European RosettaCon 2024

Knowledge Transfer for Modern Protein Modeling and Design

November 11th - 13th, 2024

Biocenter, University of Copenhagen

Hosted by ISBUC

Organising Committee:

Aleksandra Panfilova, University of Copenhagen
Marion Lucia Silvestrini, University of Copenhagen
Roland Alexander Pache, Novonesis
Johanna Katarina Sofie Tiemann, Novonesis
Ingemar André, Lund University
Che Yang, Novo Nordisk
Amelie Stein, University of Copenhagen

Hosted by:

The Integrative Structural Biology cluster at the University of Copenhagen

INTEGRATIVE STRUCTURAL BIOLOGY CLUSTER AT THE UNIVERSITY OF COPENHAGEN

Funded by:

Rosetta
Novo Nordisk Foundation

Sponsored by:

Novo Nordisk
Novonesis

Code of Conduct:

The 2024 European Rosetta Conference is committed to equal opportunity and respectful treatment for all and to providing an environment that encourages the free expression and exchange of scientific ideas. The conference is not an appropriate space for sensitive cultural or political discussions.

Program

Day 1
Monday, November 11th

8:30 - 9:00	Registration, coffee & croissants
9:00 - 9:10	Welcome
9:10 - 11:30	Session I - Enzymes
9:10-9:35	Donald Hilvert
9:35-9:45	Ramiro Illanes Vicioso
9:45-9:55	Abbie Lear
9:55-10:05	Núria Mimbrero Pelegrí
10:05-10:15	Sigrid Kaltenbrunner
10:15-10:25	Georg Künze
10:25-10:35	Cesar Antonio Ramirez-Sarmiento
10:35 - 10:55	**Coffee Break**
10:55 - 11:05	Giulia Peteani
11:05 - 11:30	Sarel Fleishman
11:30 - 14:50	Session II - Intrinsically disordered proteins, condensates & MD simulations
11:30 - 11:55	Kresten Lindorff-Larsen
11:55 - 12:05	Rasmus Krogh Norrild
12:05 - 12:30	Alena Khmelinska
12:30 - 13:30	**Lunch**
13:30 - 13:55	Matt O'Meara
13:55 - 14:05	Mykhailo Girych
14:05 - 14:30	Joseph Rogers
14:30 - 14:50	Coffee Break

14:50 - 16:20	**Session III: Small molecule interaction and design**
14:50 - 15:15	John Karanicolas
15:15 - 15:25	Niklas Gesmar Madsen
15:25 - 15:35	Christoffer Norn
15:35 - 15:45	Daniele Granata
15:45 - 15:55	Deniz Akpinaroglu
15:55 - 16:20	Rocco Moretti
16:20 - 17:00	Poster Flash Talks
17:00 - 20:00	Poster Session I - Even poster numbers
	With pizzas and drinks

Day 2
Tuesday, November 12th

9:00 - 10:25	**Session IV - ML & Protein design I**
9:00 - 9:25	Jens Meiler
9:25 - 9:50	Dek Woolfson
9:50 - 10:00	Sofia Andersson
10:00 - 10:25	Elodie Laine
10:25 - 10:50	Coffee Break
10:50 - 14:40	**Session V - Protein-Protein-Interactions & -Complexes**
10:50 - 11:15	Tina Perica
11:15 - 11:25	Martin Pačesa
11:25 - 11:35	Valeriia Hatskovska
11:35 - 11:45	Amijai Saragovi
11:45 - 11:55	Fabio Parmeggiani
11:55 - 12:20	Ora Schueler-Furman
12:20 - 13:20	**Lunch**
13:20 - 13:45	Amy Keating
13:45 - 13:55	Lee Schnaider
13:55 - 14:05	Océane Follonier
14:05 - 14:15	Alina Konstantinova
14:15 - 14:40	Tanja Kortemme
14:40 - 15:00	Coffee Break

15:00 - 16:25	Session VI - Antibodies
15:00 - 15:25	Clara Schoeder
15:25 - 15:50	Pietro Sormanni
15:50 - 16:00	Britnie Carpentier
16:00 - 16:25	Possu Huang
16:25 - 17:00	Poster Flash Talks
17:00 - 18:25	Poster Session II - Odd poster numbers
19:00 - 22:00	Gala dinner
	Restaurant Llama, Lille Kongensgade 14, 1074 København

Day 3
Wednesday, November 13th

9:00 - 12:00	Session VII - ML & Protein design II
9:00 - 9:25	Noelia Ferruz
9:25 - 9:35	Matteo Cagiada
9:35 - 9:45	Max Beining
9:45 - 10:10	Mohammed Alquraishi
10:10 - 10:30	**Coffee Break**
10:30 - 10:55	Joanna Slusky
10:55 - 11:05	Dominique Fastus
11:05 - 11:15	Thea Klarsø Schulze
11:15 - 11:25	Katharina Bachschwöller
11:25 - 11:35	Andrea Hunklinger
11:35 - 12:00	Sergey Ovchinnikov
12:00 - 12:20	Closing remarks
12:20 - 13:15	Lunch

13:15 - 14:30 Parallel workshop session 1, 2, 3

Workshop 1: Industry careers panel discussion

Moderated by Roland Pache (Novonesis)
Lundbeck Auditorium

With Che Yang (Novo Nordisk), Johanna Tiemann (Novonesis), Alexandra Chivu (Flagship Pioneering), John Karanicolas (AbbVie), Jonathan Ziegler (Cradle), Dana Cortade (Align to Innovate), Marloes Arts (Genmab), Zander Harteveld (Orbis Medicines)

Workshop 2: Adapting DNA synthesis screening to a de novo world

Sam Curtis
4-0-24

Many DNA synthesis providers routinely use screening algorithms to detect and investigate orders containing regulated pathogens and other potentially dangerous sequences. However, synthetic DNA providers and public health officials have expressed concerns that existing screening systems are increasingly ill-equipped for state-of-the-art generative models, heavily modified sequences, and de novo molecules. In this discussion, we'll explore the regulatory architecture surrounding DNA synthesis, the screening systems that exist today, and how the protein design community can contribute to the development of next-generation screening approaches.

Workshop 3: Sharing protein AI models using huggingface

Simon Duerr
4-0-10

Reproducing published computational workflows can be difficult because often even if the code is available it can be hard to figure out the exact environment and parameters to use. Webservers while practical often disappear from the web over time. This workshop will introduce methods how to quickly turn any kind of model pipeline into an easy to use webapp that other researchers can quickly deploy on their local machines and that can also be used directly in a web browser. The workshop will introduce you to the concept of a Docker container, the UI framework Gradio and how to deploy your model on HuggingFace Spaces. We will also demonstrate how one can interact with the deployed model programmatically by creating a PyMol extension using a REST Api. Target audience: People with some familiarity with programming or bash scripting that want to share model pipelines on the web. Required hardware: Laptop with code editor installed and access to a Linux terminal or alternatively a web browser.

14:30 - 14:50 Coffee Break

13:15 - 16:25 Parallel workshop session 4 (With coffee break)

Workshop 4: Rosetta Data Bazaar Hack-a-thon (3h)

Matt O'Meara
4-0-32

In teams of 2-3 participants we will collaboratively curate and make available biomolecular structure and activity datasets for the Rosetta HuggingFace (https://huggingface.co/RosettaCommons). Each team will select a published or in house dataset developed or used by Rosetta/structural biology community and create a HuggingFace Dataset so that it can be loaded for machine learning with a single line of code. Depending on the dataset, curation may require some simple scripting, but participants are welcome to participate on a team even if they have limited coding experience. At least one lap-top is required per-team.

14:50 - 16:25 Parallel workshop session 5, 6, 7

Workshop 5: Academic careers panel discussion

Moderated by Amelie Stein

Lundbeck Auditorium

With Jens Meiler, Ora Schueler-Furman, Kresten Lindorff-Larsen, Amy Keating, Tina Perica, Ajasja Ljubetič.

Workshop 6: Innovating on the dataset-to-model design cycle for protein function prediction

Dana Cortade
4-0-24

Align to Innovate creates living datasets for machine learning and deploys them via community benchmarking opportunities. If you took control of this cycle to generate better protein sequence-to-function prediction models, what kinds of protein libraries would you design to test? What published models would you want to see go head-to-head? How would you maximize the amount of information gleaned per variant tested (or per dollar spent on synthesis)? Join us to map out how the field should handle continuous evaluation and benchmarking of methods as more data comes into the ecosystem.

Workshop 7: Frame2seq: generalizable method for tuning multi-state conformational equilibria

Deniz Akpinaroglu
4-0-10

While recent advances in computational protein design have led to an increased experimental success rate designing stable single-state proteins, reliably tuning the dynamics of multi-state proteins remains a challenge. The workshop will demonstrate a generalizable method for tuning multi-state conformational equilibria using Frame2seq. We will show that model scores are consistently predictive of changes to the ratio of conformational populations. And we will use Frame2seq to identify mutation sites that will result in the most significant shits to conformational switch equilibria as computationally validated with AlphaFold2.

Keynote Speakers

1. Alena Khmelinska: Tales of Dynamic Protein Assembly Design
2. Amy Keating: Prediction and design of protein-peptide interactions
3. Clara Schoeder: Establishiment of a computational vaccine design pipeline for pandemic preparedness
4. Dek Woolfson: From peptides to proteins to functions by design
5. Donald Hilvert: Designing artificial enzymes de novo
6. Elodie Lane: From sequences to fitness and motions, protein language models to the rescue?
7. Jens Meiler: How artificial intelligence is reshaping protein structure prediction and therapeutic design – from small molecules to new modalities
8. Joanna Slusky: Protein Design Insights from Large Datasets
9. John Karanicolas: Using generative AI to design small molecules for drug discovery
10. Joseph Mathew Rogers: Targeting disordered protein using de novo designed proteins
11. Kresten Lindorff-Larsen: Towards computational design of disordered proteins and condensates
12. Matt O'Meara: Docking to Novel pocKets (DoNK): Charting Virtual Bioactivity Space Through Large-Scale Protein Design and Molecular Simulation
13. Mohammed Alquraishi: Some observations on how AlphaFold predicts structures and how it learns to predict structure
14. Noelia Ferruz: Design of functional enzymes with conditional language models
15. Ora Schueler-Furman: Different ways to bind, common way to model?
16. Pietro Sormanni: Computational Strategies for Antibody Design and Developability Optimization
17. Possu Huang: A general platform for targeting MHC-II antigens via a single loop
18. Rocco Moretti: Crowdsourcing Small Molecules: Evaluating Foldit as a platform for drug design
19. Sarel Fleishman: Computational design of functional repertoires of enzymes and antibodies
20. Sergey Ovchinnikov: Inverting Protein Structure Prediction models for protein design.
21. Tanja Kortemme: Expanding sequence & structure space, conformational switches, and synthetic cellular signaling
22. Tina Perica: Functional crosstalk between EGF and insulin signalling

1. Alena Khmelinska: Tales of Dynamic Protein Assembly Design

Ludwig Maximilian University of Munich
akhmelin@cup.lmu.de

Recent computational methods have been developed for designing novel protein assemblies with atomic-level accuracy, yet several aspects of current methods limit the structural and functional space that can be explored. I will share with you our ongoing efforts in diversifying the structural repertoire of protein assemblies and developing strategies to gain control over assembly dynamics.

2. Amy Keating: Prediction and design of protein-peptide interactions

MIT Biology
keating@mit.edu

Models trained on protein sequences and/or structures have led to exciting breakthroughs in computational structural biology. We are interested in how such models perform for protein-peptide interactions and how they might be further adapted for this specific task. I will discuss our work investigating where information comes from when using AlphaFold to dock peptides, and our latest results scoring and designing protein-peptide complexes.

3. Clara Schoeder: Establishiment of a computational vaccine design pipeline for pandemic preparedness

Leipzig University
clara.schoeder@medizin.uni-leipzig.de

Computational protein design has become a standard technology to stabilize viral glycoproteins in their prefusion conformation - the conformation necessary to elicit neutralizing and protective antibody responses. In this project, we want to take this approach one step further and systematically implement protocols and test their performance on a given prefusion stabilization task with experimental validation. Head-to-head we compared protocols from Rosetta and AI-driven sequence design.

4. Dek Woolfson: From peptides to proteins to functions by design

University of Bristol
d.n.woolfson@bristol.ac.uk

It is now possible to generate many stable peptide assemblies and proteins from scratch using rational and computational approaches. One new challenge is to move past structures found in nature and target the ‘dark matter of protein space’; that is, structures that should be possible from chemistry and physics, but which biology seems to have overlooked. This talk will illustrate what is currently possible in this nascent field using de novo designed coiled-coil peptides and proteins.

I will describe our “toolkit” of de novo coiled-coil assemblies, and how we are converting these peptides bundles and barrels into single-chain proteins through rationally seeded computational protein design. Then I will turn to subcellular applications. I will describe two new designs for (i) de novo cell-penetrating peptides, and (ii) high-affinity kinesin-binding peptides, and how these can be combined to hijack and control active motor proteins in living cells.

5. Donald Hilvert: Designing artificial enzymes de novo

ETH Zurich
donald.hilvert@org.chem.ethz.ch

Enzyme design represents a formidable challenge. We do not fully understand the rules of protein folding, and our knowledge of structure-function relationships in these macromolecules is at best incomplete. Recent progress in combining computational and evolutionary approaches for the design of artificial metalloenzymes will be discussed, together with insights into enzyme function gained from studies of the engineered catalysts.

6. Elodie Lane: From sequences to fitness and motions, protein language models to the rescue?

Sorbonne University
elodie.laine@sorbonne-universite.fr

I will present our latest work for addressing two important questions for protein engineering and human medicine. What is the impact of single-point mutations on protein functioning? How do protein move and deform to perform their functions? I will highlight the complementarity between protein language model (pLM-based predictors and evolutionary- or physics-based approaches. I will discuss some limitations linked to biases in the pLM representation spaces and the ground truth experimental data.

7. Jens Meiler: How artificial intelligence is reshaping protein structure prediction and therapeutic design – from small molecules to new modalities

Vanderbilt and Leipzig Universities
jens@meilerlab.org

AlphaFold is revolutionizing protein structure prediction. Not because of its increased accuracy in comparison to the best predictions of prior methods, but because of consistent accuracy across all folded, well-structured proteins. I will discuss some of the remaining challenges such as predicting all biologically relevant conformations of flexible membrane proteins including transporters, ion channels, or receptors.

With the availability of highly accurate structural models for most proteins, structure based drug discovery is experiencing a renaissance. The availability of large 'make-on-demand' compound libraries coupled with computational ultra-largelibrary screening fundamentally changes the paradigm in (academic) probe and drug development projects to 'in silico' first! I will introduce these concepts and detail several new algorithms to accomplish these tasks.

While AlphaFold is close to a golden bullet for protein structure prediction, computational design of protein and peptidetherapeutic candidates is much more challenging even with the use of artificial intelligence. One challenge is that the desired goal is function, the design algorithm focuses on structure implying that it might have the target function. A second challenge is the inclusion of chemical space with limited training data such as non-natural amino acids. I will give an overview of several new algorithms developed by us and others combined with illustrative applications.

8. Joanna Slusky: Protein Design Insights from Large Datasets

University of Kansas
slusky@ku.edu

Protein design relies on a deep understanding of the mimicked protein category. By constructing a large dataset of outer membrane proteins, we discovered features of outer membrane protein biogenesis and evolution. With a second large dataset—of metalloproteins—we revealed key differences between metal binding sites that can and can't catalyze reactions.

9. John Karanicolas: Using generative AI to design small molecules for drug discovery

Fox Chase Cancer Center
john.karanicolas@fccc.edu

We describe a new AI-based technique for producing novel chemical structures in silico. Through transfer learning, we demonstrate that the model can be iteratively refined to yield outputs that optimize for specific criteria, including 3D shapes. We applied the model to produce new chemical structures that recapitulate the biological activity of natural products, using much simpler readily-accessible chemical scaffolds.

10. Joseph Mathew Rogers: Targeting disordered protein using de novo designed proteins

University of Copenhagen
joseph.rogers@sund.ku.dk

Human proteins, to a surprising degree, lack folded structure. Lacking structure does not mean lacking function: these regions are important for human biology and feature prominently in proteins associated with disease. However, these dynamic, disordered peptide chains are notoriously hard to bind, and there are few research tools or drug leads targeting these regions. I will descrive our hallucination and RFDiffusion efforts to design de novo proteins that bind disordered regions, by inducing them to fold.

11. Kresten Lindorff-Larsen: Towards computational design of disordered proteins and condensates

University of Copenhagen
lindorff@bio.ku.dk

Intrinsically disordered proteins and regions (collectively IDRs) and protein with long flexible regions are pervasive across proteomes, help shape biological functions, and are involved in numerous diseases. IDRs populate a diverse set of transiently formed structures yet defy commonly held sequence-structure-function relationships. Recent developments in structure prediction and design of folded protein have led to the ability to predict the three-dimensional structures of folded proteins at the proteome scale and to design sequences that fold into specific three-dimensional structures. In contrast, knowledge of the conformational properties of IDRs is scarce, in part because the sequences of disordered proteins are poorly conserved and because only few have been characterized experimentally. In my talk I will describe how we can combine molecular simulations and machine learning methods to study the relationship between sequence, conformational properties, and functions of IDRs.

First, I will describe how we have used experimental data on more than 100 different proteins to learn a coarse-grained molecular energy function (CALVADOS) to predict conformational properties of disordered proteins including their propensity to undergo phase separation. Second, I will briefly describe how CALVADOS makes it possible to perform large-scale simulations to learn the relationship between sequence, structure, and function of IDRs at the proteome level. Third, I will present work on how we can use the information encoded in CALVADOS to design disordered proteins with desired conformational properties. Finally, I will show how we have used molecular simulations to train a model to predict the relationship between protein sequence and propensities to undergo phase separation.

12. Matt O'Meara: Docking to Novel pocKets (DoNK): Charting Virtual Bioactivity Space Through Large-Scale Protein Design and Molecular Simulation

University of Michigan
maom@umich.edu

Large scale chemical foundation models will require synthesizing empirical and theoretical knowledge. To explore the utility of simulation data we have generated DoNK, dataset of 1,000,000 in-stock ligands docked to 10,000 designed binding sites. We will describe our experience with using it to pre-train ML models for a range of virtual screening and medicinal chemistry prediction tasks.

13. Mohammed Alquraishi: Some observations on how AlphaFold predicts structures and how it learns to predict structure

Columbia University
m.alquraishi@columbia.edu

In this talk I will discuss some recent evidence regarding the degree to which AlphaFold appears to learn to do implicit physics, as well as how this knowledge is acquired during the training process. Time permitting I will also discuss differences between AlphaFold 2 and 3 with regards to the question of implicit physical knowledge.

14. Noelia Ferruz: Design of functional enzymes with conditional language models

Centre for Genomic Regulation
noelia.ferruz@crg.eu

We report the training of conditional language models for the generation of proteins in unseen regions of the protein space. We test experimentally carbonic anhydrases and lactate dehydrogenases that share little similarity with natural proteins yet preserve their activity levels. We additionally describe two new models for the design of new-to-nature enzymes and binders and their dependency on data quality. We introduce techniques for the continual learning of protein language models.

15. Ora Schueler-Furman: Different ways to bind, common way to model?

The Hebrew University of Jerusalem
ora.furman-schueler@mail.huji.ac.il

Recent advances in modeling and deep learning have made it possible to significantly improve the modeling of interactions, including those mediated by short motifs, at least those that form a defined structure. Many interactions however can be strong but still retain considerable entropy. How well can these be characterized using the protocols developed for and trained on interactions with defined structure? I will present several examples and discuss where we stand and where we can go.

16. Pietro Sormanni: Computational Strategies for Antibody Design and Developability Optimization

University of Cambridge
ps589@cam.ac.uk

Antibodies are indispensable in research, diagnostics, and as therapeutics. Despite significant advancements in antibody discovery and optimization technologies, challenges remain, particularly in the efficient targeting of predetermined epitopes and the simultaneous optimization of multiple biophysical traits. Traditional screening methods can be labor-intensive and are ineffective at navigating the complex trade-offs between critical properties such as affinity, stability, and solubility. Computational approaches present a promising solution, offering speed, cost-effectiveness, and resource efficiency. In this presentation, I will explore emerging computational methods for antibody design, which enable precise targeting of specific epitopes, accurate prediction of nativeness, nanobody humanization, and the optimization of developability potential through the simultaneous enhancement of multiple biophysical properties. These approaches hold the potential to significantly streamline antibody discovery and optimization, paving the way for rapid advancements in therapeutic and diagnostic applications.

17. Possu Huang: A general platform for targeting MHC-II antigens via a single loop

Stanford University
possu@stanford.edu

Class-II major histocompatibility complexes (MHC-IIs) are central to the communications between CD4+ T cells and antigen presenting cells (APCs), but intrinsic structural features associated with MHC-II make it difficult to develop a general targeting system with high affinity and antigen specificity. Here, we introduce a protein platform, Targeted Recognition of Antigen-MHC Complex Reporter for MHC-II (TRACeR-II), to enable the rapid development of peptide-specific MHC-II binders.

18. Rocco Moretti: Crowdsourcing Small Molecules: Evaluating Foldit as a platform for drug design

Vanderbilt University
rmorettiase@gmail.com

Foldit is an online citizen science game which allows members of the public to participate in computational structural biology research. We have recently added the ability to design druglike small molecules into the Foldit interface. Foldit game players have successfully participated in several different drug design programs, including CACHE, an independent assessment of computational small molecule design programs.

19. Sarel Fleishman: Computational design of functional repertoires of enzymes and antibodies

Weizmann Institute of Science
sarel.fleishman@weizmann.ac.il

We develop strategies that combine phylogenetic analysis, Rosetta atomistic calculations, and machine learning to optimise natural proteins. Thousands of protein scientists have used these strategies to generate stable therapeutic enzymes, vaccine immunogens, therapeutic antibodies, and membrane proteins for a range of needs in basic and applied research. A stabilised malaria vaccine immunogen designed in our lab has recently entered phase II clinical trials in West Africa. We now present a machine-learning strategy to design and economically synthesize millions of active-site variants that are likely to be stable, foldable, and active. We applied this strategy to design large libraries of enzymes and fluorescent proteins, and experimental screening revealed more than 10,000 functionally diverse proteins in each set.

20. Sergey Ovchinnikov: Inverting Protein Structure Prediction models for protein design.

Massachusetts Institute of Technology
so3@mit.edu

For this talk I'll describe some recent advances in inverting protein structure prediction models for protein design, contrasting them to more recent diffusion and flow based methods. I'll go through some examples of designing cyclic peptides, large new folds and protein binders.

21. Tanja Kortemme: Expanding sequence & structure space, conformational switches, and synthetic cellular signaling

UCSF
tanjakortemme@gmail.com

I will discuss our recent progress in deep-learning methods for de novo protein design and their applications to engineering new protein architectures, dynamic proteins, and constructing cellular signaling from the ground up.

22. Tina Perica: Functional crosstalk between EGF and insulin signalling

University of Zurich
t.perica@bioc.uzh.ch

Every cell needs to simultaneously, and in a concerted manner, respond to many different signals. Signalling pathways are, however often studied in isolation from each other. At the same time, these pathways appear over-regulated, with seemingly redundant regulatory mechanisms. I will present our preliminary work on systematically identifying functional crosstalk between pathways with proteomics as well as our ideas on how to probe the role of feedback loops in maintaining this crosstalk.

Short Talks

1. Abbie Lear: Understanding evolutionary improvements in designer enzymes to inform computational design
2. Alina Konstantinova: Designing cyclical oligomeric anchors for self-assembling protein fibers
3. Amijai Saragovi: Controlling semiconductor growth with structured de novo protein interfaces
4. Andrea Hunklinger: Protein Design with Explainable Artificial Intelligence
5. Britnie Carpentier: Incorporating Energy Calculations into AI Antibody Structure Prediction
6. Christoffer Norn: De novo design of GPCR modulators
7. Daniele Granata: Mega-scale in silico benchmark of de novo design tools for protein therapeutics
8. Deniz Akpinaroglu: Structure-conditioned masked language models for protein sequence design generalize beyond the native sequence space
9. Dominique Fastus: Synonymous codon selection bias and its impact on evolutionary co-translational protein folding
10. Fabio Parmeggiani: Predicting protein-carbohydrate interactions
11. Georg Künze: Computational engineering of a metagenome-derived polyester hydrolase for efficient PET depolymerization
12. Giulia Peteani: Enhancing metalloprotein sequence prediction with synthetic data
13. Katharina Bachschwöller: Design of a potentially novel FAD synthetase by fragment-based chimeragenesis
14. Lee Schnaider: De novo design and engineering of large functional protein complexes for sequencing and sensing
15. Martin Pačesa: BindCraft: one-shot design of functional protein binders
16. Matteo Cagiada: Predicting absolute protein folding stability using generative models
17. Max Beining: HyperMPNN – A general strategy to design thermostable proteins learned from hyperthermophilic organisms
18. Mykhailo Girych: Quality of disordered protein ensembles in coarse-grained molecular dynamics simulations
19. Niklas Gesmar Madsen: Composed Message-Passing for GPCR drug-target prediction: Integrating Knowledge Graphs and Deep Learning

20. Núria Mimbrero Pelegrí: REXzyme: A Translation Machine for the Design of Enzymes Tailored to Specific Reactions
21. Océane Follonier: Adapting and Assessing the Design of Protein-Protein Interactions for Therapeutic Applications
22. Ramiro Illanes Vicioso: Biophysical characterization and Catalysis of AI-designed carbonic anhydrases
23. Rasmus Krogh Norrild: mRNA-display based measurement of disordered protein phase separation through partitioning experiments
24. Sigrid Kaltenbrunner: Crafting Custom Catalysts: Designed Enzymes for the Morita-Baylis-Hillman Reaction
25. Sofia Andersson: De novo design of conformational changes using RFDiffusion and ProteinMPNN
26. Thea Klarsø Schulze: Learning sequence-abundance relationships across proteins from large-scale mutagenesis datasets
27. Valeriia Hatskovska: De novo design and characterization of bispecific cytokines with novel function

1. Abbie Lear: Understanding evolutionary improvements in designer enzymes to inform computational design

University of Bristol
abbie.lear@bristol.ac.uk

In this study we investigate the catalysis of the de novo designed Kemp eliminase 1A53-2, which has been improved 1,000-fold by directed evolution, to determine guidelines for future enzyme design. We calculate the reaction barrier for a range of enzyme conformations using QM/MM simulations and the adaptive string method, reproducing activity trends between variants. Analysis of transition state structures reveals features that determine activity and could be leveraged for future design.

2. Alina Konstantinova: Designing cyclical oligomeric anchors for self-assembling protein fibers

National Institute of Chemistry
konstantaalina@gmail.com

The goal of our work is to design proteins capable of moving along the tracks made of self-assembling protein fibers. Here use a combination of deep learning methods, such as RFdiffusion, ProteinMPNN and AlphaFold2, to design novel C10 oligomers that would rigidly connect to the ends of fibers and serve as anchors. The designs were ranked based on the combination of Rosetta and AlphaFold2's metrics. We present the experimental characterization (including cryoEM) of two rounds of design.

3. Amijai Saragovi: Controlling semiconductor growth with structured de novo protein interfaces

The Baker Lab, Institute for Protein Design, University of Washington
amijais@uw.edu

Protein design now enables the precise arrangement of atoms on the nanometer length scales (nanometers) of inorganic crystal nuclei, opening up the possibility of templating semiconductor growth. We designed proteins presenting regularly repeating interfaces containing functional groups that organize ions and water molecules, and characterized their ability to bind to and promote nucleation of ZnO. Utilizing the scattering properties of ZnO nanoparticles, we developed a flow cytometry-based sorting methodology and identified thirteen proteins with ZnO binding interfaces. Three designs promoted ZnO nucleation under conditions where traditional ZnO-binding peptides and control proteins were ineffective. Incorporation of these interfaces into higher order assemblies allowed the organization of defined protein-ZnO composite nanoparticles. These findings demonstrate the potential of using protein design to modulate semiconductor growth and generate protein-semiconductor composite materials.

4. Andrea Hunklinger: Protein Design with Explainable Artificial Intelligence

Centre for Genomic Regulation
andrea.hunklinger@crg.eu

The state-of-the-art protein language models (pLMs) excel at generating proficient proteins across diverse families, but they unfortunately operate as black-boxes. We apply explainable artificial intelligence (XAI) techniques like influence functions, feature attribution methods and the analysis of Transformer components to the field of enzyme design to enhance our understanding of the protein language and we use the insights to improve the models generation or for downstream prediction tasks.

5. Britnie Carpentier: Incorporating Energy Calculations into AI Antibody Structure Prediction

Johns Hopkins University
bcarpen6@jhu.edu

Due to the role of antibodies in the immune system and their specificity when binding to antigens, predicting antibody structure plays a crucial role in the design of effective therapeutics. Experimental techniques to test various structures are typically costly, creating a need for cost effective methods to predict antibody structure and their potential biophysical properties prior to experimental testing. In recent years, machine learning methods have been at the forefront of antibody structure prediction, however, not without their limitations. Energy calculations are fundamentally important in finding and understanding the most stable structures and conformations of proteins. Most machine learning structure prediction models do not include energy calculations in their training as they can be inaccurate and chemically implausible, and datasets are limited. The Rosetta Energy Approximation Network (REAN) is a new neural network that approximates the Rosetta Energy Function (REF15) energy terms to a fair degree of accuracy. I propose to integrate REAN into IgFold, a sequence-to-structure prediction model for antibodies, for training and inference, by including the energy approximation into the Invariant Point Attention (IPA) module, a component that IgFold borrowed from AlphaFold2. I will test whether incorporating energy calculations into the training for structure prediction will improve the accuracy of the structure prediction itself, particularly in the hypervariable CDR3 regions. In the long term, I will also improve the energy approximating network and investigate how it can provide insight into the structure's biophysical properties, such as, binding affinity and stability.

6. Christoffer Norn: De novo design of GPCR modulators

BioInnovation Institute
chn@biostudio.dk

G-protein coupled receptors (GPCRs) play a crucial role in various physiological processes and are major targets for drug discovery. Designing modulators that stabilize distinct functional states of G-protein coupled receptors (GPCRs) is challenging due to the small differences between these states and complex epitope shapes. We have developed novel metaproteome-based design methods and high-throughput screening techniques, resulting in the discovery of both agonists and antagonists.

7. Daniele Granata: Mega-scale in silico benchmark of de novo design tools for protein therapeutics

Novo Nordisk
dngt@novonordisk.com

To support both our internal design tasks and benchmark new tools, we developed a highly scalable and modular platform for protein design. We performed a large-scale benchmark of state-of-the-art tools for structure-based de novo binder design, generating more than 25M designs in total. The benchmark includes a systematic exploration of therapeutically relevant protein lengths (15-95 aa), the impact of hotspot information and prediction of developability properties for the designed sequences.

8. Deniz Akpinaroglu: Structure-conditioned masked language models for protein sequence design generalize beyond the native sequence space

UC San Francisco and UC Berkeley
deniz.akpinaroglu@ucsf.edu

Machine learning has enabled significant progress in protein sequence design. Frame2seq is a structure-conditioned masked language model that achieves state-of-the-art accuracy. We demonstrate Frame2seq's ability to generalize beyond natural sequences by designing for novelty, including a design with 0% sequence identity to native. Further, we show that our model is uniquely useful for control over the conformational landscape of multi-state proteins.

9. Dominique Fastus: Synonymous codon selection bias and its impact on evolutionary co-translational protein folding

Lund University
dominique_philip.fastus@biochemistry.lu.se

Computational analysis to understand how certain motifs in codons and protein structures relate are limited, as these two levels have been mostly separated in existing in silico studies. Here we obtained nucleotide sequences for a large-scale analysis on codon usage bias in correlation to secondary structures based on AlphaFold predictions. We also studied the conservation of rare codons of different protein families and investigated patterns of certain motifs and synonymous codons.

10. Fabio Parmeggiani: Predicting protein-carbohydrate interactions

Cardiff University
parmeggianif@cardiff.ac.uk

Protein-carbohydrate interactions are ubiquitous and fundamental for the function of proteoglycans. However, due to low affinity and similarity between carbohydrates, prediction of specificity, validation of docking models and design of protein binders are still poor. Moreover, successful machine learning tools, employed for protein structure prediction and design, require large amount of high-resolution data that are simply not available for protein-carbohydrate complexes.
In this work, we have developed a 3D graph neural network that process geometry and interactions of protein-sugar interfaces in three-dimension space, using limited and curated sets of high-resolution structures. The predictor takes advantage of sampling atomic features of carbohydrates, making possible to generalize its application also to sugars not included in the training set.
This tool allows us to rapidly classify structures and models of protein-carbohydrate complexes with high accuracy and efficiency, providing guidance to experimental testing of potential binders and design of novel carbohydrate binders.

11. Georg Künze: Computational engineering of a metagenome-derived polyester hydrolase for efficient PET depolymerization

University of Leipzig
georg.kuenze@uni-leipzig.de

Current research focuses on the development of efficient polyester hydrolases for the recycling of polyester plastics such as Poly(ethylene terephthalate) (PET). The recently discovered Polyester Hydrolase Leipzig 7 (PHL7), isolated from a compost metagenome, completely degrades amorphous PET from post-consumer plastic waste in less than a day. However, its application at an industrial level has been hampered by its low stability and short lifetime at low salt concentrations and high temperatures, which are required for efficient PET hydrolysis. Here, we have applied a computer-aided design workflow to engineer PHL7 for improved stability and activity. Multiple mutations were predicted by Rosetta energy calculations and combined, increasing the thermal melting temperature to more than 95°C. Subsequent iterative mutagenesis around the PET binding pocket yielded a PHL7 variant with a >120-fold increased activity and 12.2°C higher melting temperature in 0.1 M buffer. X-ray structure and MD simulation analyses revealed the mechanisms underlying the enzyme's elevated activity and stability. The new PHL7 variant is better than ICCG in terms of its PET hydrolyzing activity and comparable to the recently reported PETases LCC-A2 and TurboPETase at 65°C. Our results constitute a significant advancement for the engineering of industrially applicable PET hydrolases.

12. Giulia Peteani: Enhancing metalloprotein sequence prediction with synthetic data

Universität Bern
giulia.peteani@unibe.ch

Large language models (LLMs) show promise in protein modeling, yet they struggle with metalloproteins due to limited data. Thus, we propose a fixed backbone design to generate synthetic metalloprotein sequences, expanding the training set. Fine-tuning ProtGPT2 on these sequences, we developed models that better generate sequences likely to contain metal-binding sites while preserving quality. Our work shows the power of synthetic data augmentation for training LLMs on specific protein classes.

13. Katharina Bachschwöller: Design of a potentially novel FAD synthetase by fragment-based chimeragenesis

University of Bayreuth
katharina.bachschwöller@uni-bayreuth.de

The chimeric protein TyrAFld was constructed, comprising an ATP-binding fragment derived from the Rossmann fold domain of HiTyrA recombined with a fragment from the FMN-binding flavodoxin-like domain of DgFld, to catalzye the formation of FAD. After initial solubility problems and sequence optimization with PROSS and ProteinMPNN, the designed TyrAFld variants are expressed soluble and in high yields. Biochemical characterization indicates correct folding, high thermal stability and ATP binding.

14. Lee Schnaider: De novo design and engineering of large functional protein complexes for sequencing and sensing

UCSF
lee.schnaider@ucsf.edu

Protein nanopores are large channel-forming complexes which have considerable potential, from single-model detection of small molecules to commercially successful nucleic acid sequencing applications. Sequencing using protein nanopores relies on measuring the ionic current through a nanometer-scale protein pore embedded in a membrane. The baseline current is governed by the physical dimensions, chemical characteristics, and stability of the protein complex. The CsgG:CsgF protein-peptide pore is an 18-mer protein-peptide assembly (9xCsgG + 9xCsgF), which is part of the E. coli curli biogenesis system. Derivatives of this pore have been used for DNA sequencing, which places high demands on the structural stability and homogeneity of the complex. To increase the robustness of the pore we employed two methods (i) incorporating protein engineering with proximity labeling to design derivatives of CsgF-bearing sulfonyl fluorides, which react with CsgG in very high yield. While proximity labeling is primarily used analytically, we implemented it preparatively, to direct covalent bond formation between the subunits of this 280 kDa protein-peptide and covalently stabilize it. (ii) de novo design of highly stable protein-peptide nanopores with predetermined structural features, optimal channel conductance, and enhanced stability. Through this work, we were able to shed light on specific nanopore properties that govern membrane insertion propensity, signal uniformity and stability. Excitingly, derivatives of these designs are now in development at Oxford Nanopore Technologies. I would be very happy to present the methodologies that enabled these designs in an oral presentation.

15. Martin Pačesa: BindCraft: one-shot design of functional protein binders

EPFL
martin.pacesa@epfl.ch

Protein-protein interactions (PPIs) are at the core of all key biological processes. However, the complexity of the structural features that determine PPIs makes their design challenging. We present BindCraft, an open-source and automated pipeline for de novo protein binder design with experimental success rates of 10-100%. BindCraft leverages the trained deep learning weights of AlphaFold2 to generate nanomolar binders without the need for high-throughput screening or experimental optimization, even in the absence of known binding sites. We successfully designed binders against a diverse set of challenging targets, including cell-surface receptors, common allergens, de novo designed proteins, and multi-domain nucleases, such as CRISPR-Cas9. We showcase their functional and therapeutic potential by demonstrating that designed binders can reduce IgE binding to birch allergen in patient-derived samples. This work represents a significant advancement towards a "one design-one binder" approach in computational design, with immense potential in therapeutics, diagnostics, and biotechnology.

16. Matteo Cagiada: Predicting absolute protein folding stability using generative models

University of Copenhagen
matteo.cagiada@bio.ku.dk

While there has been substantial progress in our ability to predict changes in protein stability due to amino acid substitutions, progress has been slow in methods to predict the absolute stability of a protein. In outrwork, we showed how a generative model for protein sequence can be leveraged to predict absolute protein stability. We benchmarked our predictions across a broad set of proteins and find a mean error of 1.5~kcal/mol and a correlation coefficient of 0.7 for the absolute stability across a range of small--medium sized proteins up to ca. 150 amino acid residues. We analysed current limitations and future directions including how such model may be useful for predicting conformational free energies. Our approach is simple to use and freely available via an online implementation.

17. Max Beining: HyperMPNN – A general strategy to design thermostable proteins learned from hyperthermophilic organisms

University Leipzig
beining@izbi.uni-leipzig.de

Deep learning protein design approaches like ProteinMPNN have shown strong performance both in creating novel proteins or stabilizing existing ones, with stability being a key factor to enable the use of recombinant proteins in therapeutic or biotechnological applications. Nevertheless, the resulting stability of designs is unlikely to surpass significantly that of natural proteins in the training set, which tend to be only marginally stable. Here, we collected predicted protein structures from hyperthermophilic organisms, which differ significantly in their amino acid composition from mesophilic organisms. We show that ProteinMPNN fails at recovering this unique amino acid composition and subsequently retrained the network on hyperthermophilic proteins. The result, termed HyperMPNN, not only recovers this unique amino acid composition but can also be applied to non-hyperthermophilic proteins. Next, we experimentally verified our approach by stabilizing existing proteins. In conclusion, we created a new way to design highly thermostable proteins through self-supervised learning on data from hyperthermophilic organisms.

18. Mykhailo Girych: Quality of disordered protein ensembles in coarse-grained molecular dynamics simulations

VTT Technical Research Centre of Finland
mykhailo.girych@vtt.fi

Machine learning (ML) models trained on coarse-grained (CG) simulation data are gaining traction for predicting intrinsically disordered protein (IDP) ensembles. Here, we evaluate CG force fields (FFs) accuracy by comparing CG IDP ensembles to atomistic simulations and NMR spin relaxation data. Results show that while some CG FFs capture key IDP features, performance varies across different proteins. This highlights the need of CG FFs evaluation for reliable prediction of IDPs.

19. Niklas Gesmar Madsen: Composed Message-Passing for GPCR drug-target prediction: Integrating Knowledge Graphs and Deep Learning.

Technical University of Denmark (DTU) Biosustain
nikma@dtu.dk

Drug-target interaction (DTI) databases are vast but sparse, leaving many off-target and adverse drug effects unexplored. We developed a composed message-passing approach to robustly interpolate within this sparse data distribution for G protein-coupled receptors (GPCRs), starting at the molecular graph level and then extending to the chemical neighborhood level (also a graph). Finally, we correlated predictions with over 3,000 DTIs and validated 14 novel DTIs using a GPCR-yeast biosensing platform. The results demonstrate the inherent graph homophily in the data and likewise reconciles knowledge graphs and deep learning that operates on them.

20. Núria Mimbrero Pelegrí: REXzyme: A Translation Machine for the Design of Enzymes Tailored to Specific Reactions

Centre for Genomic Regulation (CRG)
nuria.mimbrero@crg.eu

REXzyme is an innovative enzyme design approach inspired by natural language processing and chemical and protein language models. It uses the T5 encoder-decoder model to translate desired chemical reactions into amino acid sequences. Trained on a dataset of 21 million non-redundant reactions-enzyme sequence pairs, REXzyme generates biocatalysts with natural-like properties: globular, ordered, and whose predicted functionalities match their intended catalytic reactions. Its applications span from biomedicine to environmental sciences.

21. Océane Follonier: Adapting and Assessing the Design of Protein-Protein Interactions for Therapeutic Applications

University of Basel
oceane.follonier@unibas.ch

Structural insights into virus interactions with the host immune system are key to developing novel antivirals and vaccines. Despite the availability of advanced deep learning methods to generate novel proteins, accurately scoring and validating the functionality of these designs remains a significant challenge. We aim to identify the critical factors in evaluating protein designs to ensure they meet specific target properties, such as preserved antigenic interfaces, solubility and stability.

22. Ramiro Illanes Vicioso: Biophysical characterization and Catalysis of AI-designed carbonic anhydrases

Centre for Genomic Regulation
ramiro.illanes@crg.eu

Our project focuses on the ability of ZymCTRL, a conditional language model trained on the enzyme sequence space, to generate stable and active proteins with catalytic performance close to those found in Nature. We were able to isolate seven carbonic anhydrases designed by ZymCTRL with an aminoacidic sequence identity below 50%, and define their oligomeric state, purity, stability and catalytic yield. In the end, the seven candidates proved to be monomeric (SEC-MALS), folded (CD, SAXS) and active (Wilbur-Anderson assay), accomplishing catalytic perfomances in line with their natural counterparts, such as the carbonic anhydrase from E. coli. The obtained results evidence ZymCTRL's capabilities at generating proficient enzymes without further training of rounds of optimization.

23. Rasmus Krogh Norrild: mRNA-display based measurement of disordered protein phase separation through partitioning experiments

Technical University of Denmark
rkrno@dtu.dk

Biomolecular condensates have emerged as a new class of cellular membrane-less organelles involved in compartmentalisation, regulation, and signalling. Relying on a multitude of interactions, phase separation of dynamic and weakly interacting protein and RNA molecules contribute to assembly. These interactions are challenging to quantify by traditional methods limiting thermodynamic data available for predictive models. Here we present an mRNA-display-based approach to directly measure the energetic contributions of peptide and RNA molecules to the formation of biomolecular condensates, at a large scale. Using the intrinsically disordered region (IDR) of Dead-box helicase 4 (DDX4N1), which is a central germ-granule component, we measure partitioning coefficients of almost 100,000 peptides and corresponding mRNA molecules between the protein-rich and depleted phases. By using partitioning as a proxy for phase separation, we show that peptide fragments of DDX4N1 itself provide high-resolution data on regions of the protein domain responsible for homotypic condensate formation. Additionally, peptide tiles of all other known IDRs form a catalogue of potential clients of the condensates. The combined data informs on general properties of partitioning peptides which might extend to also rationalise phase separation of unrelated IDRs. RNA partitioning is disfavoured upon secondary structure formation, and we show that purine content promotes partitioning. The unprecedented scale of the data generated by this method allows quantitative evaluation of how condensation behaviour and potential specificity are encoded in protein and RNA.

24. Sigrid Kaltenbrunner: Crafting Custom Catalysts: Designed Enzymes for the Morita-Baylis-Hillman Reaction

Graz University of Technology
sigrid.kaltenbrunner@student.tugraz.at

We report on the use of Riff-Diff, a modeling strategy for scaffolding catalytic arrays in de novo protein backbones to design an enzyme for the Morita-Baylis-Hillman reaction. Using Riff-Diff, 18 de novo sequences derived from 14 unique backbones were created. While the experimental characterization of activity is under investigation, we believe that the approach taken can straightforwardly be applied to different reactions, paving the way to fast solutions to biotechnological challenges.

25. Sofia Andersson: De novo design of conformational changes using RFDiffusion and ProteinMPNN

Lund University
sofia.andersson@biochemistry.lu.se

RFDiffusion is used to design two structurally unique backbones, where one is a de novo template structure and the other is a redesigned version. This is done using an iterative process based on sequence similarity and structural difference. ProteinMPNN is then used to design sequences for both structures, and the amino acid probabilities are combined to create a sequence which encodes both structures. Two such sequences have been expressed experimentally and CD spectra and NMR show a reversible conformational change between two unique states.

26. Thea Klarsø Schulze: Learning sequence-abundance relationships across proteins from large-scale mutagenesis datasets

University of Copenhagen
thea.schulze@bio.ku.dk

Accumulation of data from multiplexed assays of variant effects (MAVEs) has enabled large-scale analyses of variant effects in proteins and created the opportunity to train supervised models directly against experimental data to learn sequence-function relationships, including how missense variants affect protein stability and abundance. We have used data obtained by variant abundance by massively parallel sequencing (VAMP-seq), a MAVE technique that quantifies the steady-state cellular abundance of protein variants, to create a model for predicting the impact of missense variation on cellular abundance across proteins. We used data reporting on the effects of ca. 32,000 missense variants on the abundance in six proteins as training data for a deep learning model. The model consists of (i) an inverse folding model taking protein structure as input, (ii) a supervised model for effects on protein stability and abundance, (iii) a function mapping stability and abundance effects to folding probability, and (iv) a downstream model that describes the experimental process. Our model predicts variant abundance with state-of-the-art accuracy.

27. Valeriia Hatskovska: De novo design and characterization of bispecific cytokines with novel function

MPI for Biology, Tuebingen
valeriia.hatskovska@tuebingen.mpg.de

The undifferentiated hematopoietic and leukemic stem cells are difficult to target due to their low abundance and lack of unique cell surface markers. To target these cells using bivalent binders of more than one cytokine receptor can be an effective strategy. To achieve this, we apply de novo design of bivalent, single-domain binders, using Damietta protein design. Specifically, here we report proteins capable of associating two key hematopoietic receptors (IL-3Ra and G-CSFR), simultaneously. Upon experimental characterization of five design candidates, they were well-expressed and highly thermostable, and two candidates bound both targets at nanomolar affinities. Our results also showed one of our designs to be capable of associating both receptors simultaneously at nanomolar concentrations.

Posters

1. Adrian Bunzel: Creating Photoenzymes by Design and Evolution
2. Ajasja Ljubetic: Designing and tracking a completely de novo random protein walker
3. Alex Cahill: De Novo Design of Metallocorrin Enzymes
4. Alexia Taylor: Structural homology search as a tool for nature-inspired de novo protein design
5. Ali Asghar Hakami Zanjani: Engineering Annexin A3 to Mimic Annexin A4: Insights into Plasma Membrane Repair Mechanisms
6. Alvaro Martin: De novo design of transmembrane fluorescent activating proteins (tmFAPS)
7. Anna Schrüfer, De novo design of thermostable LPMO (Lytic Polysaccharide Monooxigenase)
8. Anna Nobis: Establishment and validation of a computational vaccine development pipeline for emerging viral diseases – a case study on Nipah Virus
9. Armen Sargsyan: Modeling, synthesis and in vitro testing of unusual amino acid and peptides as potential antibacterial agents
10. Arriën Symon Rauh: Probing the Interactions in Biomolecular Condensates using Simulations of Double Mutant Cycles
11. Aswathy Muttathukattil: Design of Viral Capsid Assembly
12. Benedikt Singer: Expanding ModelArchive for De Novo Protein Design
13. Benjamin James Hardy: Designing transmembrane cytochromes for synthetic bioenergetics
14. Bhoomika Basu Mallik: Towards predicting oligomorphism in de novo designed protein assemblies
15. Bram Mylemans: De novo design of minimal proteins for in cell protein degradation
16. Bruce Lichtenstein: Binding and manipulating the properties of bound cofactors in computationally designed proteins
17. Carlos Cruz: Design of antimicrobial peptides self-assembled into virus-like capsids
18. Carlos Josué Alvarez Quispe: Unlocking dynamic-function relationships in UDP-glucose 4-epimerases
19. Cesar Antonio Ramirez-Sarmiento: Improving the activity of PETases and MHETases at varying temperatures using MD simulations and ThermoMPNN
20. Charlotte Crauwels: Designing chimeric G-Protein Coupled Receptors: Past, Present and Future
21. Christian Tüting: Conserved hydrophilic checkpoints tune FocA-mediated formate:H+ symport
22. Christopher Frank: Machine learning methods for the design of large proteins

23. Dan Pike: Correlated dynamics and activation heat capacity in the adaptation of enzyme activity to low temperatures: triosephosphate isomerase
24. Dana Cortade: The Open Datasets Initiative & Protein Engineering Tournament: Generating living datasets & benchmarking opportunities for machine learning models
25. Daria Gusew: Investigating the dynamics of protein - ligand interactions
26. Dimitrios Kolokouris: Investigating the role of cholesterol in regulating the SLC7 family of amino acid transporters
27. Dina Listov: A fully computational design of high-efficiency Kemp eliminase
28. Dominic Rieger: De novo design of protein-binding proteins targeting viral attachment proteins of the paramyxovirus family
29. Elizaveta Maltseva: Design of Degenerate Interfaces for Multiple Geometry-Forming Scaffolds Using Diffusion Models
30. Ethan Bungay: In silico directed evolution of de novo flavocytochromes
31. Eva Rajh: Kinetic control of Coiled coil complex dissociation
32. Eva Smorodina: Structural modeling of antibody variant epitope specificity with complimentary experimental and computational techniques
33. Fabian Schuhmann: The Automated Ligand Searcher (ALISE)
34. Fan Cao: A coarse-grained model for disordered and multi-domain proteins
35. Felipe Akihiro Melo Otsuka: Single synonymous scanning mutagenesis of E. coli's protein expression and cell stress profile to measure translation efficiency
36. Fernando Meireles: Guiding structure-based protein design with UniProt annotations
37. Francesco Altiero: In-silico Workflow to Design De-novo Antigens by Epitope Scaffolding
38. Giulio Tesei: Conformational ensembles of the human intrinsically disordered proteome
39. Gustav Oberdorfer: Computational design of highly active de novo enzymes
40. Gustavo Araiza: Epitope-focusing and viral escape prediction for the design and selection of robust influenza antibodies
41. Han Tang: A Diffusion Model Approach to Propose Non-Canonical Amino Acid Substitutions in Protein Engineering
42. Haye Nijhuis: Producing and stabilizing recombinant surface glycoproteins from highly pathogenic zoonotic viruses
43. Ho Yeung Chim: Computational design of small molecule-dependent cyclic oligomers.
44. Holly Ford: Computational design and biochemical characterisation of a molecular wire
45. Horst Lechner: Improving protein expression using ProteinMPNN
46. Hubert Klein: In-Silico Design of Foldamer – Protein Interactions
47. Ida Kjærsgaard Grene: De novo protein design for neuroscience
48. Ignacio Retamal: Autonomous motor design, modulating track binding affinity through allostery

49. Inken Fender: Predicting Molecular Hotspots for Efficient Drug Discovery using Deep Learning
50. Lori Morita: De novo design of a programmable sequence-specific ssRNA binding protein
51. Jacopo Sgrignani: AI based design of human lipocalin ligands
52. Jakob Riccabona: Development of a computational vaccine design pipeline using machine learning – focus on stabilizing proline substitutions
53. Jakob Nilsson: Substrate recognition principles for the PP2A-B55 protein phosphatase
54. Janina Sörmann: Leveraging Deep Learning tools to design hASIC1a miniprotein modulators for stroke treatment
55. Jannick Prentoe: The hepatitis C virus envelope protein complex is a dimer of heterodimers
56. Joel James Chubb: RASSCoL: Simplifying Computational Design of Small-Molecule Binding Pockets
57. Johannes A. Klier: Exploring de novo-likeness of natural proteins to improve protein design
58. Johannes Peterlechner: Enhancing Protein Loop Model Quality with a Rotationally Equivariant Variational Autoencoder (REVA)
59. Johannes Schweininger: In vitro and in silico inhibitor design against the HCMV pUL50-pUL53 interaction
60. Johnny Alexander Jimenez Siegert: Modeling ASIP-Melanocortin Receptor Interactions for Anti-Obesity Drug Design
61. Jonas Gunkel: Computational design and experimental testing of Polyester Hydrolase Leipzig 7 (PHL7) mutants on different bioplastics
62. Jonathan Funk: ProteusAI: An Open-Source and User-Friendly Platform for Machine Learning-Guided Protein Design and Engineering
63. Jonathan Hungerland: The rhodopsin GPCR complex in European Robin: computational construction and validation
64. Jonathon Liston: Designing cobalt specific protein “Velcro”
65. Jonathan Ziegler: Conditioning Generative Models for Protein Optimization
66. Joshua Behringer: Protein functionalization towards enzyme selectivity
67. Joy Lyu: AI-guided identification of novel extracellular protein interactions between wheat and Zymoseptoria tritici
68. Julia K. Varga: Substrate recognition principles for the PP2A-B55 protein phosphatase
69. Julian Beck: Scaffolding a Kemp eliminase activity into an idealized de novo TIM barrel
70. Juno Underhill: Spectral tuning in a de novo designed riboflavin protein

.

71. Justin Booth: Evolving Allostery: Harnessing the Power of Continuous Evolution for Dynamic Protein Engineering
72. Karel van der Weg: Towards the construction of a large foundation model for protein structures with message-passing Graph Neural Networks
73. Kristoffer E. Johansson: Global Analysis of Multi-Mutants to Improve Protein Function and Stability
74. Leonardo De Maria: PepInvent: generative models for therapeutic peptide optimization
75. Leoni Abendstein: Designing Synthetic Ortholog Heterodimers: Challenges and Opportunities in De Novo Protein Engineering
76. Leonie Windeln: Turning toxins into targets: design of α-conotoxin capturing proteins.
77. Liza Ulčakar: Characterization of designed cyclic oligomers that walk on designed protein fibers
78. Lorenzo Scutteri: De Novo Design Light-Responsive Protein Switches
79. Lucien Krapp: PeSTo and CARBonAra: the secret sauces for protein design
80. Luis I. Gutierrez-Rus: PiLGriM: A computational pipeline for metalloprotein design
81. Luise Jacobsen: Clues about Uncoupling Protein 1 Mechanisms from Molecular Simulations
82. Lukas Milles: De novo design of intra- and intermolecular autocatalytic isopeptide bonds
83. Mads Jeppesen: Title: Shape-Directed Protein Assembly Design
84. Mads Mørup Nygaard: De novo designed functional peptide binders for novel targets
85. Marco Orlando: In silico tools to mine glycoside hydrolase enzymes from nature
86. Maria Artigues Llexia: Embedding Annotation Transfer (EAT): a contrastive learning implementation to functionally annotate prokaryotic transcription factors for bioengineering
87. Marin Matic: Dissecting structure determinants of GPCR-G protein selectivity via structural bioinformatics
88. Marius Amann: Enabling Heterologous Expression and Enhancing Stability of Plasmodium falciparum Proteins through ProteinMPNN-Driven Mutagenesis
89. Mark Kriegel: De novo prediction of explicit water molecule positions by a novel algorithm within the protein design software MUMBO
90. Marvelous Chikerema: Towards a cDNA display platform for the directed evolution of membrane proteins
91. Matej Milijaš-Jotić: Computational Design of a Bivalent Cytokine Receptors Antagonist
92. Matteo Tiberti: MAVISp: Multi-layered Assessment of Varlants by Structure for proteins
93. Matthew Ellis: De novo design of synthetic dye binders
94. Maximilian Salomon: GPCRFlow: GPCR Ensemble Prediction of Metastable States
95. Michael Westberg: Real-time control of endogenous proteins by photoswitchable binders

96. Miguel Atienza Juanatey: Towards meta-stable protein assemblies
97. Mikel Martinez Goikoetxea: Old folds can learn new tricks: Alphafold-driven insights on coiled-coil structure
98. Moritz Ertelt: Combining machine learning with structure-based protein design to predict and engineer post-translational modifications of proteins
99. Muhammad Yasir Ateeque: Modeling the Structural Dynamics of the DnaB-DNA Complex using AI Structure Prediction Tools and Solid-State NMR Data
100. Nicolas Deutschmann: Kickstarting ML-guided antibody optimization without affinity data
101. Nina Schuch: The Impact of Epistasis and Covariation on Enzyme Stability and Activity through Ancestral Reconstruction
102. Palina Pliushcheuskaya: Evaluation of binding site prediction methods for ligands targeting the protein-membrane interface region
103. Parth Bibekar: RISoTTo: Context-aware geometric deep learning for RNA sequence design
104. Pasquale Miglionico: Predicting Stable Protein Complexes with Co-evolution and Co-expression Data
105. Paul Eisenhuth: Flexible E(3) equivariant interaction kernels for biomolecular affinity prediction
106. Paul Kluge: Parametrical design of surface binding helical bundles for the elucidation of the CISS-effect
107. Pauline Hermans: Exploring evolution to enhance mutational stability prediction
108. Piotr Smieja: deepBBQ: a deep learning approach to the protein backbone reconstruction
109. Qihao Zhang: A Heme Location Prediction Tool with 3dCNN/Docking Procedure
110. Rajashi Sinharoy: Studying the Association of Blood Plasma Proteins on Lipid Nanoparticles with Brownian Dynamics Simulations
111. Renzo Condori: Machine Learning-Guided In Silico Design of EGFR Binders: A Promising Alternative for Treating Triple-Negative Breast Cancer
112. Robert Jefferson: Computational design of dynamic receptor—peptide signaling complexes applied to chemotaxis
113. Rodrigo Ochoa: Open-source tools for peptide analysis and design: PepFuNN and mPARCE
114. Sajith Kolathuparambil: Design of noval metalloproteases
115. Samuel Curtis: Adapting DNA synthesis screening to a de novo world

116. Sarah Badawy Ahmed Ossman: Coarse-Grained Umbrella Sampling Simulations for Binding Affinity Prediction of SARS-CoV-2 spike-Nanobody Complexes
117. Sebastian Lindner: Expanding the natural chorismate mutase verse through AI-guided design and wet lab evolution
118. Sevilay Gülesen: Prediction of Different Conformational States of Class I Fusion Protein
119. Shengxiao Qi: Cloning and Expression of ASMT for the Identification of Pharmacological Probes
120. Shuhao Zhang: Deep-Learning-Designed Nanobodies Targeting AT1R Show Enhanced Binding Affinity and Antagonistic Effects
121. Simon Duerr, EPFL: Predicting metal-protein interactions using cofolding methods: Status quo
122. Sören von Bülow: Prediction of phase separation propensities of disordered proteins from sequence
123. Teemu Kalle Eemeli Rönkkö: Structure-based predictions as a novel means to pair cone snail-derived peptide toxins with G-protein coupled receptors
124. Thomas Lavstsen: Harnessing Novel Broadly Neutralizing mAbs Against P. falciparum Virulence Proteins for Vaccine Design
125. Till El Harrar: Decoding the Dual Role of Aqueous Ionic Liquids in Protein Stability: Insights from Energy Profiling and Perturbation Pathways
126. Timothy Patrick Jenkins: Crack the Code, Craft the Cure: How Machine Learning is Redefining Target Discovery and Therapeutic Innovation
127. Tobias Wörtwein: A web-based toolkit for protein design
128. Victor Klein-Sousa: Towards a complete phage tail fiber structure atlas
129. Yasser Almeida: PPI-Affinity: A Tool for the Prediction and Optimization of Protein–Peptide and Protein–Protein Binding Affinity
130. Yuval Fishman: Exploring the Functional Diversity of Natural Enzyme Families Using Protein Design
131. Yuyang Zhang: TopoDiff: Improving diffusion-based protein backbone generation with global-geometry-aware latent encoding

1. Adrian Bunzel: Creating Photoenzymes by Design and Evolution

ETH Zurich
adrian.bunzel@bsse.ethz.ch

The energy crisis challenges us to produce energy more sustainably. Our group combines computational design and directed evolution to create photoenzymes for biohybrid solar cells. Here, we show how we designed photosensitizer binding sites into existing protein scaffolds, significantly enhancing photostability and improving electricity generation. We are currently optimizing solar efficiency through directed evolution, which may lead to solar cells competitive with established photovoltaics.

2. Ajasja Ljubetic: Designing and tracking a completely de novo random protein walker

National Institute of Chemistry
ajasja.ljubetic@ki.si

Deep learning methods have greatly increased experimental success rates for de novo design of single chain proteins; however large dynamic protein mechanisms have not yet been designed. I will present the design and characterization of a random protein walker that can diffuse along micro-meter long fibers. I will outline the different walker scaffolds we have designed and present several experimental 3D structures and trajectories obtained from single molecule TIRF microscopy experiments.

3. Alex Cahill: De novo design of metallocorrin enzymes

University of Portsmouth
alex.cahill@port.ac.uk

Vitamin B12 is an essential natural cobalamin cofactor able to catalyse chemical transformations that have proven challenging in chemistry and industry, such as dehalogenation and selective isomerisation, and is vital for DNA methylation within the body. The role vitamin B12 and its derivatives play in these processes is well established, however how proteins control its chemistry is still not well understood. De novo designed cobalamin binding proteins will allow close study of the fundamental engineering principles underlying the complex enzymes using cobalamin that are found in nature. This research aims to use recently developed deep learning tools, such as RFdiffusion all-atom and LigandMPNN, to design proteins capable of binding and utilising cobalamin, simplifying the study of the fundamentals of B12 chemistry without the complexities found within biological systems. The protein designs were validated using the molprobity tool, and designs will be expressed in the laboratory whereafter their binding affinities and chemical activities will be characterised in order to tease out the engineering rules for controlling B12 chemistry.

4. Alexia Taylor: Structural homology search as a tool for nature-inspired de novo protein design

University of Bristol
alexia.taylor@bristol.ac.uk

A FoldSeek structural-homology search of the AlphaFold database against the de novo transmembrane cytochrome CytbX revealed thousands of uncharacterised integral membrane proteins, all of which are predicted to contain heme binding sites. Characterisation of these proteins will deepen our understanding of electron transfer reactions in membranes, and enable us to elucidate design principles for tuning redox properties in de novo membrane cytochromes.

5. Ali Asghar Hakami Zanjani: Engineering Annexin A3 to Mimic Annexin A4: Insights into Plasma Membrane Repair Mechanisms

University of Southern Denmark
zanjani@sdu.dk

Annexin proteins are crucial in regulating various cellular processes, including plasma membrane repair (PMR). Annexin A4 is known to form trimeric structures on the membrane surface, which induce high negative curvature and accelerate PMR. However, Annexin A3 does not form trimers. We investigated the factors driving annexin-mediated PMR by using MD simulations to engineer a mutant form of Annexin A3 that can trimerize and induce high curvature on the membrane surface, similar to Annexin A4

6. Alvaro Martin: De novo design of transmembrane fluorescent activating proteins (tmFAPS)

VIB-VUB Center for Structural Biology
alvaro.martin@vub.be

Fluorescence activating proteins (FAPs) are non-fluorescent proteins which are engineered to bind and activate a small fluorogenic ligand. Here, I present a computational strategy to design a de novo transmembrane protein capable of binding a fluorogen of interest. I provide experimental evidence for in vitro fluorescence as well as binding specificity between chemically similar ligands. Finally, I devise a strategy for in vivo expression of the tmFAPs into mitochondrial membranes to optimise fluorescence through rational mutations and directed evolution.

7. Anna Schrüfer, De novo design of thermostable LPMO (Lytic Polysaccharide Monooxigenase)

Graz University of Technology
a.schruefer@tugraz.at

We designed new LPMO variants and expressed them in E. coli to improve their activity and stability. All variants were active, some potentially more than the wild type, and showed better thermal stability. However, purification was difficult because the enzymes degrade the cellulose materials used in the lab. Despite this, we solved the crystal structure of one variant using X-ray data from ESRF, giving us insights into its properties.

8. Anna Nobis: Establishment and validation of a computational vaccine development pipeline for emerging viral diseases – a case study on Nipah Virus

Leipzig University, Medical Faculty, Institute for Drug Discovery
anna.nobis@medizin.uni-leipzig.de

During the Covid-19 pandemic, the importance of epidemic preparedness was clearly shown. We created a computational vaccine design pipeline, applied it to Nipah virus, and stabilized its fusion protein using insertion of disulfide bridges and single-point mutations. Leveraging AlphaFold, ProteinMPNN, and Rosetta, we identified vaccine candidates quickly and cost-effectively. This pipeline enables rapid development of novel vaccines based on predicted protein structures within a few weeks.

9. Armen Sargsyan: Modeling, synthesis and in vitrio testing of unusual amno acid and peptides and potential antibacterial agents

SPC "Armbiotechnology" NAS RA
armenssargsyan@gmail.com

Many proteases are of great interest for pharmacology due to their key role in various pathologies. Bacterial collagenase is a quite attractive target for drug development as the inhibitors of bacterial collagenolytic protease may stop disease propagation caused by infections. A search for compounds with the ability to inhibit bacterial collagenases was conducted. For this purpose, multiple structures were generated and screened in silico. Leeds were synthesized and tested in vitro.

10. Arriën Symon Rauh: Probing the Interactions in Biomolecular Condensates using Simulations of Double Mutant Cycles

University of Copenhagen
arrien.rauh@bio.ku.dk

The relationship between the sequence and and phase separation (PS) behaviour of intrinsically disordered proteins (IDPs) is often interrogated through mutagenesis experiments [1-4]. In these experiments either the interactions of specific amino acid residues, or sequence patterns in general are perturbed. To get a more detailed understanding of how these mutations affect the biomolecular condensates we explore the application of double mutant cycles (DMCs) in molecular dynamics (MD) simulations to quantify the impact of sequence perturbations [5-7]. To this end, we use the effective residue-level coarse-grained MD model CALVADOS [8,9]. Here we present the case of tyrosine-arginine interaction(s) in the hnRNPA1-LCD (A1). We can quantify a contribution to stabilising condensates of A1. However, interpreting the alterations in interaction pattern in the condensate is not straightforward, due to condensate concentration differences.

[1] Wang et al. Cell (2018) [2] Borcherds et al. Curr. Op. in Str. Bio. (2021)[3] Martin et al. Science (2020) [4] Bremer et al. Nat.Chem. (2022)[5] Carter et al. Cell (1984)[6] Fersht et al. J. Mol. Bio. (1992) [7] Horovitz Folding and Design (1996)[8] Tesei et al. PNAS (2021)[9] Tesei & Lindorff-Larsen Open Res. Eur. (2022)

11. Aswathy Muttathukattil: Design of Viral Capsid Assembly

Lund Univeristy
aswathy.muttathukattil@biochemistry.lu.se

Natural capsids are stabilized by protein–protein interfaces that have evolved to be efficient self-assembling units. The hierarchy of strengths of interactions of interfaces are carefully selected by nature to allow error-corrected assembly without any off-pathway aggregated states. Our research aims to unravel the mechanism governing viral capsid assembly using molecular simulations. Based on the principles obtained from the efficient natural system using minimal coarse-grained models, we intend to modify the monomers to redesign a capsid assembly process for biomedical applications.

12. Benedikt Singer: Expanding ModelArchive for De Novo Protein Design

EPFL
benedikt.singer@epfl.ch

Since more than a decade, ModelArchive as an analogue to the PDB has enabled open deposition of computed structures of biomolecules, filling a gap in structural biology's open data efforts. In close collaboration with the curators of ModelArchive at the Swiss Institute of Bioinformatics (SIB) with Torsten Schwede's group, our aims are to facilitate the deposition of de novo designed proteins to ModelArchive by leveraging and extending the already published ModelCIF format. The original development of this format involved the wwPDB/mmCIF-developers. Its ability to store metadata enables an automated validation of already deposited structures. As both ModelArchive and ModelCIF already display interoperability and representability when it comes to predicted structures of proteins (e.g. originating from AlphaFold) there exists a lack thereof for designed proteins. The aim is to improve the reuse of structure models by providing accuracy estimates and metadata while still being fully interoperable with the structures in the PDB and further enabling reuse and reproducibility.

13. Benjamin James Hardy: Designing transmembrane cytochromes for synthetic bioenergetics

University of Bristol
ben.hardy@bristol.ac.uk

Protein design enables the elucidation of underlying principles of electron transfer (ET) at the membrane and the construction of bioenergetic pathways from scratch. Our core module is cytochrome bX (CytbX), a de novo di-heme membrane cytochrome capable of cross-membrane ET. I will share recent design efforts to study heme-heme interactions in the membrane and to enable light-powered ET, and describe an in-silico directed evolution pipeline for the design of protein:protein interactions.

14. Bhoomika Basu Mallik: Towards predicting oligomorphism in de novo designed protein assemblies

LMU Munich
bhoomika.basu@lmu.de

Over the past decade, advancements in protein engineering have enabled the creation of novel polyhedral protein assemblies with precisely tailored structural properties for specific applications. However, current design strategies are mainly limited to highly symmetric icosahedral, octahedral, and tetrahedral assembly geometries. Recently we have shown that structural flexibility drives assembly into multiple well-defined geometries, rather than aggregation, exploring a wider repertoire of architectures. Informed by experimental screening, we are developing a computational pipeline that combines physics-based and AI-driven methods to rapidly predict structural flexibility and consequently control assembly outcome. Here, I will present our preliminary results for two de novo designed assembly families.

15. Bram Mylemans: De novo design of minimal proteins for in cell protein degradation

University of Bristol
bram.mylemans@bristol.ac.uk

We designed a robust protein scaffold based on an antiparallel coiled-coil dimer. Using a hybrid of classical motif grafting and protein MPNN, binding sites for the anti-apoptotic BClxl and MCL1 proteins were introduced on the solvent exposed side of the scaffold. Specific binding was achieved and was proven by ITC. These scaffolds were further functionalized to degrade BCLxL through the ubiquitination pathway, this was validated by cellular assays. To ubiquitinate the target, the loop was redesigned using the same motif grafting method to bind to the KLHL20 E3 ligase.

16. Bruce Lichtenstein: Binding and manipulating the properties of bound cofactors in computationally designed proteins

University of Portsmouth
bruce.lichtenstein@port.ac.uk

Cofactors significantly expand the functional repertoire of proteins beyond the chemical and physical properties of natural amino acids. Despite their critical role in nearly all biological processes, the influence of the protein environment on cofactor behaviour remains poorly understood. To address this gap, we employ de novo design to systematically investigate how protein sequence and structure modulate the properties of bound cofactors, allowing us to isolate these effects from evolutionary constraints. In this study, we are concurrently developing synthetic cofactors and de novo-designed scaffolds to enhance and regulate their activities. We present our latest findings on novel fluorogenic proteins and oxidoreductases, emphasizing recent advancements in designing cofactor-binding proteins for fully synthetic dyes, a new class of phenazines, and naturally inspired metallocorrins

17. Carlos Cruz: Design of antimicrobial peptides self-assembled into virus-like capsids

University College London
carlos.cruz@ucl.ac.uk

Sophisticate DL-based protein design methods were applied to tailor novel trigonal building blocks, structures and sequences, using the topology of 12-mer beta-annulus peptides from the Sesbania virus as a seed of our design strategy. With high prediction accuracy, these peptides can form the framework needed to nano-capsule assembling. Our hypotheses have been explored by extensive full-atom and coarse-grained molecular dynamics simulations.

18. Carlos Josué Alvarez Quispe: Unlocking dynamic-function relationships in UDP-glucose 4-epimerases

Ghent University
carlosjosue.alvarezquispe@ugent.be

UDP-glucose 4-epimerases (Gal4E) form a very interesting group of enzymes, as they can invert the configuration of a specific hydroxyl group without prior activation or protection steps. As a result, an unusual sugar structure can be generated from a more common counterpart, leading to new properties and potential applications in the food and pharma industries. Despite the intricate mechanism in Gal4Es involving oxidation, sugar ring rotation, and reduction, the role of protein dynamics in this process remains poorly understood. In this study, we employ molecular dynamics simulations with biochemical characterization to uncover the dynamic-function relationships within these enzymes

19. Cesar Antonio Ramirez-Sarmiento: Improving the activity of PETases and MHETases at varying temperatures using MD simulations and ThermoMPNN

Pontificia Universidad Catolica de Chile
cesar.ramirez@uc.cl

PET hydrolases (PETases) and MHET hydrolases (MHETases) are exciting potential biocatalysts for the recycling of PET. Most known efficient PETases operate at high temperatures, whereas only a few mesophilic MHETases are known. The design of PETases that efficiently degrade PET at lower temperatures is desirable to enable whole-cell biocatalysis, whereas the design of thermostable MHETases could aid in enhancing PET degradation at high temperatures using PETase-MHETase mixtures. In this short talk, we will describe the discovery of a novel PETase found in an Antarctic organism, Mors1, that operates at 25 ºC. Using MD simulations to compare the flexibility of the active site between thermophilic and mesophilic PETases, we generated a Mors1 chimera by loop swapping with an extended loop from the thermophilic PETase LCC that has 5-fold more activity and an optimal temperature for activity at 45ºC. Reasoning that thermostable MHETases are required for generating synergistic PETase-MHETase systems to degrade PET at high temperature, we used ThermoMPNN to increase the thermal stability of a mesophilic MHETase. While the wild-type enzyme exhibits higher activity ~40 ºC, all tested single-point mutants generated by ThermoMPNN show higher activity at 50ºC, with one of such mutants exhibiting ~10-fold higher activity than the wild-type MHETase.

20. Charlotte Crauwels: Designing chimeric G-Protein Coupled Receptors: Past, Present and Future

Vrije Universiteit Brussel
charlotte.crauwels@vub.be

We developed GPCRchimeraDB, the first database to centralize and describe chimeric G-Protein Coupled Receptors (GPCRs). By merging two GPCRs, these chimeras help gaining insights into GPCRs, proteins involved in many diseases. We also generated a novel alignment method, SIMSApiper, that enhances the reliability of sequence alignment of undruggable GPCRs. These results are now being employed to define the sequence, structural and biophysical space of GPCRs and suggest new chimeric designs.

21. Christian Tüting: Conserved hydrophilic checkpoints tune FocA-mediated formate:H+ symport

Martin-Luther University Halle-Wittenber
christian.tueting@bct.uni-halle.de

FocA belongs to the widespread, evolutionarily ancient formate-nitrite transporter (FNT) family of pentameric anion channels and translocates formic acid bidirectionally. Here, we identify compartmentalized polarity distribution across the complete FocA pore structure – resolved at 2.56 Å – mirrored against a two-fold axis with H209 at its center. The FocA-H209N efflux-only variant reveals a density consistent with formic acid located directly at N209, breaking local polarity distribution. Pyruvate formate-lyase, generating formate, orients at the cytoplasmic face where formate delivery is regulated by conformational changes in the FocA vestibule. Comparisons with other FNTs suggest a tuning mechanism of formate-specific transport via checkpoints enriched in hydrophilic residues.

22. Christopher Frank: Machine learning methods for the design of large proteins

Technical University of Munich
chris.j.frank@tum.de

Machine-learning (ML) based design approaches have advanced the field of de novo protein design, with diffusion-based generative methods increasingly dominating protein design pipelines. Here, we report two approaches which build on the structure prediction network Alphafold2 (AF2) and turn it into a state-of-the-art protein design algorithm without any need for retraining or finetuning. We accomplish this by firstly introducing Relaxed Sequence Optimization, a method that utilizes a relaxed or continuous sequence space, enabling the efficient design of high-quality protein backbones over multiple scales and with broad scope of application. We experimentally produced and characterized over 100 proteins using this approach. Three high-resolution crystal structures and two cryo EM density maps of designed single-chain proteins comprising up to 1000 amino-acids validate the accuracy of the method. Our pipeline can also be utilized to design synthetic protein-protein interactions, as validated experimentally by a set of protein heterodimers. Secondly, we propose a new way of how AF2 can be combined with ProteinMPNN and be directly used as a highly efficient denoising diffusion model to recover suboptimal backbone designs from Chroma. We verify this by rigorous computational benchmarking as well as by experimental characterization, generating very large proteins resembling different distinct shapes. All together we show that structure prediction networks can be readily and easily modified into state-of-the-art protein design algorithms with a broad application regime, requiring no retraining and only slight code modifications.

23. Dan Pike: Correlated dynamics and activation heat capacity in the adaptation of enzyme activity to low temperatures: triosephosphate isomerase

University of Bristol
ze23981@bristol.ac.uk

In enzyme catalysis, activation heat capacity (ΔC‡) is increasingly recognized as a key factor for understanding the temperature dependence of enzyme-catalysed reaction rates, particularly in the context of thermoadaptation and enzyme evolution. Understanding the underlying principles and dynamical origins would enable the prediction and manipulation of enzyme temperature optima. Here, we combine simulation and experiment to study three variants of triosephosphate isomerase (TIM) from psychrophilic, mesophilic, and thermophilic organisms. Kinetic data shows clear curvature in Eyring plots for psychrophilic and mesophilic enzymes. This non-Arrhenius behaviour can be accounted for by macromolecular rate theory (MMRT). ΔC‡ values calculated from extensive molecular dynamics (MD) simulations of the Michaelis complex and a model of the transition state show good agreement with experimental kinetic data. Heat capacity differences are borne out in root mean square fluctuation (RMSF) analysis, showing the psychrophile to be the most flexible and the thermophile the least. Both clustering analysis and principal component analysis (PCA) show conformational distributions indicative of the measured and calculated heat capacities for each variant. Correlation and network analysis, in combination with projection of the principal components, also reveal key residues and structural movements that may be important in modulating the dynamical behaviours involved in transition state stabilisation and therefore differences in ΔC‡ between these variants. This helps us to progress our understanding of the dynamical origins that arise as a result of thermoadaption and thus enable the tuning of temperature optima for enzymes with biotechnological applications.

24. Dana Cortade: The Open Datasets Initiative & Protein Engineering Tournament: Generating living datasets & benchmarking opportunities for machine learning models

Align to Innovate
dana@alignbio.org

Public databases, such as the PDB, and community-driven benchmarking, such as CASP, have advanced the state of the art of sequence-to-structure prediction. Align to Innovate, a scientific non-profit, is building upon these successes to address sequence-to-function prediction. We will discuss initial results from our suite of programs aiming to create living datasets for machine learning and benchmarking opportunities for sequence-to-function ML models, supported by automated, open-source methods.

25. Daria Gusew: Investigating the dynamics of protein - ligand interactions

University of Copenhagen
daria.gusew@bio.ku.dk

Understanding the processes that control protein dynamics and protein-ligand interactions is important as they occur in all major life processes. Methods that integrate experiments and simulations to refine simulations or interpret experimental data show that tight integration can provide an accurate description of the structural and dynamic processes of proteins. The project aims at combining MD simulations, NMR spectroscopy and X-ray crystallography to quantify the functions of protein motions.

26. Dimitrios Kolokouris: Investigating the role of cholesterol in regulating the SLC7 family of amino acid transporters

University of Copenhagen
dimitrios.kolokouris@sund.ku.dk

Human solute carriers (SLC) are a broad superfamily of transporters with a diverse substrate set. The SLC7 family of amino acid transporters is linked with several pathologies, such as cystinuria - SLC7A9 (b0+AT), autism-related disorders and cancer - SLC7A5 (LAT1), and human pancreatic cancer - SLC7A11 (xCT). These transporters primarily function on the plasma membrane, where cholesterol comprises 20-40% of lipids. Notably, some members of the SLC7 family, which function as heterodimers in the plasma membrane, were solved with steroid densities in their cryoEM structures (rBAT/b0,+AT, 4F2hc/LAT1, 4F2hc/xCT) captured in the inward-open state. In vitro assays of 4F2hc/LAT1 have demonstrated cholesterol-dependent transport activity. We hypothesize that cholesterol may preferentially bind and stabilize the inward-open state via specific sites available on these transporters despite the lack of specific molecular rules for cholesterol binding in the literature. We quantified cholesterol binding patterns with kinetic modeling of three SLC7 human transporters in the membrane and proposed physiological binding sites. Multiscale (atomistic and coarse-grained) MD data suggests these binding pockets to be topologically conserved amongst inward-open structures despite sequence divergence. Geometrical absolute binding free energy calculations and photoclick mass spectrometry demonstrate conserved affinity across two stable, high-residency sites between different SLC7s. Cholesterol binding to these sites was abolished in site mutants, in silico and in vitro, indicative of specific binding. Notably, one of the two sites demonstrated conformational dependence in binding cholesterol. We report that the conformational change, outward → inward-open, results in more stable cholesterol binding, validating and quantifying the ensemble observations from cryoEM. We propose a model where cholesterol preferably binds SLC7s in the inward-open state. These proteins may have evolved to include cholesterol as part of their functional units.

27. Dina Listov: A fully computational design of high-efficiency Kemp eliminase

Weizmann Institute of Science
dina.listov@weizmann.ac.il

We designed a highly efficient Kemp eliminase by computational methods only, achieving kcat=5.5s^{-1}, KM=0.43mM, kcat/KM=13,000$M^{-1}s^{-1}$. The enzyme, a 256 amino-acid TIM-barrel, exhibits Tm=87°C, and high modeling accuracy (<0.5Å RMSD). This was achieved without experimental optimization, ML methods, or evolutionary data in the active site. We focused on foldability and precise active-site design, showing that computational approaches can produce new, stable enzymes with high catalytic efficiency

28. Dominic Rieger: De novo design of protein-binding proteins targeting viral attachment proteins of the paramyxovirus family

University Leipzig
Dominic.Rieger@medizin.uni-leipzig.de

Langya Virus (LayV) and Mojang Virus (MojV) belong to the paramyxovirus family, specifically the henipavirus genus, alongside Nipah and Hendra viruses. While LayV and MojV show milder symptoms, the rapid mutation rate of RNA viruses underscores the need for further research. Currently, no antibodies are known to bind the attachment domains of LayV and MojV. The goal is to computationally design small protein domains that bind these glycoproteins, followed by experimental screening and optimization, to develop a specific, high-affinity, stable, and efficiently expressed protein binder.

29. Elizaveta Maltseva: Design of Degenerate Interfaces for Multiple Geometry-Forming Scaffolds Using Diffusion Models

LMU (University of Munich)
eli.maltseva@campus.lmu.de

My project focuses on designing protein structures with degenerate interfaces that enable the formation of multiple geometries using diffusion models like RFDiffusion. By alternating applied symmetry constraints during denoising, I generate scaffolds matching two distinct assembly geometries from a single block, optimizing the efficiency and adaptability of protein cage self-assembly. This innovative strategy opens new avenues for designing complex and functional protein materials.

30. Ethan Bungay: In silico directed evolution of de novo flavocytochromes

University of Bristol
ethan.bungay@bristol.ac.uk

De novo redox protein design provides an alternate route towards sustainable energy generation and the understanding of natural bioenergetic pathways. The Anderson Group utilises the maquette method to protein design; the use of small, genetically tractable proteins which avoid evolutionary complexity. The diheme four-helix bundle 4D2 has acted as the group's blueprint, being subsequently lengthened and aberrated to configure alternative ligands, with previous work investigating its amenability for covalent flavin attachment. Flavins boast the ability to partake in photomediated electron transfer and bifurcation, yet its amphipathic structure had made attempts to sequester the cofactor homogenously in the lumen of a four-helix bundle difficult.
This work presents a novel pipeline of in silico directed evolution which redesigns an input structure with LigandMPNN and refolds the sequence with ESMFold, iterating upon the best scoring outputs. This method was utilised to create a flavin-binding domain ab initio, yielding high-confidence 4D2-based flavocytochromes. Wet-lab studies prove these designs bolster impressive thermostability, heme-affinity, and structural homogeneity. The latter being evident in the design H60, which is the first de novo flavocytochrome to be structurally resolved through crystallography. The designs contribute to the 'toolbox' of 4D2-originating bioenergetic machinery by providing light-harvesting and energy transduction capabilities.

31. Eva Rajh: Kinetic control of Coiled coil complex dissociation

National Institute of Chemistry, Slovenia
eva.rajh@ki.si

We aim to introduce toehold-mediated strand displacement into coiled coil (CC) protein complexes with overhangs based on short orthogonal coil pairs, enabling precise kinetic control over assembly. To achieve longer dissociation times and more stable complexes without off target states, we developed novel designed CCs with up to 15 heptads or highly stable CC bundles with interface H-bond networks. Kinetic parameters of displacement will be monitored using FRET, split luciferase assays, and SPR.

32. Eva Smorodina: Structural modeling of antibody variant epitope specificity with complimentary experimental and computational techniques

University of Oslo
eva.smorodina@medisin.uio.no

Antibodies are key immunotherapeutic biomolecules with a hallmark feature of antigen-specific binding. However, the principles governing the ability of diverse paratopes to bind to the same epitope with comparable affinity and specificity remain largely unexplained. An insufficient understanding of the structural rules behind antibody-antigen binding, due to a lack of experimentally resolved structures, leads to the current inability to characterize antibody variants binding in silico. Here we propose a rule-based antibody design that relies on a thorough understanding of epitope-paratope interactions, in contrast to generative design based on millions of trials and errors. We identified the binding region of five affinity-verified Trastuzumab variants using cryo-EM and position-resolved HDX-MS and modeled their antibody-antigen interactions in silico. Rigid models alone were insufficient for accurate antibody-antigen representation while molecular dynamics simulations with computational analysis of the complex conformations succeeded in replicating and complimenting experimental findings. Structural parameters calculated based on geometry, surface, and biochemical properties were able to distinguish between high and low variant binders. In addition, we underlined the possibilities of AI in antibody and antibody-antigen structure modeling, demonstrating the limitations of language-based models to predict and understand antibody variants. Overall, our study explains the binding mechanisms of the variant sequences, showing how antibodies with diverse sequences share similar antigen-binding rules.

33. Fabian Schuhmann: The Automated Ligand Searcher (ALISE)

University of Copenhagen, Niels Bohr Institute
fabian.schuhmann@nbi.ku.dk

The Automated Ligand Searcher (ALISE) is designed as an automated computational drug discovery tool. To approximate the binding free energy of ligands to a receptor, ALISE includes a three-stage workflow, with each stage involving an increasingly sophisticated computational method: molecular docking, molecular dynamics, and free energy perturbation, respectively. To narrow down the number of potential ligands, poorly performing ligands are gradually segregated out. The performance and usability of ALISE are benchmarked for a case study containing known active ligands and decoys for the HIV protease. The example illustrates that ALISE filters the decoys successfully and demonstrates that the automation, comprehensiveness, and user-friendliness of the software make it a valuable tool for improved and faster drug development workflows.

34. Fan Cao: A coarse-grained model for disordered and multi-domain proteins

Structural Biology and NMR Laboratory & the Linderstrøm-Lang Centre for Protein Science, Department of Biology, University of Copenhagen
fan.cao@bio.ku.dk

Multi-domain proteins consist of folded domains connected by flexible linkers. Because of their substantial dynamics, characterising the conformational ensembles of multi-domain proteins by simulations may be difficult. We present a coarse-grained model for multi-domain proteins that provides an accurate description of the global conformational properties, as a starting point for understanding interactions between folded and disordered regions and for possible applications on phase separation.

35. Felipe Akihiro Melo Otsuka: Single synonymous scanning mutagenesis of E. coli's protein expression and cell stress profile to measure translation efficiency

Lund University
felipe.otsuka@biochemistry.lu.se

Gene expression is a fundamental process of the cell, making functional proteins while minimizing any burden to the cell. Here is investigated the effect of all single synonymous codon mutants in an E. coli gene known as D-alanine—D-alanine ligase A (ddlA) to understand the impact of codon usage on protein expression and its impact on cell stress. Moreover, prediction of 5′ UTR mRNA structures from synonymous variants will be modelled to identify key interactions with the ribosome binding site.

36. Fernando Meireles: Guiding structure-based protein design with UniProt annotations

EPFL
fernando.meireles@epfl.ch

Structure-based protein design methods allow us to explore the sequence space that should fold into a specific 3D conformation. These methods design sequences only constrained by the geometry of the input protein, which may not retain functional information. We aim to enrich our recently developed protein design method, CARBonAra, with features extracted from the UniProt database, such as GO terms, organism, and localization to ensure the designed proteins retain functional information.

37. Francesco Altiero: In-silico Workflow to Design De-novo Antigens by Epitope Scaffolding

Università degli Studi di Napoli "Federico II"
francesco.altiero@unina.it

De-novo protein design allows a more accurate exploration of the folding space than directed evolution and has provided remarkable benefits to several fields of medicine and biochemistry, such as in-vitro diagnostics (IVD). Specifically, the IVD industry still relies on antibodies due to their cheap price and established production methods. The design of de-novo antigens with high specificity can reduce the time of production of novel antibodies, while increasing their effectiveness. Thanks to the recent advances in the AI field, the chance of success-rate of in-silico de-novo design significantly increased, particularly due to the different models available in the Rosetta suite. We present a workflow focused on the generation of small-protein scaffolds (~50AA) around a known antigen epitope to design de-novo antigens which can elicit higher immune response in host organisms. The workflow begins with a generation phase, using RFdiffusion to generate the backbone structure and ProteinMPNN to sample several sequences from these backbones. A subsequent validation phase leveraging AI folding models, as OmegaFold and AlphaFold, aims to filter the sequences which are most likely folded to their original backbones. These two phases are cyclically iterated until a high rate of folding accuracy is predicted. The obtained proteins are subsequently screened in-silico to evaluate their immunogenicity, to retain a small set of (3 to 6) designed antigens with the highest predicted values. The workflow is fully generalizable to any given epitope and, as a proof of concept, we applied this workflow to scaffold around two epitopes of HPV-16 and one epitope of SARS-CoV-2 obtaining small de-novo antigens for each epitope. For one epitope of HPV-16, the resulting set of 6 de-novo antigens is currently being screened in-vitro.

38. Giulio Tesei: Conformational ensembles of the human intrinsically disordered proteome

University of Copenhagen
giulio.tesei@bio.ku.dk

Intrinsically disordered regions (IDRs) comprise about one third of the human proteome and play key roles in biological processes. While lacking fixed 3D structures, IDRs adopt conformational ensembles determined by their amino acid sequences, which current prediction methods yet cannot capture. Using simulations and bioinformatics, we generated a proteome-wide database of IDR ensembles and explored sequence-ensemble-function relationships as well as the evolutionary conservation of IDRs' conformational properties.

39. Gustav Oberdorfer: Computational design of highly active de novo enzymes

Institute of Biochemistry, Graz University of Technology
gustav.oberdorfer@tugraz.at

Reliably introducing function into genetically encodable de novo proteins is still a challenging task. Current design methods mostly produce de novo enzymes with low reactivities. As a result, they require costly experimental optimization and high-throughput screening to be industrially viable. We developed rotamer inverted fragment finder–diffusion (Riff-Diff), a hybrid machine learning and atomistic modelling strategy for scaffolding catalytic arrays in de novo protein backbones. We show that proficient enzymes can be generated with Riff-Diff while screening as little as 35 designs. The talk will highlight challenges and findings during scaffolding active sites for catalyzing the retro-aldol, Morita Baylis-Hillman reaction and metal cofactors of increasing complexity.

40. Gustavo Araiza: Epitope-focusing and viral escape prediction for the design and selection of robust influenza antibodies

Vanderbilt University
gustavo.araiza@vanderbilt.edu

Influenza is a seasonal disease caused by Influenza A viruses (IAVs) that accounts for up to 500,000 deaths per year, globally.1 Healthy individuals can fight off IAV infection through their adaptive immune systems that most commonly target the immunodominant protein hemagglutinin (HA), of which there are 18 known subtypes (H1-H18). Immunocompromised populations are at higher risk of hospitalization and death resulting from influenza. Therapeutic antibodies (Abs) can be administered to infected.

41. Han Tang: A Diffusion Model Approach to Propose Non-Canonical Amino Acid Substitutions in Protein Engineering

Department of Computer Science (DIKU), University of Copenhagen
ht@di.ku.dk

Current inverse folding models can propose candidate mutations for amino acids at specific protein sites to enhance functionality. However, these models are restricted to the standard set of 20 natural amino acids. Studies have shown that modifying natural amino acids or incorporating non-canonical amino acids can significantly enhance protein performance in specific contexts. To overcome this limitation, we propose a strategy that leverages molecular generative diffusion models to suggest non-canonical amino acid substitutions. Our approach involves training a continuous sampling space for side-chain atom sets conditioned on the proteomic pocket and validating this model using Deep Mutational Scan (DMS) datasets. This continuous sampling space is generated by training a molecular diffusion model to inpaint side-chain atoms using data from multiple sources. Our results indicate that the continuous sampling space could align with known mutational data from DMS experiments, demonstrating its potential to identify novel and effective amino acid substitutions.

42. Haye Nijhuis: Producing and stabilizing recombinant surface glycoproteins from highly pathogenic zoonotic viruses

University of Amsterdam, Department of Medical Micriobiology and Infection Prevention, Institute for Infection and Immunity, Infectious Diseases.
h.nijhuis1@amsterdamumc.nl

Zoonotic viruses cause many minor and major highly pathogenic outbreaks each year. Since the hosts of such dangerous viruses generally exist in warmer climates, those viruses pose an increasing threat with the rise of global temperatures. As such, the WHO designated several viruses as priority pathogens. Among them are Arenaviruses, Henipaviruses, Marburg virus, and Crimean-Congo Hemorrhagic Fever virus (CCHFV), which combined cause thousands of deaths annually. Suitable therapeutics are rare, not yet officially commercially available, only used sporadically, and only used when absolutely necessary, underscoring the need for better countermeasures against these viruses. In this study, we successfully produced Old- and New-world Arenavirus glycoprotein complex trimers (GPC), Hendra and Nipah virus fusion glycoprotein trimers (F), Marburg virus glycoprotein trimers (GP) and CCHFV Gn, Gc, and GP38 glycoprotein monomers. Trimeric constructs were designed by linking the extracellular domain of the protein to an I53-50A trimerization domain. In addition, Henipavirus F and Arenavirus GPC were stabilized by introducing helix-breaking and cavity-filling amino-acid substitutions, and Arenavirus GPC and Marburg virus GP proteolytic cleavage was improved by introducing or optimizing an already existing furin cleavage site. SDS-PAGE, Blue Native-PAGE, size exclusion chromatography, and ELISA with neutralizing and/or quaternary-structure-specific monoclonal antibodies (mAbs) were used to evaluate protein folding. Our constructs appear to be native-like and harboring known neutralizing epitopes. By adding I53-50B, we successfully assembled the proteins into I53-50 nanoparticles, which show a favorable immunogenicity profile compared to recombinant protein alone. Our nanoparticles serve as interesting candidates for future vaccination studies, for some viruses being the first of their kind. Additionally, recombinant glycoproteins can be used to isolate and characterize new and potentially protective mAbs from convalescent patient- or vaccinated animal PBMCs which in turn can identify viral glycoprotein vulnerabilities. Taken together, novel mAbs and vaccine candidates can restrain the rising global health threat that these viruses pose.

43. Ho Yeung Chim: Computational design of small molecule-dependent cyclic oligomers

Ludwig Maximilian University of Munich
h.chim@lmu.de

Designing protein oligomers responsive to small molecules is a key challenge in dynamic protein assembly design. I developed a new RFdiffusion module for designing cyclic oligomers by using pre-existing interfaces, focusing on scaffolding symmetrized target interfaces to maintain interface properties in the final design. I computationally designed cyclic oligomers with different symmetries derived from neomorphic PPIs, as a first step towards creating controllable and dynamic protein assemblies.

44. Holly Ford: Computational design and biochemical characterisation of a molecular wire

University of Bristol
h.ford@bristol.ac.uk

A single chain tetrahelical bundle binding two b-type hemes has been well characterised. Duplication of the helical segments yields longer proteins that bind 4 or even more hemes, forming linear redox chains. But these proteins are highly insoluble, prone to aggregate and reluctant to bind heme at every site. New deep learning based protein sequence design and structure prediction methods that model ligands, were used to redesign these proteins.

45. Horst Lechner: Improving protein expression using ProteinMPNN

Graz University of Technology
horst.lechner@tugraz.at

We want to produce successfully fungal proteins in non-fungal heterologous expression hosts. We are developing a method based on ProteinMPNN, RFdiffusion and Rosetta to adapt the sequences of a model protein to preserve its function and selectivity but enable high production yields from bacterial host organisms

46. Hubert Klein: In-Silico Design of Foldamer – Protein Interactions

Ludwig-Maximilians-Universität
hubklph@cup.uni-muenchen.de

Protein-protein interactions are common biomedical targets, as they are involved in various diseases. We address the challenge of targeting flat protein-protein interaction surfaces, which are inherently difficult to inhibit, by proposing the use of non-natural aromatic delta peptides, namely foldamers. Foldamers give us complete freedom of choice relative to amino acid sidechains, while retaining design simplicity, as the foldamer secondary structure is robust to sidechain identity changes.
Here, we explore the use of Rosetta for the design of foldamer-binding proteins, to experimentally verify whether the Rosetta scoring function can properly rank interactions between proteins and foldamers. In the future, this will help us establish a computational pipeline for the design of foldamers to target specific PPIs.

47. Ida Kjærsgaard Grene: De novo protein design for neuroscience

Aarhus University
ikg@mbg.au.dk

Developing de novo designed protein binders of neuronal targets using RFdiffusion and Protein MPNN. Testing the affinity and specificity of these using both in-vitro biophysical methods and cell-culture assays. Future goals include testing the effect of using binders to block and/or force protein-protein interactions on cell signaling.

48. Ignacio Retamal: Autonomous motor design, modulating track binding affinity through allostery

University of New South Wales
i.retamal@unsw.edu.au

We are taking a modular approach to designing autonomous motors. We will start by selecting non-motor parts and assembling them from the bottom up. Our goal is to understand the mechanochemical cycle and energy transduction events of natural linear motor proteins. In this proposal, we aim to use effector domains to control binding affinity, while utilizing stalk-coiled coils as intermediate structures to allosterically modulate track binding affinities

49. Inken Fender: Predicting Molecular Hotspots for Efficient Drug Discovery using Deep Learning

University of Bern
inken.fender@unibe.ch

The limited availability of detailed protein-ligand complex data hinders the development of highly selective drugs with strong binding affinity to their target proteins. Recent research showed position and identity of an interacting chemical group can be predicted from local backbone position and identity alone. Our method extends on this idea and harnesses a CNN trained on micro-environments to predict molecular hotspots.

50. Iori Morita: De novo design of a programmable sequence-specific ssRNA binding protein

Institute of Science and Technology Austria
iori.morita@ista.ac.at

Detection and regulation of ssRNA is of great interest in synthetic and cell biology. In this context, it is highly desirable to develop a programmable ssRNA binding protein. Taking inspiration from natural RNA binding proteins, we propose to develop a ssRNA binding protein that is highly programmable.

51. Jacopo Sgrignani: AI based design of human lipocalin ligands

AInstitute for Research in Biomedicine
jacopo.sgrignani@irb.usi.ch

Neuroinflammation, a complex immune response in the CNS, involves glial cell activation and inflammatory mediator release, playing roles in both protection and disease progression, particularly in conditions like Alzheimer's disease. Lipocalin-2 (LCN2) is a key player in neuroinflammation, contributing to blood-brain barrier disruption and exacerbating inflammation through its interaction with astrocytes.
Starting from this experimental observation, we used AI based tools, to design mini protein able to bind LCN2 and block its interaction with the receptor.
The more promising binders were expressed and purified and their ability to bind the target protein verified by biophysical assay. The experimental results revealed that one of the mini protein can bind LCN2 with an affinity constant in the low nanomolar range.

52. Jakob Riccabona: Development of a computational vaccine design pipeline using machine learning – focus on stabilizing proline substitutions

ScaDS.AI Dresden/Leipzig (Center for Scalable Data Analytics and Artificial Intelligence)
Jakob.Riccabona@medizin.uni-leipzig.de

Targeting viral surface proteins is a key cornerstone of therapeutic antibody and vaccine development. To increase the stability of these proteins, proline mutations are strategically introduced, leveraging on the unique structural constraints of this amino acid. Here, we present proliNNator, a graph neural network capable of predicting position-specific proline probabilities within protein structures. Our tool outperforms existing AI-based sequence design tools in predicting engineered proline.

53. Jakob Nilsson: Substrate recognition principles for the PP2A-B55 protein phosphatase

University of Copenhagen
jakob.nilsson@cpr.ku.dk

The PP2A-B55 phosphatase regulates a plethora of signalling pathways throughout eukaryotes. How PP2A-B55 selects its substrates presents a severe knowledge gap. By integrating AlphaFold modelling with comprehensive high resolution mutational scanning, we show that a-helices in substrates bind B55 through an evolutionary conserved mechanism. Despite a large diversity in sequence and composition, these a-helices share key amino acid determinants that engage discrete hydrophobic and electrostatic patches. Using deep learning protein design, we generate a specific and potent competitive peptide inhibitor of PP2A-B55 substrate interactions. With this inhibitor, we uncover that PP2A-B55 regulates the nuclear exosome targeting complex (NEXT) by binding to an a-helical recruitment module in the RNA-binding protein 7 (RBM7) a component of the NEXT complex. Collectively, our findings provide a framework for the understanding and interrogation of PP2A-B55 function in health and disease.

54. Janina Sörmann: Leveraging Deep Learning tools to design hASIC1a miniprotein modulators for stroke treatment

University of Copenhagen
janina.sormann@sund.ku.dk

Stroke is a leading cause of death and adult disability, but current treatment options are limited despite the pressing demand for novel therapeutic strategies. Acid sensing ion channels (ASICs), particularly ASIC1a, are promising targets in the pathophysiology of ischemic stroke, which accounts for ~80% of all stroke cases. ASICs are primarily expressed in the central and peripheral nervous systems, where they contribute to synaptic transmission and thereby mediate crucial processes such as synaptic plasticity, learning, memory, fear, and pain processing. However, prolonged tissue acidification as a result of stroke or inflammation can lead to pathological activation of ASICs, contributing to tissue damage and cell death. ASIC1a knock out or pharmacological inhibition by venom-derived peptides like PcTx1 or Hi1a significantly reduces infarct volumes, underscoring the channel's role in acid-induced neurotoxicity and highlighting its potential as a therapeutic target. Nevertheless, ASICs remain underexplored as drug targets due to the lack of potent and selective inhibitors. To address this, we computationally designed miniproteins specifically targeting hASIC1a using a deep learning-based pipeline based on RFdiffusion, ProteinMPNN, and AlphaFoldMultimer. The miniprotein designs were targeted against different areas in the hASIC1a extracellular domain to achieve hASIC1a inhibition or enhancement, subtype selectivity, and high affinity. They were optimised for in vivo stability and solubility, making them promising candidates for therapeutic development. The miniproteins were subcloned, expressed and purified before being screened in automated patch-clamp experiments. A screen of 96 miniproteins yielded in 5 inhibitors and potentiators, some with nanomolar potency. Notably, the pipeline we developed for hASIC1a modulators, integrating computational tools with a versatile expression and purification workflow and experimental validation, can be adapted to other membrane protein targets. This creates a flexible platform for drug discovery across conditions involving ion channel dysfunction.

55. Jannick Prentoe: The hepatitis C virus envelope protein complex is a dimer of heterodimers

ISIM, SUND, UCPH
jprentoe@sund.ku.dk

Worldwide, 58 million individuals suffer from chronic hepatitis C virus (HCV) infection, which is a primary driver of liver cancer and for which no vaccine is available. The HCV envelope proteins, E1 and E2, form a heterodimer, which is the target for neutralizing antibodies (NAbs). However, the higher order organization of these E1/E2 heterodimers, as well as that of any Hepacivirus envelope protein complex, remains unknown. Here, we determined a ~3.5 Å cryo-electron microscopy structure of two E1/E2 heterodimers in a homodimeric arrangement. We reveal how the homodimer is established at the molecular level and provide insights into HCV NAb evasion and membrane fusion, as orchestrated by E2 motifs such as hypervariable region 1 and antigenic site 412, as well as the organization of the transmembrane helices, including two internal to E1. The study addresses long-standing questions on the higher order oligomeric arrangement of Hepacivirus envelope proteins and provides a critical framework in the design of novel HCV vaccine antigens.

56. Joel James Chubb: RASSCoL: Simplifying Computational Design of Small-Molecule Binding Pockets

University of Bristol
joel.chubb@bristol.ac.uk

Designing proteins with precise small-molecule binding capabilities is a current challenge in protein design, with implications for drug discovery, biosensing, and synthetic biology. Current methods often involve complex, time-consuming approaches requiring levels of in vitro screening not widely accessible and that may not yield the desired specificity or affinity. To address these limitations, we present RASSCoL (Rapid Assessment of Size and Shape Complementarity of Ligands). RASSCoL is a two-stage computational approach that integrates volume displacement with fast computational screening to efficiently design small-molecule binding pockets in de novo proteins. In the first stage, we employ exhaustive sampling and docking of potential sequence candidates generated using a volume displacement technique that reshapes binding pockets to enhance their complementarity with target molecules. This method allows us to rapidly generate a pool of candidates, each tailored to accommodate specific small molecules. The second stage involves accelerated molecular dynamics simulations to evaluate and refine these candidates, ensuring both accuracy and binding affinity. Our approach has successfully produced binders with nanomolar (nM) to micromolar (μM) affinities for a selection of fluorescent molecules, demonstrating specificity and selectivity. Furthermore, we have engineered dual binders from these designs capable of energy transfer between molecules, illustrating the functional versatility of our designs. Building on these successes, we are now extending our pipeline to more complex targets, including pharmacologically relevant compounds. This highlights the potential of our design pipeline to contribute to the development advanced biosensors and function assays. Our streamlined approach represents an advance in the field of protein design, offering a new tool for the rational design of protein-based binders with high specificity and functionality.

57. Johannes A. Klier: Exploring de novo-likeness of natural proteins to improve protein design

Leipzig University, Medical Faculty
klier@izbi.uni-leipzig.de

De novo proteins often display high thermostability and rigidity, limiting their flexibility. To design more natural proteins in the future, we study de novo-like proteins as a bridge between natural and de novo proteins. Using a classifier to assess the de novo-likeness of natural proteins, we find that early evolutionary organisms have a higher proportion of these proteins. This suggests that current de novo proteins may represent an early evolutionary stage in protein development.

58. Johannes Peterlechner: Enhancing Protein Loop Model Quality with a Rotationally Equivariant Variational Autoencoder (REVA)

Institute of Biochemistry, Graz University of Technology
peterlechner.johannes@gmail.com

Despite recent advancements in protein structure prediction, accurately modeling protein loops - particularly those that are surface-accessible - remains challenging. These regions are often crucial for protein function, serving as active sites, binding interfaces, or determinants of molecular specificity, such as in antibodies. Hence, optimizing loop structures could significantly enhance our understanding of protein interactions and facilitate drug discovery. In this study, we propose a graph-based variational autoencoder (VAE) for refining the structure of loop models. Our method utilizes an equivariant 3D-CNN encoder that captures an all-atom representation of the loop and encodes it into a latent space. A decoder then reconstructs a refined loop structure from this latent representation. So far, we compiled a dataset including 5087 CDR loops and initiated our first training runs, setting the foundation for our Rotationally Equivariant Variational Autoencoder (REVA).

59. Johannes Schweininger: In vitro and in silico inhibitor design against the HCMV pUL50-pUL53 interaction

Yves Muller Lab, FAU Erlangen-Nuremberg
johannes.schweininger@fau.de

As a herpesvirus, the human cytomegalovirus HCMV is a ubiquitous pathogen. The proteins pUL50 and pUL53 form the basis of the HCMV nuclear egress complex, which mediates transport of capsids from the host cell nucleus to the cytoplasm. Directed evolution was used to identify inhibitory peptides against the pUL50-pUL53 interaction. RFdiffusion and ProteinMPNN are now explored for the generation of binders and scaffolds for affinity maturation using fluorescence complementation and yeast display.

60. Johnny Alexander Jimenez Siegert: Modeling ASIP-Melanocortin Receptor Interactions for Anti-Obesity Drug Design

Leipzig University
johnny.jimenez_siegert@uni-leipzig.de

The Melanocortin-4 Receptor (MC4R), one of five human melanocortin receptors, plays a crucial role in energy homeostasis and its impairment is a leading cause of monogenic obesity. Overexpression of Agouti Signaling Protein (ASIP), a natural antagonist to melanocortin receptors, has recently been identified as a novel cause of obesity in humans. However, no experimental structure of an ASIP-receptor complex is available to guide drug design. Using AlphaFold-Multimer, AlphaFold3 and Rosetta, we modeled the interaction of ASIP with melanocortin receptors, recapitulating observations from various mutagenesis studies and identifying key interaction sites. These models provide a basis for the design of receptor-specific peptide and small molecule ligands for the treatment of ASIP overxpression.

61. Jonas Gunkel: Computational design and experimental testing of Polyester Hydrolase Leipzig 7 (PHL7) mutants on different bioplastics

Institute for Drug Discovery (University Leipzig)
jonasgunkel99@gmail.com

Plastics are widely used due their properties and versatilities, making them an integral part of everyday life. Nevertheless, a significant quantity of plastics ends up littering the environment, highlighting the need for closed-loop recycling methods such as enzyme-based hydrolysis of post-consumer plastics.
In this work we applied a computational method to a known PET-degrading enzyme, Polyester Hydrolase Leipzig 7 (PHL7), with the objective to increase it's activity towards different plastics such as PET, PBS, and PLA. The computational method used for the diversification of the enzyme is an automated method that combines phylogenetic analysis and Rosetta design calculation called FuncLib. A ligand docking procedure with three plastic (PBS, PET, and PLA) was carried out to select mutants that would likely be stable and maintain or increase its activity compared to PHL7. Mutants with a predicted improved interaction with the plastic ligands were subjected to experimental testing, which included the expression of the mutants in bacterial cultures, their purification in small scales, measurement of stability and plastic degradation rates of various plastic films using electrochemical impedance spectroscopy (EIS).

62. Jonathan Funk: ProteusAI: An Open-Source and User-Friendly Platform for Machine Learning-Guided Protein Design and Engineering

DTU Bioengineering
jonfu@dtu.dk

Protein design and engineering are crucial for advancements in biotechnology, medicine, and sustainability. Machine learning (ML) models are used to design or enhance protein properties such as stability, catalytic activity, and selectivity. However, many existing ML tools require specialized expertise or lack open-source availability, limiting broader use and further development. To address this, we developed ProteusAI, a user-friendly and open-source ML platform to streamline protein engineering and design tasks. ProteusAI offers modules to support researchers in various stages of the design-build-test-learn (DBTL) cycle, including protein discovery, structure-based design, zero-shot predictions, and ML-guided directed evolution (MLDE). Our benchmarking results demonstrate ProteusAI's efficiency in improving proteins and enyzmes within a few DBTL-cycle iterations. ProteusAI democratizes access to ML-guided protein engineering and is freely available for academic and commercial use. Future work aims to expand and integrate novel methods in computational protein and enzyme design to further develop ProteusAI.

63. Jonathan Hungerland: The rhodopsin GPCR complex in European Robin: computational construction and validation

University of Oldenburg
jonathan.hungerland@uol.de

We present the first complete model for an avian rhodopsin complex. Due to a lack of both experimental and computational data for avian rhodopsin GPCR models, it is an intriguing benchmark system for protein complex prediction. Besides this, the protein complex may be involved in avian magnetoreception, possibly allowing the magnetosensitive Cryptochrome4 to piggy-back on the signal transduction mechanism of the rhodopsin GPCR complex.

64. Jonathon Liston: Designing cobalt specific protein “Velcro”

Durham University, UK
Jonathon.D.Liston@durham.ac.uk

It is desirous when building protein cages and other structures to have designed connections between monomers which are specific and enable controlled assembly and disassembly. By using a cobalt binding protein from Salmonella typhimurium as a template for RFDiffusion, we have attempted to design a pair of proteins which do not interact without metal present, but dimerise in the presence of cobalt. The release of AlphaFold3 facilitated the in silico screening of these designs. The ability of these proteins to dimerise were assessed via Native PAGE and DLS. Next steps upon successful dimerization is to design novel protein cages which would assemble in a cobalt-mediated manner with these proteins as the interfaces.

65. Jonathan Ziegler: Conditioning Generative Models for Protein Optimization

Cradle
jonathan@cradle.bio

Pretrained protein large language models (PLMs) have become successful at generating function-preserving variants of a template protein. This makes them very useful for lead optimization, where they can be combined with property-predicting models to define new candidates to evaluate in vitro. However, plausible mutations suggested by PLMs aren’t necessarily beneficial for the optimization across properties, leading to an inefficient search in protein space. We show how conditioning these models to bias toward optimization goals can significantly increase the efficiency of the search space exploration. We present our latest improvements on a lab-in-the-loop active learning protein optimization pipeline. We obtained significant performance gains by combining semi-supervised training on evolutionary data with supervised fine tuning using Direct Preference Optimization to enhance protein language models for sequence generation. This approach improves the alignment of model likelihoods with experimental fitness data, facilitating protein optimization tasks by generating libraries of stronger candidates. We have demonstrably applied this method to over 25 in-silico protein generation campaigns on both internal and publicly available data, where our automated pipeline consistently produced robust models with a measurable increase of pseudo likelihood correlation with variant fitness. (Continued on next page).

To further validate this approach and test the generative ability of the models, we conducted in-vitro experiments across five distinct protein optimization projects, covering peptides, enzymes, and antibody fragments. The score distribution of the candidate libraries shifted significantly toward higher-fitness proteins, confirming the effectiveness of this fine-tuning method in producing strong leads without decreasing diversity.

66. Joshua Behringer: Protein functionalization towards enzyme selectivity

Technische Universität München
joshua.behringer@tum.de

While the field of protein design has seen significant process, the de novo design of functional enzymes remains challenging. Here, we describe the functionalization of a de novo designed lanthanide-binding protein, turning it into a cerium-dependent photoredox enzyme. Most recently, we achieved significant improvements in kcat/Km and enantioselectivity by computational redesign with RFDiffusion and ProteinMPNN.

67. Joy Lyu: AI-guided identification of novel extracellular protein interactions between wheat and Zymoseptoria tritici

University of Oxford
joy.lyu@stx.ox.ac.uk

We use AlphaFold Multimer to screen for putative interactions between wheat extracellular hydrolases and small secreted proteins from its pathogen Zymoseptoria tritici. We have established high through-put pipelines to run AlphaFold as a fast screening tool for putative interactions. We then select candidates based on the model scores (ipTM+pTM, pLDDT, etc.) from AlphaFold, and manually pick ones that are likely to be enzyme-inhibitor interaction to validate with biochemical assays

68. Julia K. Varga: Substrate recognition principles for the PP2A-B55 protein phosphatase

Hebrew University of Jerusalem
julia.varga@mail.huji.ac.il

The PP2A-B55 phosphatase regulates a plethora of signaling pathways throughout eukaryotes. How PP2A-B55 selects its substrates presents a severe knowledge gap. By integrating AlphaFold modeling with comprehensive high-resolution mutational scanning, we show that α helices in substrates bind B55 through an evolutionary conserved mechanism. Despite a large diversity in sequence and composition, these α helices share key amino acid determinants that engage discrete hydrophobic and electrostatic patches. Using deep learning protein design, we generate a specific and potent competitive peptide inhibitor of PP2A-B55 substrate interactions. With this inhibitor, we uncover that PP2A-B55 regulates the nuclear exosome targeting (NEXT) complex by binding to an α-helical recruitment module in the RNA binding protein 7 (RBM7), a component of the NEXT complex. Collectively, our findings provide a framework for the understanding and interrogation of PP2A-B55 function in health and disease.

69. Julian Beck: Scaffolding a Kemp eliminase activity into an idealized de novo TIM barrel

University Bayreuth
julian.beck@uni-bayreuth.de

One milestone in protein design was the design of the first de novo TIM barrel. However, the introduction of functions into de novo TIM barrels is challenging due to their idealized scaffolds compared to natural TIM barrels. We developed a multistep design workflow combining physics- and AI-based tools including Triad and RFdiffusion to scaffold a Kemp eliminase activity into a de novo TIM barrel. 1 out of 5 designs showed an activity comparable to previous first design rounds of Kemp eliminases.

70. Juno Underhill: Spectral tuning in a de novo designed riboflavin protein

University of Bristol
juno.underhill@bristol.ac.uk

The de novo “maquette” heme b binding protein 4D2, and its designed and evolved variants, provide an expandable, modular de novo protein platform for precision redox engineering. 4D2 has been redesigned to covalently bind a riboflavin chromophore in order to introduce photoresponsivity to the system. In this work we use a QM/MM model that can accurately describe spectral properties of this de novo riboflavin protein which can provide dynamic insight into the "structure-spectrum problem".

71. Justin Booth: Evolving Allostery: Harnessing the Power of Continuous Evolution for Dynamic Protein Engineering

The Biodesign Lab | Francis Crick Institute
justin.booth@crick.ac.uk

Allostery is the constant dialogue between natural proteins and their surroundings, the key secret underpinning the dynamic functionality we observe in evolution’s most prolific high-achievers. Continuous directed evolution techniques have long promised to be groundbreaking tools in protein engineering, but their inability to evolve complex allosteric function has limited their utility. As they can only provide evolutionary pressure in a single environment, they cannot select for phenotypes which require varied activity in different environments — which constitute some of the most interesting and high-value targets for optimisation. We present early-stage work on a system based on Phage and Robotics Assisted Near-Continuous Evolution (PRANCE) to enable the high throughput evolution of proteins towards diverse allosteric functions. By combining lab automation with phage-assisted continuous evolution, iterative cycles of simultaneous mutagenesis and selection can be performed to provide strong evolutionary pressure towards protein activity that changes according to the environment, allowing a protein’s allosteric response itself to optimised. We demonstrate this system with the evolution of an allosteric transcription factor and an aminoacyl-tRNA synthetase, and expect to further apply it towards the evolution of novel targets including optogenetic proteins and allosteric enzymes. We are also looking with particular fascination — in the context of recent advances towards de-novo ligand binder design — at its potential as a platform for the rapid development of novel single-protein biosensors.

72. Karel van der Weg: Towards the construction of a large foundation model for protein structures with message-passing Graph Neural Networks

Institute of Bio- and Geosciences (IBG-4: Bioinformatics)
k.van.der.weg@fz-juelich.de

Recent advances in biology and machine learning have revolutionized the field by enabling the development of large foundation models (LFMs) trained on vast biological data sets. Notable LFMs such as Evolutionary Scale Modeling (ESM) train using billions of sequences. Due to the increased availability of computational generated structure data, we now have the possibility to investigate LFMs for protein structure. In this project we create an LFM trained on data from AlphaFold2, ESM, and experimental structures, totaling 986 million structures. By incorporating structure in the LFM we aim to establish a pre-trained model for researchers to use on any structural related protein task. We will assess the quality of the LFM on three different structural protein tasks: Enzyme commission number prediction, biological assembly prediction and generative function design. Using a message passing GNN we plan to learn protein structures at different levels of representation. Our first accomplishment is the creation of a ready-to-train database containing this information. Secondly, we are investigating three different methods to scale up our LFM, two message passing GNNs (ProNet, GearNet) and one 3D transformer-based model (Uni-Mol). We are testing the models for scaling along the network depth, database size and input model protein quality. While the GNNs show promising results, we see no scaling behavior while the 3D transformer-based models scale with the input data size.

73. Kristoffer E. Johansson: Global Analysis of Multi-Mutants to Improve Protein Function and Stability

University of Copenhagen
kristoffer.johansson@bio.ku.dk

Stabilizing proteins without otherwise hampering their function is a central task in protein engineering and design. Multiplexed assays can screen thousands of protein variants for properties and at conditions that are relevant for the engineering task. We demonstrate that assay data of multi-mutant libraries are highly informative for engineering purposes. To do this, we have developed a global multi-mutant analysis (GMMA) that can disentangle the effects of individual amino acid substitutions.

74. Leonardo De Maria: PepInvent: generative models for therapeutic peptide optimization

AstraZeneca
leonardo.demaria@astrazeneca.com

The therapeutic potential of peptides can be further unlocked with the help of artificial intelligence. Non-natural amino acids (NNAAs), amino acids not encoded by DNA, have been used to enhance many peptide properties from binding affinity, protease, and plasma stability to half-life. Incorporating NNAAs has enabled venturing beyond the chemical space dictated by the 20 natural amino acids and into the small molecule-like chemical space. Recently we showed how very large libraries of completely new alpha-amino acids can be enumerated and in-silico screened for improved binding affinities. In this work, we present a novel generative AI-based tool, PepINVENT as an extension to the small molecule molecular design platform, REINVENT. (Continued on next page).

Our tool aims to facilitate de novo design of peptides with enhanced properties through reinforcement learning driven generation. With its chemistry-aware generative capabilities, PepINVENT explores chemical spaces beyond those of enumerated amino acid libraries. It produces novel, diverse and chemically feasible amino acids with chirality mutations, backbone modifications, and sidechain alterations. In addition to the novel amino acid designs, our generative framework offers structural flexibility by exploring diverse peptide topologies or constraining the generation to a specific topology. We have developed a range of learning objectives that are specific for multi-parameter optimization of properties on both amino acid- and peptide-level and applied it to several therapeutically relevant peptides. To enable this work in the laboratory, we have also developed tools for chem-informatically protect any novel amino-acid to be used in solid-phase synthesis and to score the synthesizability of the so obtained molecule by a suitable retrosynthesis score.

75. Leoni Abendstein: Designing Synthetic Ortholog Heterodimers: Challenges and Opportunities in De Novo Protein Engineering

Institute of Science and Technology
leoni.abendstein@ist.ac.at

Ortholog heterodimers, found in various species, play critical roles in numerous biological processes, making them an intriguing target for synthetic biology. These natural protein pairs exhibit conserved interactions across species, offering valuable insights for designing de novo heterodimers with similar or enhanced functionalities. However, replicating the precise structural and functional compatibility of these natural pairs presents significant challenges. Our study addresses these challenges by developing innovative design strategies that aim to create synthetic ortholog heterodimers, with the goal of expanding their applications in biotechnology and synthetic biology.

76. Leonie Windeln: Turning toxins into targets: design of α-conotoxin capturing proteins

University of Southampton
l.windeln@soton.ac.uk

α-Conotoxins are naturally occurring peptides, that are of pharmacological interest, but their potency also poses risks without available anti-toxins. Inspired by recent advances in AI-driven protein design, we generated proteins that bind to α-Conotoxins. For this we tested different scaffold constrains and applied stringent filtering for relevant protein-peptide interactions. We plan to validate our best designs experimentally and improve them using AI and medicinal chemistry methods.

77. Liza Ulčakar: Characterization of designed cyclic oligomers that walk on designed protein fibers

National Institute of Chemistry
liza.ulcakar@ki.si

The goal of our group is to design walker proteins that can move randomly along a designed protein fiber. We have successfully designed fibers and walkers with corresponding binding sites (‘feet’). The rollers differ in the binding affinity and number of feet. We hypothesize that the walkers with higher affinity for the fibers will move slower compared to the walkers with lower affinity. We characterize the system using single molecule TIRF tracking, negative-stain-EM, cryo-EM and SPR.

78. Lorenzo Scutteri: De Novo Design Light-Responsive Protein Switches

EPFL
lorenzo.scutteri@epfl.ch

Designing proteins that can dynamically and reversibly switch between different states remains challenging. To address this, we developed a computational framework for engineering light-mediated switchable properties in target molecules. We leveraged the LOV2 optogenetic domain to control the conformational dynamics of de novo designed scaffolds, creating light-responsive protein switches. We anticipate that functionalizing these switches will lead to impactful biological applications.

79. Lucien Krapp: PeSTo and CARBonAra: the secret sauces for protein design

EPFL
lucien.krapp@epfl.ch

In structural biology, predicting protein interactions and designing sequences using backbone scaffolds are crucial yet challenging tasks. The Protein Structure Transformer (PeSTo) and its derivative, CARBonAra, both built on the same deep learning framework, tackle these challenges effectively. PeSTo utilizes geometric transformers to accurately predict a variety of protein binding interfaces, setting new standards in both accuracy and computational efficiency. It supports high-throughput analyses and integrates seamlessly with the extensive AlphaFold foldome. CARBonAra, derived from PeSTo, focuses on sequence recovery from backbone scaffolds and uniquely considers non-protein entities such as nucleic acids and ligands. These methods offer a combination of speed, accuracy, and broad applicability, paving the way for significant advancements in structural biology and biotechnology.

80. Luis I. Gutierrez-Rus: PiLGriM: A computational pipeline for metalloprotein design

University of Bristol
luis.gutierrezrus@bristol.ac.uk

De novo metalloprotein design is a promising approach to address the design of new enzymes from scratch. Here we present PiLGriM, a computational pipeline that combines rational design and AI approaches to generate catalytic metal-binding sites. PiLGriM scans protein scaffolds for sites that are best suited for metal binding by considering all combinations of ligands and geometries for a specific target metal. We present the method and experimental characterization of our first designs.

81. Luise Jacobsen: Clues about Uncoupling Protein 1 Mechanisms from Molecular Simulations

University of Bristol
luis.gutierrezrus@bristol.ac.uk

De novo metalloprotein design is a promising approach to address the design of new enzymes from scratch. Here we present PiLGriM, a computational pipeline that combines rational design and AI approaches to generate catalytic metal-binding sites. PiLGriM scans protein scaffolds for sites that are best suited for metal binding by considering all combinations of ligands and geometries for a specific target metal. We present the method and experimental characterization of our first designs.

82. Lukas Milles: De novo design of intra- and intermolecular autocatalytic isopeptide bonds

MPI of Biochemistry
lukas.milles@gmail.com

Isopeptide bonds are amide bonds formed between amino acids side chains, prominently used as the basis for the covalent SpyTag/Catcher labeling system. They can be formed autocatalytically and are irreversible making them interesting tools as intramolecular bonds to stabilize proteins, or for intermolecular protein-protein crosslinks. Here, we present de novo designed isopeptide bond forming domains (DIPS) generated through a variety of methods. The designs hint at their autocatalytic enzymatic mechanism and design principles of such bonds, but an exact formula for isopeptide generation remains challenging. We can split these DIPS into a peptide tag and small protein catcher that will bind to each other and form a covalent isopeptide bond at their interface. The crosslinking reaction can be controlled by temperature. Finally, we extend DIPS to form large, fully covalently and irreversibly linked protein assemblies.

83. Mads Jeppesen and René Onken: Shape-Directed Protein Assembly Design

Lund University
mads.jeppesen@biochemistry.lu.se
rene.onken@biochemistry.lu.se

Protein assemblies are central to cellular function and provide frameworks for engineering new functionalities. Drawing from the principle of shape complementarity, we developed a shape-directed design method, selecting natural proteins as building blocks based on their precise shape matching within assemblies. This approach is currently applied to designing icosahedral structures inspired by viral capsids and simulated shapes with prespecified features such as porosity and overall geometry.

84. Mads Mørup Nygaard: De novo designed functional peptide binders for novel targets

Gubra
mny@gubra.dk

Therapeutic peptides have received increasing attention from the pharmaceutical industry in recent years. In particular, hormone-derived peptide drugs have been demonstrated as effective treatments for metabolic diseases, including obesity. Peptides are attractive drug modalities since they have high receptor potency and selectivity, minimizing off-target effects, and tuneable circulating half-life. However, the druggable protein space that can be targeted with hormone-derived peptides is limited to fewer than a hundred GPCRs.
To exploit peptides as drug modalities for targets without a native peptide ligand,
we have leveraged a hallucination protocol in combination with AlphaFold to generate peptide binders against different target classes. The pipeline generates and scores peptide binders in silico using a combination of structural metrics from AlphaFold and proteinMPNN.
We validated the protocol by in silico generation of 3000 peptide binders from which the best 192 de novo peptide sequences were synthesized as crudes and characterized for functional receptor modulation, and physiochemical properties. The initial hit rate of the 192 synthesized peptides was 74%, with multiple peptides in the nM potency range. Subsequently, we demonstrated that the de novo peptides could be optimized in a discovery maturation project, leading to a 10- to 100-fold potency increase compared to the initial in silico hit peptide.
We demonstrate that a deep learning-based structural biology approach combined with high-throughput peptide synthesis and functional testing can be used to create a robust foundation for identifying de novo peptide-based hit molecules for drug discovery projects.

85. Marco Orlando: In silico tools to mine glycoside hydrolase enzymes from nature

University of Milano-Bicocca
marco.orlando@unimib.it

Glycoside hydrolases (GHs) are enzymes active in the hydrolysis of poly- and oligo-saccharides. They form one of the greatest classes of enzymes, with the possibility to find, within the same GH family, many different substrate specificities and promiscuity, tunable by few key amino acids substitutions or the addition of substrate-binding loops. The identification of GH highly specialized for a substrate of interest is relevant for new biotechnological applications but may require expanding the “known enzymatic toolbox” in new regions of the sequence space. This will require specificity studies on purified enzymes on the several thousands of new enzymes added each year; as high-throughput cell-free specificity assays may have high technical requirements, do not exist for all enzyme classes and may suffer from enzyme solubility issues, several in-silico sequence-based approaches were developed, but they do not allow predictions that traverse the sequence space of GH families. In this work Blast and end-to-end deep learning (DL) predictors were tested for the task of predicting the substrate specificity of one of the biggest GH family. The results suggest that DL methods perform similarly to Blast top-1 hit approach for GHs and are prone to fail on sequences distant from those included in the training sets. An alternative workflow was proposed to solve those cases by explicitly modelling the enzyme-substrate complex with DL-based docking and couple it with physics-based refinement. To exemplify its effectiveness, it was applied to identify and characterize two evolutionary “orphan” GHs from an Antarctic Pseudomonas strain.

86. Maria Artigues Llexia: Embedding Annotation Transfer (EAT): a contrastive learning implementation to functionally annotate prokaryotic transcription factors for bioengineering

Universitat Pompeu Fabra
maria.artigues@upf.edu

Transcription factor-based biosensors have been key in the creation of genetically encoded biosensors in synthetic biology. Nonetheless, the characterisation of new regulatory proteins in new organisms and their corresponding inducers is still laborious and time-consuming.Classically, new regulatory proteins are identified considering sequence homology in annotation pipelines, but there is insufficient data to pinpoint what inducer they might bind. Recent advances in Artificial Intelligence (AI) models have revolutionised bioinformatics. Here, we are presenting a Contrastive Learning implementation that, combined with embeddings output by protein Language Models (pLMs), have improved the results obtained by homology-based inference when annotating non-model organism Cutibacterium acnes. The results lead us to believe that this method has the potential to annotate which inducer the new transcription factors may bind.

87. Marin Matic: Dissecting structure determinants of GPCR-G protein selectivity via structural bioinformatics

Scuola Normale Superiore Pisa
marin.matic@sns.it

G protein-coupled receptors (GPCRs) regulate cellular signaling by selectively coupling with distinct G-proteins. This study presents a computational analysis of the structural determinants shaping the G-protein-coupling repertoire in both experimentally resolved and predicted 3D GPCR–G-protein complexes. Interface contact analysis highlights key structural features, including transmembrane helices 5 and 6 (TM5 and TM6) and intracellular loops (ICLs), which define coupling specificity. By using these interface contacts as fingerprints, we clustered Gs- and Gi-coupled complexes, suggesting that specific residues drive selective G-protein coupling. We also compared active GPCR complexes to inactive states from PDB structures annotated in GPCRDb, exploring how intramolecular contacts within the transmembrane bundle signal varying G-protein preferences. This analysis highlights the importance of conformational changes in regulating GPCR–G-protein interactions, offering insights for drug design. Experimentally, we validated these findings using the promiscuous receptor CCKAR. Mutations at specificity-determining positions, designed with Rosetta Multi-State Design, shifted G-protein coupling preferences. Notably, Gs-coupled complexes displayed more conserved interfaces, while Gi/o complexes exhibited greater variability in docking poses. Binding energy calculations indicated higher stability for Gs-coupled complexes. AlphaFold2 predictions further supported these structural features and expanded our understanding of less-characterized complexes, including those involving G12/13 proteins.

88. Marius Amann: Enabling Heterologous Expression and Enhancing Stability of Plasmodium falciparum Proteins through ProteinMPNN-Driven Mutagenesis

Albert-Ludwig-University Freiburg
marius.amann@pharmazie.uni-freiburg.de

We used ProteinMPNN to enable expression and enhance stability of three Plasmodium falciparum proteins in E. coli for small-molecule drug discovery. Previously unstable and inexpressible proteins were successfully produced, enabling biophysical studies via ITC and X-ray crystallography. Notably, X-ray crystallography revealed an unexpected conformational change outside the mutated regions, presenting new challenges for understanding protein dynamics.

89. Mark Kriegel: De novo prediction of explicit water molecule positions by a novel algorithm within the protein design software MUMBO

University of Bayreuth
mark.kriegel@uni-bayreuth.de

Despite water plays a major role in protein interaction interfaces, explicit water molecules are usually not considered in protein design. Thus, we implemented the prediction of water molecules in the protein design software MUMBO. To reduce the computational effort we incorporated explicit water through the solvation of rotamer pairs. Our extensive validation demonstrates the potential of the algorithm, e.g. achieving recovery rates of 67% for bridging and up to 86% for fully coordinated water.

90. Marvelous Chikerema: Towards a cDNA display platform for the directed evolution of membrane proteins

VIB/VUB Center for Structural Biology
Marvelous.Chikerema@vub.be

Membrane proteins constitute 30% of natural proteins, our understanding of their sequence determinants is limited. I aim to characterize effects of mutations on folding Transmembrane β-barrel proteins (TMBs). Through directed evolution of TMBs by cell-free, high-throughput cDNA approach. This will result in protein sequences optimized for folding, leading to a better understanding of thermodynamic principles governing fold stability and a reference assay for future engineering membrane proteins.

91. Matej Milijaš-Jotić: Computational Design of a Bivalent Cytokine Receptors Antagonist

Max Planck Institute for Biology
matej.jotic@tuebingen.mpg.de

Targeting multiple receptors with a single therapeutic agent can enhance treatment efficacy and overcome the limitations of monotherapy. We aimed to develop single-domain, bivalent blockers for IL-1RI and G-CSFR, two cytokine receptors implicated in several inflammatory and immune-related disorders. By combining receptor-binding fragments, we created single-domain proteins capable of forming stable receptor complexes. The designs were optimized using the Damietta protein design software for enhanced binding affinity and solubility, while molecular dynamics was employed to select the candidates for experimental testing. Preliminary results show that the constructs are soluble, thermostable, and bind both receptors with high affinity, offering a promising therapeutic strategy.

92. Matteo Tiberti: MAVISp: Multi-layered Assessment of Varlants by Structure for proteins

Danish Cancer Institute
tiberti@cancer.dk

The accumulation of genomic alterations over time, especially in tumor suppressor genes or oncogenes, is the prime cause of cancer. Such alterations influence disease development and outcome, risk of relapse and response to treatment. While the number of known disease-associated mutations is growing at an astonishing pace, the effect of many of them is currently unknown or uncertain (Variants of Unknown Significance, VUS). We introduce MAVISp (Multi-layered Assessment of Varlants by Structure for proteins), a modular framework for the prediction of the impact of variants affecting protein-coding genes and their corresponding protein products, based on the analysis of protein structures or protein structural ensembles. Our approach is based on high-throughput calculations of changes in folding free energy, biomolecular simulations of different types and other methods. It focuses on delivering a mechanistic interpretation of the effect of each mutation. In this contribution, we highlight the role of Rosetta and Rosetta-based tools in our framework. We use the Rosetta cartesian_ddg and flex ddg protocols for the prediction of changes of free energy of folding or binding upon mutation, respectively, and for protein structure prediction using RoseTTAFold All-Atom. We plan on further improving MAVISp acting on features including those to some extent depending on Rosetta, such as improving our protein modeling protocols, prediction of the effect of indel mutations, support for predicting changes of stability upon mutations including a cell membrane environment. MAVISp includes data for more than 200 different proteins and tens of thousands of variants. We plan on expanding our dataset and variants will be deposited on a regular basis. We welcome displays of interest for new protein targets.

93. Matthew Ellis: De novo design of synthetic dye binders

University of Portsmouth
up2011220@myport.ac.uk

Over billions of years, natural photosystems have evolved as highly efficient apparatus for absorbing and transferring light energy – possessing near-perfect quantum yields. However, despite our fundamental understanding of the underlying processes, comparable efficiencies have yet to be achieved within a de novo system. This knowledge gap presents a potent opportunity to develop synthetic binders that outperform natural proteins for various biocatalytic applications. To address this, we utilise cutting-edge de novo design tools, RFDiffusionAA and LigandMPNN, to create simplified protein binders for synthetic dyes. We anticipate that spectroscopic analysis of these designs will reveal strategies to fine-tune the spectral properties of the dyes.

94. Maximilian Salomon: GPCRFlow: GPCR Ensemble Prediction of Metastable States

Leipzig University
maximilian.salomon@uni-leipzig.de

MD simulations of GPCR complexes are computational expensive. To study GPCR - G-protein coupling, we aim to predict conformational ensembles, focusing on meta-stable states along the coupling pathway. These states will guide MD simulations to explore the conformational landscape. We plan to refine AlphaFlow (Jing et al., 2024) using microsecond-long GPCR MD simulations. Additionally, we aim to replace AlphaFold2 with a model better suited for multimeric predictions, like AlphaFold-Multimer.

95. Michael Westberg: Real-time control of endogenous proteins by photoswitchable binders

Aarhus University
westberg@chem.au.dk

We report the design of single-chain photoswitchable DARPins (psDARPins) for bidirectional optical control of endogenous proteins. We computationally redesigned the DARPin scaffold so that attached photodissociable protein domains occlude target recognition in the dark but permit binding upon illumination. The redesign leaves the DARPin paratope intact to allow modular paratope exchange with existing or new DARPins. We validate the approach by temporally controlling endogenous kinase activity.

96. Miguel Atienza Juanatey: Towards meta-stable protein assemblies

LMU-München
miguelatienzajuanatey@gmail.com

Self-assembling polyhedral protein assemblies are great platforms for engineering the targeted encapsulation and transport of biomolecules. Current successful de novo protein assemblies are extremely stable, limiting their application as cargo delivery vehicles. Designing such functional protein nanoparticles will require understanding and controlling not only structure, but also their stability and assembly dynamics. I will present our approach for the design of meta-stable protein assemblies.

97. Mikel Martinez Goikoetxea: Old folds can learn new tricks: Alphafold-driven insights on coiled-coil structure

Max Planck Institute for Biology
mikel.martinez@tuebingen.mpg.de

Coiled coils are a widespread protein structure motif, found throughout the proteome of life. Their short sequence repeats give rise to considerable structural diversity, which has facilitated the development of numerous prediction and modeling programs. In this study, we demonstrate the application of AlphaFold for both the discovery and accurate modeling of coiled-coil structures, highlighting its potential to deepen our understanding of this versatile motif.

98. Moritz Ertelt: Combining machine learning with structure-based protein design to predict and engineer post-translational modifications of proteins

University of Leipzig
moritz.ertelt@gmail.com

Post-translational modifications (PTMs) of proteins play a vital role in their function and stability. These modifications influence protein folding, signaling, protein-protein interactions, enzyme activity, binding affinity, aggregation, degradation, and much more. To date, over 400 types of PTMs have been described, representing chemical diversity well beyond the genetically encoded amino acids. Such modifications pose a challenge to the successful design of proteins, but also represent a major opportunity to diversify the protein engineering toolbox. To this end, we first trained artificial neural networks (ANNs) to predict eighteen of the most abundant PTMs, including protein glycosylation, phosphorylation, methylation, and deamidation. In a second step, these models were implemented inside the computational protein modeling suite Rosetta, which allows flexible combination with existing protocols to model the modified sites and understand their impact on protein stability as well as function. Lastly, we developed a new design protocol that either maximizes or minimizes the predicted probability of a particular site being modified. We find that this combination of ANN prediction and structure-based design can enable the modification of existing, as well as the introduction of novel, PTMs. The potential applications of our work include, but are not limited to, glycan masking of epitopes, strengthening protein-protein interactions through phosphorylation, as well as protecting proteins from deamidation liabilities. These applications are especially important for the design of new protein therapeutics where PTMs can drastically change the therapeutic properties of a protein. Our work adds novel tools to Rosetta's protein engineering toolbox that allow for the rational design of PTMs.

99. Muhammad Yasir Ateeque: Modeling the Structural Dynamics of the DnaB-DNA Complex using AI Structure Prediction Tools and Solid-State NMR Data

Institute for Drug Discovery, Medical Faculty, University of Leipzig
muhammad.ateeque@medizin.uni-leipzig.de

H. pylori DnaB is a 335kDa large, hexameric helicase which catalyzes the ATP-driven unwinding of double stranded DNA into single stranded DNA. The structural dynamics of DnaB, particularly upon DNA interaction and the switch between open and closed conformations, remain challenging to decipher. While solid-state NMR data on H. pylori DnaB indicate a closed ring conformation, different experimental datasets from homologous DnaB proteins suggest that a transition to a more open, staircase-like conformation is possible as well. However, the intricate structural details and their relation to the function of DnaB have yet to be fully uncovered. In this study, we used different AI structure prediction tools and experimental NMR restraints to predict the structural ensemble of the DnaB-DNA complex with unprecedented accuracy. We chose a conformational selection approach, in which a distribution of realistic structural models was first generated, followed by the selection of a representative set of conformers that best agreed with the NMR data. We began by modeling the hexameric complex, including single-stranded DNA, ADP molecules, and Mg^{2+} ions, using AlphaFold2, RosettaFold2 Nucleic Acid and RosettaFold All-atom techniques. To delve deeper into the open and closed conformations of DnaB, we utilized AlphaFold2-Multimer to sample a distribution of possible conformations and analyze their changes with reference to homologous DnaB structures. Atomic distances from the predicted structures were compared with NMR-derived restraints to identify the set of models that best satisfied the experimental data. The selected structural ensemble will be studied using molecular dynamics simulations, offering deeper insights into the dynamic behavior of the DnaB-DNA complex. This integrative approach not only advances our understanding of DnaB's conformational states and functional mechanisms in DNA replication but also opens new avenues for targeting this crucial protein in therapeutic applications.

100. Nicolas Deutschmann: Kickstarting ML-guided antibody optimization without affinity data

Cradle
nicolas@cradle.bio

We explore to what extent ML-driven antibody optimization cycles can be initiated in the absence of affinity data or relevant diversity. Unlike many protein classes, the evolutionary context summarized in protein language models is unlikely to be directly relevant to complementarity-determining regions matured through somatic hypermutation. Despite these expectations, low-signal data such as phage display from immunized organisms can be leveraged to generate function-preserving diversity, both in CDRs and in the framework region of an anti SARS-Cov-2 VHH, which we demonstrate with in-vitro validation of ML-designed variants. This opens an exciting path for active-learning-driven antibody optimization starting from an immunization campaign or even a single lead.

101. Nina Schuch: The Impact of Epistasis and Covariation on Enzyme Stability and Activity through Ancestral Reconstruction.

Biochemistry, Lund University
nina.schuch@biochemistry.lu.se

We investigate the role of epistasis and covariation in protein evolution by testing all possible mutations at selected sites in an enzyme across multiple homologs. Ancient sequence reconstruction and comparative analysis of covariant and non-covariant sites will reveal how these forces shape stability and activity over time. Rosetta's modeling tools will be crucial for predicting structural impacts, offering insights into the evolution of protein sequence landscapes

102. Palina Pliushcheuskaya: Evaluation of binding site prediction methods for ligands targeting the protein-membrane interface region

Leipzig University
palina.pliushcheuskaya@uni-leipzig.de

Increasing structural and biophysical evidence shows that many drugs bind to the protein-membrane interface region in membrane protein structures. To assess correctly pharmacological profile of a potential drug, it is important to consider the influence of the membrane, which represents a special chemical and physical environment, impacting ligand distribution and binding. In this study, we tested the performance of available computational methods for ligand binding site prediction in the protein transmembrane interface region. We compiled non-redundant protein datasets containing GPCRs and ion channels, and compared method performance relative to a soluble protein dataset obtained from PDBBind. We tested state-of-the-art geometry-based (Fpocket, ConCavity), energy probe-based (FTSite), machine learning-based (P2Rank, GraSP) and deep learning-based (PUResNet, DeepPocket, PUResNetV2.0) methods and evaluated them using the center-to-center distance (DCC) and discretized volume overlap (DVO) between the predicted binding site and the actual ligand position. The three best-ranking methods based on success rates on GPCRs were DeepPocket, PUReNetV2.0, and Concavity, and for ion channels these were DeepPocket, PUResNetV2.0, and FTSite. However, average DCC and DVO values were lower for all methods compared to the soluble protein dataset for which DVO and normalized DCC values ranked between 0.96 and 0.29 in their best case. In conclusion, this study provides an overview of the performance of the state-of-the-art binding site prediction methods on their ability to identify pockets at the protein-membrane interface. It also underscores the need for further method development in prediction of the protein-membrane ligand binding sites.

103. Parth Bibekar: RISoTTo: Context-aware geometric deep learning for RNA sequence design

EPFL
parth.bibekar@epfl.ch

RNA design is essential in synthetic biology, but most methods focus on secondary structures, neglecting 3D structure and biological context. To address this, we introduce RISoTTo, a context-aware RNA design method inspired by the success of CARBonAra, which considers the environment in protein design. In-silico validation of RISoTTo shows promising results, suggesting its potential for further experimental validation in complex biological environments.

104. Pasquale Miglionico: Predicting Stable Protein Complexes with Co-evolution and Co-expression Data

Scuola Normale Superiore
pasquale.miglionico@sns.it

In this study, we employed a data-driven approach to explore the evolutionary and genetic determinants of stable complex formation in the human proteome. We found that simple co-evolutionary and co-expression metrics are highly informative of stable complex formation. We used this information to train supervised binary classifiers to predict interactions directly involved in forming a stable complex (as annotated in Complex Portal). Our approach outperformed the STRING score, which we employed as a baseline. From these pairwise predictions, we generated a human proteome-wide network that we clustered to assess the recovery of known complexes from Complex Portal within network communities. We could recover known complexes with higher accuracy than clustering the STRING networks. Furthermore, we employ Alphafold-Multimer to structurally model small, uncharacterized clusters, generating new candidate complexes for experimental validation. This method enables the refined stratification of molecular interaction networks and facilitates the discovery of novel functional complexes

105. Paul Eisenhuth: Flexible E(3) equivariant interaction kernels for biomolecular affinity prediction

ScaDS.AI, Meiler Lab, Leipzig University
eisenhuth@cs.uni-leipzig.de

Biomolecular interaction are often decomposed into frequently observed geometric formations like hydrogen bonds or pi-pi stacks. These are useful to explain and describe differences in binding affinities, but are pre defined following strict physics based criteria. We propose a flexible kernel architecture which can freely learn observable interaction patterns and are thereby useful building blocks for larger machine learning models. These kernels adapt to geometries observed during training.

106. Paul Kluge: Parametrical design of surface binding helical bundles for the elucidation of the CISS-effect

Leipzig University, Medical Faculty
pk12jizo@studserv.uni-leipzig.de

This project focuses on the design of helical proteins that bind to a gold surface. The aim is to create helical bundles with varying parameters and to test the influence of these parameters on the CISS effect. The CISS effect (chirality-induced spin selectivity) refers to the phenomenon where chiral molecules preferentially interact with electrons of a specific spin orientation, leading to spin-polarized electron transport through the chiral structure.

107. Pauline Hermans: Exploring evolution to enhance mutational stability prediction

Université Libre de Bruxelles
pauline.hermans@ulb.be

Since protein stability is a major driver of evolution, evolutionary data are often used to guide stability predictions. Many state-of-the-art stability predictors extract evolutionary information from multiple sequence alignments (MSA) of proteins homologous to a query protein, and leverage it to predict the effects of mutations on protein stability.
To evaluate the power of such methods and their limitations, we used the massive amount of stability data recently obtained by deep mutational scanning to study how best to construct MSAs and optimally extract evolutionary information from them. The parameters considered include the protein sequence dataset used for the homologous search, as well as MSA depth, E-value and curation criteria.
We also tested different evolutionary models and unexpectedly found that independent-site models achieve the similar accuracy as more complex epistatic models.
Interestingly, by combining any of the evolutionary features with a simple structural feature, the relative solvent accessibility of the mutated residue, we obtained similar prediction accuracy of supervised, machine learning-based, protein stability change predictors.
Our results provide new insights into the relationship between protein evolution and stability, and show how evolutionary information can be exploited to improve the performance of mutational stability prediction.

108. Piotr Smieja: deepBBQ: a deep learning approach to the protein backbone reconstruction

University of Warsaw
p.smieja2@uw.edu.pl

We propose a new machine learning model, deepBBQ, capable of reconstructing the full protein backbone from its Cα coordinates alone. Our approach utilizes a convolutional neural network to predict a single internal coordinate for each residue, relying on geometric patterns derived from the Cα trace. Assuming the planarity of peptide plates, these predictions are converted into Cartesian coordinates. Comparison against existing methods reveals deepBBQ's high accuracy and computational efficiency.

109. Qihao Zhang: A Heme Location Prediction Tool with 3dCNN/Docking Procedure

EPFL
qihao.zhang@epfl.ch

Heme-containing proteins are involved in many important metabolic processes like oxygen transport in blood or gas sensing. However, most methods for heme prediction either only predict binding residues or require the number of heme as the input. In this work, we develop a 3dCNN-based method for blind heme prediction given a single structure as input and heme binding density as output, and we explore multiple approaches like rigid docking and molecular dynamics flexible fitting to dock the heme.

110. Rajashi Sinharoy: Studying the Association of Blood Plasma Proteins on Lipid Nanoparticles with Brownian Dynamics Simulations

Universität Leipzig
sinharoy.rajarshi@uni-leipzig.de

mRNA vaccines have revolutionized the field of vaccination, offering a novel and highly effective way to prevent infectious diseases, as demonstrated during the COVID-19 pandemic. The development of these vaccines has been significantly supported by advancements in lipid nanoparticle (LNP) technology, which has proven to be a crucial delivery system for nucleic acids, including mRNA. LNPs, upon intravenous injection, undergo a series of complex biological interactions, including the detachment of PEG chains from the LNP our lipid shell and the subsequent adsorption of proteins and lipoproteins on the LNP surface, leading to the formation of a protein corona. This protein corona is supposed to play a critical role in determining the biodistribution, cellular uptake, and overall efficacy of the LNPs. However, there is still a substantial knowledge gap in our understanding of the specific interactions between LNPs and blood plasma proteins, particularly concerning how these interactions vary with different LNP lipid compositions, surface charges and PEG chains lengths.
To address this challenge, we developed a computational framework leveraging Brownian Dynamics (BD) simulations to investigate the diffusional association processes between plasma proteins and LNPs. The BD system contains over 1000 particles of the most abundant blood plasma proteins to mimic a composition that closely matches physiological conditions. This approach allows for a detailed examination of how variations in the LNP composition, lipid charge and PEG chain length will influence the interaction landscape with plasma proteins.

111. Renzo Condori: Machine Learning-Guided In Silico Design of EGFR Binders: A Promising Alternative for Treating Triple-Negative Breast CancerSimulations

University of Ghent
RenzoAntonio.CondoriTolentino@UGent.be

Triple-negative breast cancer (TNBC) is an aggressive subtype of breast cancer with limited treatment options, as it lacks targeted therapies due to the absence of hormone receptors and HER2 expression. Recent advances in machine learning (ML) offer a promising avenue for the design of epidermal growth factor receptor (EGFR) binders, which have the potential to inhibit EGFR signaling pathways. Our aim is to design new ligands with high affinity for EGFR through machine learning tools, opening new pathways to improve drug discovery efficiency.

112. Robert Jefferson: Computational design of dynamic receptor—peptide signaling complexes applied to chemotaxis

King's College London
rob.jefferson@kcl.ac.uk

Engineering protein biosensors that sensitively respond to specific biomolecules by triggering precise cellular responses is a major goal of diagnostics and synthetic cell biology. Previous biosensor designs have largely relied on binding structurally well-defined molecules. In contrast, approaches that couple the sensing of flexible compounds to intended cellular responses would greatly expand potential biosensor applications. Here, to address these challenges, we develop a computational strategy for designing signaling complexes between conformationally dynamic proteins and peptides. To demonstrate the power of the approach, we create ultrasensitive chemotactic receptor—peptide pairs capable of eliciting potent signaling responses and strong chemotaxis in primary human T cells. Unlike traditional approaches that engineer static binding complexes, our dynamic structure design strategy optimizes contacts with multiple binding and allosteric sites accessible through dynamic conformational ensembles to achieve strongly enhanced signaling efficacy and potency. Our study suggests that a conformationally adaptable binding interface coupled to a robust allosteric transmission region is a key evolutionary determinant of peptidergic GPCR signaling systems. The approach lays a foundation for designing peptide-sensing receptors and signaling peptide ligands for basic and therapeutic applications.

113. Rodrigo Ochoa: Open-source tools for peptide analysis and design: PepFuNN and mPARCE

Novo Nordisk
raoc@novonordisk.com

Computational methods are required to study peptides and gain insights into rational design strategies. Open-source frameworks can accelerate the analyses of complex peptides, reducing human error during the assembling of molecules. Some open-source protocols for peptide research will be presented. PepFuNN is a Novo Nordisk package to run cheminformatics analysis for peptides. mPARCE is a evolutionary protocol to design peptide sequences using Rosetta and external scoring functions.

.

114. Sajith Kolathuparambil: Design of noval metalloproteases

Institute of Biochemistry, Graz University of Technology
sajith.kolathuparambil@tugraz.at

Efficient recombinant protein purification is pivotal in biological research, predominantly relying on affinity tags. However, the complete removal of C-terminal tags is problematic since most endoproteases cut toward the C-terminus of the recognition sequence. To this end, we introduce a new approach that harnesses the power of de novo design to develop new proteases. Our preliminary results show a gradual improvement in the designed models, which signifies the efficiency of our pipeline.

115. Samuel Curtis: Adapting DNA synthesis screening to a de novo world

Rosetta Commons / OMSF
samuelmcurtis@gmail.com

Many DNA synthesis providers routinely use screening algorithms to detect and investigate orders containing regulated pathogens and other potentially dangerous sequences. However, synthetic DNA providers and public health officials have expressed concerns that existing screening systems are increasingly ill-equipped for state-of-the-art generative models, heavily modified sequences, and de novo molecules. In this discussion, we'll explore the regulatory architecture surrounding DNA synthesis, the screening systems that exist today, and how the protein design community can contribute to the development of next-generation screening approaches.
Sebastian Lindner

116. Sarah Badawy Ahmed Ossman: Coarse-Grained Umbrella Sampling Simulations for Binding Affinity Prediction of SARS-CoV-2 spike-Nanobody Complexes

Research Scientist 2
sarah@proteinea.com

TThe accurate prediction of binding affinities between engineered nanobodies and SARS-CoV-2 spike protein is critical for the development of effective therapeutics. We present a pipeline leveraging coarse-grained umbrella sampling (CG-US) simulations, utilizing the SIRAH force field, which offers a high-resolution residue-based coarse-graining approach. This method strikes a balance between computational accuracy and speed, efficiently capturing key thermodynamic parameters while reducing computational costs. Our approach provides a scalable solution for screening nanobody candidates, facilitating rapid in-silico assessments.

.

117. Sebastian Lindner: Expanding the natural chorismate mutase verse through AI-guided design and wet lab evolution

University Heidelberg/ ETH Zurich
sebastian.lindner@stud.uni-heidelberg.de

Chorismate mutase evolved in two distinct scaffolds. Although multiple endeavors tried to develop new scaffolds, no de novo design with adequate catalytic efficiency exists. By deploying RFdiffusion and LigandMPNN with various downstream in-silico validation techniques like ChemNet, we designed new variants in collaboration with the Baker lab that yield promising in-silico results. Selected designs are subjected to rigorous in vivo selection and in vitro characterization with subsequent rounds of directed evolution.

118. Sevilay Gülesen: Prediction of Different Conformational States of Class I Fusion Protein

Leipzig University
sevilay.guelesen@medizin.uni-leipzig.de

Viruses, such as MARV, use their surface glycoprotein (GP) to perform the essential process of attaching to a host cell and fusing the viral membrane with it. Epitopes on the viral GP are the target of recognition of the humoral immune response, specifically antibodies. During the fusion processes, the GP undergoes conformational changes because of difference of pH. Structures through these states are mainly named as pre-fusion structure, middle-state structure and post-fusion structure. Nevertheless, ratio of solved post-fusion structures is less than the pre-fusion in RSCB-PDB, especially for class I fusion proteins. Structures of post-fusion states are important to predict possible epitope regions by computational methods. Thanks to recent development in computationally structural prediction of protein enables to predict structures much easier with high accuracy. In this project, we observed that prediction of different conformations of GPs is possible with AlphaFold2-multimer (AF2-M) but the overall prediction ratio of post-fusion states is very low. Therefore, we hypothesized that we need to adapt the AF2-M sampling to enrich for these desired conformations. The major two aims of this project are to manipulate AF2-M to predict different conformations of a given GP and benchmark the AxIEM (Accelerated class I fusion protein Epitope Mapping) protocol using these predicted conformations. AF2-M requires a sequence as input and uses internally either 1) a multiple sequence alignment, 2) just it's trained weights or 3) template information. In this project, we will test which of these parameters need to be adapted to enrich for pre-or post-fusion conformations.

119. Shengxiao Qi: Cloning and Expression of ASMT for the Identification of Pharmacological Probes

Department of Drug Design and Pharmacology, University of Copenhagen
nrz989@sund.ku.dk

Melatonin (MT) is the key hormone of the circadian clock. The acetylserotonin methyltransferase (ASMT) plays a key role in the last step of its synthesis in all mammals. Recent evidence suggests that malfunction of circadian clock is closely related to many neurological and psychiatric disorders such as depressive disorders, schizophrenia and sleep disorders. Consequently, there is an increasing interest in understanding how the disruption of the circadian clock affects health states and results in diseases, which remain largely enigmatic. For such studies the availability of pharmacological probes for further studies is essential. Here, we cloned human and rat ASMT (hASMT and rASMT) (full length and a truncation corresponding to the human enzyme) by in-Fusion cloning4 and expressed them in three expression system including E. coli, insect cells and HEK 293F cells. Currently, we have already gotten pure hASMT and found that two of newly designed melatonin analogs have inhibition effect through HPLC-fluorescence-based enzyme assay.

120. Shuhao Zhang: Deep-Learning-Designed Nanobodies Targeting AT1R Show Enhanced Binding Affinity and Antagonistic Effects

EPFL
shuhao.zhang@epfl.ch

GPCRs are crucial regulators of cellular activity and represent a major class of therapeutic drug targets. In this study, we used deep-learning-based protein design methods and MD simulations to design novel nanobodies targeting the angiotensin receptor. The designed nanobodies exhibited varying levels of antagonistic effects and achieved increasing binding affinity compared to a nanobody engineered from a yeast surface display library, offering potential new tools for modulating GPCR activity.

121. Simon Duerr: Predicting metal-protein interactions using cofolding methods: Status quo

EPFL
simon.duerr@epfl.ch

Metals play important roles for enzyme function and many therapeutically relevant proteins. Despite the fact that the first drugs developed via computer aided drug design were metalloprotein inhibitors, many computational pipelines for drug discovery still discard metalloproteins due to the difficulties of modelling them computationally. New “cofolding” methods such as AlphaFold3 (AF3) (Abramson et al., 2024) and RoseTTAfold-AllAtom (RFAA) (Krishna et al., 2024) promise to improve this issue by being able to dock small molecules in presence of multiple complex cofactors including metals or covalent modifications. In this talk I want to provide an overview of how these methods perform for metal ion prediction. The talk will cover generalization to novel metal binding sites, sensitivity to mutations, fold-switching metalloproteins and stoichiometry of predictions.

122. Sören von Bülow: Prediction of phase separation propensities of disordered proteins from sequence

Structural Biology and NMR Laboratory, Linderstrøm-Lang Centre for Protein Science, Department of Biology, University of Copenhagen
soren.bulow@bio.ku.dk

Phase separation (PS) is one possible mechanism governing the selective cellular enrichment of biomolecular constituents for processes such as transcriptional activation, mRNA regulation, and immune signalling. PS is mediated by multivalent interactions of biological macromolecules including intrinsically disordered proteins and regions (IDRs). Despite considerable advances in experiments, theory and simulations, the prediction of the thermodynamics of IDR phase behaviour remains challenging.

123. Teemu Kalle Eemeli Rönkkö: Structure-based predictions as a novel means to pair cone snail-derived peptide toxins with G-protein coupled receptors

University of Copenhagen
teemu.ronkko@sund.ku.dk

In this study, we developed a computational workflow to identify novel interactions between conotoxins and human G protein-coupled receptors (GPCRs). We retrieved 118 conotoxin sequences from ConoServer and selected 137 class A and B1 human GPCRs. Using AlphaFold 2 (AF2), we modelled over 16,000 GPCR–conotoxin complexes. Conserved GPCR residue positions from GPCRdb were employed to filter out predictions in which the conotoxin was not interacting with the orthosteric binding pocket. We applied the Local Interaction Score (LIS) to identify high-confidence GPCR-conotoxin pairs, validating the workflow with 29 known interacting pairs. Our results yielded potential novel pairings of conotoxins with human GPCRs, which will undergo experimental validation. Future work includes integrating DeepPeptide in the workflow for mature conotoxin prediction from venom gland transcripts. This method offers a robust framework for the efficient screening of GPCR–peptide interactions, enhancing drug discovery and therapeutic applications of conotoxins.

124. Thomas Lavstsen: Harnessing Novel Broadly Neutralizing mAbs Against P. falciparum Virulence Proteins for Vaccine Design

University of Copenhagen
thomasl@sund.ku.dk

Malaria remains a significant global health burden, primarily due to the virulence of P. falciparum parasites, which sequester infected erythrocytes to endothelial cells via specific variants of their PfEMP1 adhesion proteins that bind to the human endothelial protein C receptor (EPCR). Immunity to malaria is mediated by antibodies acquired through exposure that target and inhibit PfEMP1 from binding to EPCR. In recent work, we identified the first two broadly neutralizing antibodies (bNAbs) that specifically target the EPCR-binding domains of PfEMP1. We are now seeking to exploit the structural resolution insights from these mAbs' interactions with multiple PfEMP1 variants for the design of immunogens and antibody therapeutics. By focusing on conserved epitopes across the PfEMP1 family, this approach has the potential to significantly enhance immunity to malaria and provide broad protection

125. Till El Harrar: Decoding the Dual Role of Aqueous Ionic Liquids in Protein Stability: Insights from Energy Profiling and Perturbation Pathways

Forschungszentrum Jülich
t.el.harrar@fz-juelich.de

Ionic Liquids (ILs) and aqueous ILs (aILs) are attractive solvents for biocatalysis due to their unique properties, but enzyme activity often decreases when incubated in these solvents due to various - often poorly understood - molecular mechanisms. Our recent study revealed “Perturbation Pathways” as a molecular mechanism of one of these effects, where local structural perturbations that originate from energetically favorable aIL interactions with enzyme surface residues propagate through the enzyme, potentially destabilizing distant regions such as the catalytic site. However, the thermodynamic basis of these changes at the origin-residues was unclear.
In this study, we systematically analyzed changes in the free energy profiles of 50 relevant pairwise protein residue interactions using extensive Potential of Mean Force (PMF) computations in aILs, salt solutions, and water, in some cases in multiple conformations. As principal results, we showed that aILs and salt solutions can induce substantial solvent-, concentration-, conformation-, and system-specific effects on intramolecular interactions between protein residues. Surprisingly, these effects can be of destabilizing or stabilizing nature by acting as competitors for interaction partners or solvent bridge-like stabilizers, respectively.
This comprehensive dataset of PMF profiles establishes the thermodynamic basis for linking specific aIL-residue interactions to the structural changes observed in Perturbation Pathways. Moreover, this data can be used to enhance the predictive accuracy of computational tools that rely on internal force fields for predicting changes in protein stability, such as FoldX or our in-house tool CNA.

126. Timothy Patrick Jenkins: Crack the Code, Craft the Cure: How Machine Learning is Redefining Target Discovery and Therapeutic Innovation

DTU
tpaje@dtu.dk

Machine learning (ML) is ushering in a new era in target identification and therapeutic discovery, with profound implications across various scientific domains. In the realm of proteomics, we have leveraged the power of ML to develop a leading-edge deep learning model called InstaNovo. This model enables high-precision de novo peptide sequencing, eliminating many of the constraints of conventional methods and opening up new opportunities in antibody sequencing, identification of neo-epitopes in cancer, and the exploration of the dark proteome. However, beyond target identification, ML is also proving promising in the rapid discovery and development of therapeutics. Particularly with the rise of generative de novo protein design, design of functional binders entirely in silico has been brought within reach. We have taken advantage of these developments, to design minibinders (small binding proteins primarily comprised of beta sheets and alpha helices) that can neutralise snake venom toxins, industrial enzymes, and immunotherapy targets. These findings hold great promise for the rapid development of next-generation therapeutics and bio-industrials.

127. Tobias Wörtwein: A web-based toolkit for protein design

Max Planck Institute for Biology, University Clinic Tübingen
tobias.woertwein@tuebingen.mpg.de

We have recently developed a fast, physics-based approach for protein design, which leverages tensorized energy calculations. Here, we present a toolkit of different applications using our tensorized design engine as well as third-party tools, such as OpenMM and ProteinMPNN. Our toolkit seamlessly integrates different applications, allowing interoperability and pipelining. The toolkit is implemented as a web server that enables structure-centric manipulations through an intuitive graphical interface; available at damietta.de. Our future work involves improving the performance of this platform and expanding the available tools set.

128. Victor Klein-Sousa: Towards a complete phage tail fiber structure atlas

University of Copenhagen
victor.klein@cpr.ku.dk

Bacteriophages use receptor-binding proteins (RBPs) to adhere to bacterial hosts. Understanding the structure of these RBPs can provide insights into their target interactions. Tail fibers, a prominent type of RBP, are typically elongated, flexible, and trimeric proteins, making it challenging to obtain high-resolution experimental data of their full-length structures. Recent advancements in deep learning-based protein structure prediction, such as AlphaFold2-multimer (AF2M) and ESMfold, now allow for the generation of high-confidence predicted models of complete tail fibers. In this paper, we introduce RBPseg, a method that combines monomeric ESMfold predictions with a novel sigmoid distance pair (sDp) protein segmentation technique. This method segments the tail fiber sequences into smaller fractions, preserving domain boundaries. These segments are then predicted in parallel using AF2M and assembled into a full fiber model. We demonstrate that RBPseg significantly improves AF2M v2.3.1 in terms of model confidence, running time, and memory usage. To validate our approach, we used single-particle cryo-electron microscopy to analyze five tail fibers, from three phages of the BASEL collection. Additionally, we conducted a structural classification of 66 fibers and their domains, which identified 16 well-defined tail fiber classes and 89 domains. Our findings suggest the existance of modular fibers as well as fibers with different sequences and sharedstructure, indicating possible sequence convergence, divergence, and domain swapping. We further demonstrate that these structural classes account for at least 27% of the known tail fiber universe.

129. Yasser Almeida: PPI-Affinity: A Tool for the Prediction and Optimization of Protein–Peptide and Protein–Protein Binding Affinity

Technical University of Dortmund
yasser.almeida@tu-dortmund.de

We present PPI-Affinity, a support vector machine (SVM) tool for the prediction of protein-protein and protein-peptide binding affinities. The tool allows ranking, screening and designing peptides and proteins with optimized binding affinity. The performance of the SVM models was assessed on four benchmark datasets, which include protein-protein and protein-peptide data. Furthermore, the models were evaluated on a set of mutants of EPI-X4, an endogenous peptide inhibitor of the chemokine receptor CXCR4 as well as on complexes of the serine proteases HTRA1 and HTRA3 with peptide inhibitors.

130. Yuval Fishman: Exploring the Functional Diversity of Natural Enzyme Families Using Protein Design

Weizmann Institute of Science
yuval.fishman@weizmann.ac.il

Enzymes are vital in biological and industrial processes, and enzyme discovery is critical for the green transition in the chemical industry. However, most natural enzymes exhibit low or no expression in lab strains, limiting discovery. We present a workflow combining AlphaFold2 modeling, structural bioinformatics, and PROSS stability design to generate homologous enzymes with diverse activity profiles. This scalable approach expands enzyme functionalities for research and biocatalyst discovery.

131. Yuyang Zhang: TopoDiff: Improving diffusion-based protein backbone generation with global-geometry-aware latent encoding

Tsinghua University
zhangyuy21@mails.tsinghua.edu.cn

TopoDiff is a novel diffusion-based framework bridging the representation learning and generative modeling of protein structures. TopoDiff significantly improves fold coverage of generated backbones and opens up numerous possibilities for a more controllable generation process. We validate TopoDiff with experimental characterization of novel mainly-beta proteins, which are among the first successful designs of such proteins without predefined topology blueprint.