utilized the EVcoupling pipeline, which performs well for globular proteins, but its performance is rather limited for IDPs. We now employ coevolution values derived from transcript-aware multiple sequence alignments using GaussDCA, as they can be more reliable on IDPs. This shift has produced remarkably improved results over the previous version.

Within this framework, GNM decomposes the system into dynamic modes, enabling the analysis of information propagation across distinct dynamic subsets. From the resulting transfer entropy network, we compute residue collectivity, which quantifies the ability of a residue to function as an information hub coordinating global dynamic communication. SeqGNM-TE generates a directional map of net information transfer flows across the IDP and identifies residues that exert strong dynamic influence over others, prioritizing them as potential binding hotspots.

Validation across a set of IDP–protein complexes demonstrates that these predicted binding residues engage their partners in poses that maximize cooperative dynamic association across the IDP–protein complex. These results demonstrate strong predictive power and highlight the interplay between evolutionary constraints and allosteric communication in defining functional binding sites. This perspective provides a dynamics-aware prediction of binding sites in IDPs.

Organization

This event is part of the activities of the COST Action ML4NGP, CA21160, supported by COST (European Cooperation in Science and Technology).

Patronage

Honorary patronage of the Mayor of Warsaw

Local partners and sponsors

Warsaw

Warsaw Tourism Organization

INSTITUTE OF BIOCHEMISTRY AND BIOPHYSICS POLISH ACADEMY OF SCIENCES

Ministry of Science and Higher Education Republic of Poland

WEKTORY NAUKI