European RosettaCon 2025

Oct 20-22, 2025

Ljubljana, Slovenia

Book of Abstracts

Table of Contents

European RosettaCon 2025 Abstract Book

Editors:

Ajasja Ljubetič, Dominik Gront

Organising Committee:

Ajasja Ljubetič, Dominik Gront

Local Organizing Committee:

Federico Olivieri
Eva Rajh
Liza Ulčakar
Alina Konstantinova
Katja Prek
Piotr Śmieja
Žan Žnidar, Teo Stupar, Nej Bizjak

Issued by:

Slovenian Biochemical Society

Ljubljana, 2025

Price: Free
URL: https://europeanrosettacon.org/eu_rosetta_con_2025_abstract_book/
Kataložni zapis o publikaciji (CIP) pripravili v Narodni in univerzitetni knjižnici v Ljubljani
COBISS.SI-ID 256647939
ISBN 978-961-95941-3-1 (PDF)

OUR SPONSORS

Platinum Level

NOVARTIS | Reimagining Medicine

https://www.novartis.com/si-en/

Gold Level

Fidabio

https://www.fidabio.com/

inito

https://www.inito.com/

Organizational Sponsors

Rosetta Commons

https://rosettacommons.org/

instruct ERIC Slovenia

https://instruct-eric.si/

SLOVENIAN BIOCHEMICAL SOCIETY

https://sbd.si/

ROSETTA DESIGN GROUP

https://rosettadesigngroup.com/

	Monday 20th	Tuesday 21st	Wednesday 22nd
08:15	**Registration**		
08:20			
08:25			
08:30		**Registration**	**Registration**
08:35			
08:40			
08:45			
08:50			
08:55			
09:00	Welcome	**Roman Jerala** (Sythetic biology) chair: Dominik Gront	**Tanja Kortemme** (Multistate proteins) chair: Enrique Marcos
09:05			
09:10	**Birte Höcker** (Molecular machines) chair: Ajasja Lubetic		
09:15			
09:20			
09:25			
09:30			
09:35			
09:40			
09:45		Ben Hardy Design and Engineering of Artificial Membrane Cytochromes for Synthetic Bioenergetics	Adam Broerman Design of facilitated dissociation enables timing of cytokine signaling
09:50			
09:55	Liza Ulčakar		
10:00	De novo designed random walker	Miguel Atienza Juanatey Towards meta-stable de novo protein assemblies	Eva Rajh Toehold mediated strand displacement as a tool for kinetic control of protein dissociation
10:05			
10:10	Alena Khmelinskaia		
10:15	From Nature to the drawing board: design of dynamic protein assemblies	Alexandra Dashevsky Unexpected modes of cooperativity in the enzyme glutathione reductase	Marian Novotny Bound or not bound - annotating conformational state of protein structure models
10:20			
10:25	Anna Backeberg		
10:30	De novo design of multiple-geometry forming protein scaffolds	Florian Pretorious (DNA and heterodimers)	Lorenzo Scutteri De Novo Design Of Light-Regulated Dynamic Proteins Using Deep Learning
10:35			
10:40	Federico Alberto Olivieri		
10:45	Prosculpt: An Open-Source Pipeline for Accessible Protein Design		Kathryn Shelley (Light switches)
10:50			
10:55	Michael Adrian Jendrusch	Leoni Abendstein	
11:00	Efficient and Versatile Protein Structure Generation with SALAD	Designing for exclusivity: Can we engineer orthogonal heterodimers de novo using sequence-based design approaches?	
11:05			
11:10	Coffee break Sponsored by Fidabio	Coffee break Sponsored by Novartis	Coffee break
11:15			
11:20			
11:25			
11:30			
11:35			
11:40	Gustav Oberdorfer (Enzyme design) chair: Paola Panizza	Jeff Gray (Antibodies) chair: Georg Kuenze	Amelie Stein (PPI prediction & binders) chair: Alena Khmelinskaia
11:45			
11:50			
11:55			
12:00			
12:05	Sajith Kolathuparambil	Pedro Moreira	Gustavo Avelar Molina
12:10	Computational design of new proteases	Computational design of monobody binders to target viral epitopes	Versatile In-Vitro Validation of AI-Designed Protein Binders
12:15			
12:20	Rajarshi SinhaRoy	Tomer Cohen	Enrique Marcos
12:25	pHoptNN: Predicting Enzyme pH-Optima from Structure Using Equivariant Graph Neural Networks	Epitope-specific antibody design using diffusion models on the latent space of ESM embeddings	De novo design principles for stabilizing β-sheet frameworks to enable novel binding functions
12:30			
12:35	Salomé Guilbert	Klara Kropivsek	Simon Kohl
12:40	Binding affinities and computational optimisation of fluoroacetate dehalogenase to break carbon-fluorine bonds in "forever chemicals"	Multi-Objective Active Learning for Nanobody Development	Generating All-Atom De Novo Protein Binder Designs with High Lab Success Rates
12:45			
12:50	Georg Kuenze	Natalia Rincon	Xinchen Lyu
12:55	Computational design of high-efficiency PET-degrading enzymes with Rosetta and structure-guided fragment recombination	Evaluating The Biological Realism Of Iglm-Generated Antibody Sequences	De novo design of mini-protein binders broadly neutralizing Clostridioides difficile toxin B variants
13:00			
13:05	Natasha Poppy Murphy	Tadej Medved	Kasper H Björnsson
13:10	Engineering and application of a thermostable MHETase for PET de-polymerization	Analysis of Pareto optimality at the antibody-antigen interface	A strategy for the de novo design of highly polyspecific proteins targeting coral snake three-finger toxins
13:15			
13:20	Group photo		
13:25			
13:30	Lunch	Lunch	Lunch
13:35			
13:40			
13:45			
13:50			
13:55			
14:00			
14:05			
14:10			
14:15			

	Monday 20th	Tuesday 21st	Wednesday 22nd
14:20	Jens Meiler (AI & small molecule design) chair: Marian Novotny	Patrick Bryant (Non-canonicals & AI) chair: Florian Praetorius	Moritz Ertelt Practically useful computational liability analysis for protein-based binders
14:25			
14:30			
14:35			Chentong Wang Generative Design of Functional Proteins by Fusing Sequence and Structural Modalitie
14:40			
14:45	Johnny Alexander Jimenez Siegert	Alex Li	
14:50	Scalable and Energy-Efficient Ultra-Large Library Screening on a SpiNNaker2 Neuromorphic Chip	All-atom protein design via SE(3) flow-matching with ProteinZen	Rita Melo Targeting the RANK-TRAF6 Axis: A Molecular Approach
14:55			
15:00	Andrea Hunklinger	Tereza Neuwirthova	
15:05	Explainable AI Insights for Biologically Informed Training, Evaluation, and Interpretation of Protein Language Models	Prediction of Amino Acid Dispensability in Model Organisms	Yang Yang AI protein design at ACE2 binding interface
15:10			
15:15	Henry Dieckhaus	Tadeas Kalvoda	
15:20	Building hydrogen bond networks in silico with HBDesigner	Amino Acid Alphabet (R)Evolution: Changing the Molecular Basis of Life	Anastassia Vorobieva (Membrane proteins) chair: Adam Broerman
15:25			
15:30	Ana Paola Panizza Scasso	Tal Levin	
15:35	Leveraging Computational tools for Cost-Effective Enzyme Design in Uruguay, South America.	Combinatorial assembly and design of novel enzyme specificities	
15:40			
15:45	Bijoy Desai	Qihao Zhang	Alvaro Martin
15:50	The Open Datasets Initiative and Protein Engineering Tournament	Heme3D: A Heme Location Prediction Tool with 3dCNN/Docking Procedure	Expression of a de novo fluorescence-activating transmembrane β-barrel in mitochondrial membranes
15:55			
16:00	Coffee break Sponsored by Fidabio	Coffee break Sponsored by Novartis	Closing
16:05			
16:10			
16:15			Coffee break
16:20			
16:25			
16:30	Flash talks 1 (Odd poster numbers)	Flash talks 2 (Even poster numbers)	**Industry career panel** Julian Englert Marina Klemenčič Franziska Geiger Daniella Pretorius George Sutherland Adam Kral Anna Macintyre
16:35			
16:40			
16:45			
16:50			
16:55			
17:00			
17:05			
17:10			
17:15	Poster session 1 (Odd poster numbers) (With beer, wine or drinks)	Poster session 2 (Even poster numbers) (With beer, wine or drinks) (Light snack -- pizza and štruklji)	
17:20			
17:25			
17:30			
17:35			
17:40			
17:45			
17:50			
17:55			
18:00			
18:05			
18:10			
18:15			
18:20			
18:25			
18:30			
18:35			
18:40			
18:45			
18:50			
18:55			
19:00			
	PI meeting 19:30-20:00		
	Gala dinner 20:00-22:00	David Baker (remote lecture streamed in Grand hall) 20:00-21:00	

KEYNOTE SPEAKERS

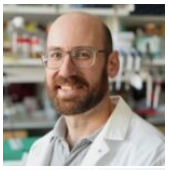
David Baker
University of Washington
(remote lecture)

Birte Höcker
University of Bayreuth, Germany

Roman Jerala
National institute of Chemistry, Slovenia

Tanja Kortemme
University of California San Francisco

INVITED SPEAKERS



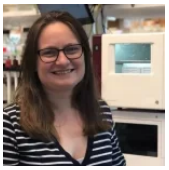
Florian Praetorius
Institute of Science and Technology Austria

Patrick Bryant
Vanderbilt University

Jeffrey Gray
Johns Hopkins University

Jens Meiler
Vanderbilt University

Kathryn Shelley
University of Washington



Anastassia Vorobieva
VIB-VUB Center for Structural Biology

Gustav Oberdorfer
Technical University in Graz

Amelie Stein
Copenhagen University

ORGANIZERS

Ajasja Ljubetič
National Institute of Chemistry, Slovenia

Dominik Gront (co-organizer)
University of Warsaw, Poland

LOCAL ORGANIZING COMMITTEE

Federico Olivieri

Eva Rajh

Liza Ulčakar

Alina Konstantinova

Katja Prek

Piotr Śmieja

Žan Žnidar, Teo Stupar, Nej Bizjak

Monday, Oct 20

Session 1: Molecular machines

KEYNOTE LECTURE

Designing artificial protein machines

Birte Höcker[1]

[1] Department of Biochemistry, Universität Bayreuth, Bayreuth, Germany

birte.hoecker@uni-bayreuth.de

Nature has evolved proteins into sophisticated molecular machines that sustain life. Today, advances in computational modelling and artificial intelligence are transforming how we design such systems, opening new avenues for engineering protein function. A key challenge in synthetic biology is to construct an artificial protein motor that is capable of autonomously stepping along a track.

We developed a roadmap to address this challenge using a modular design strategy that combines natural, non-motor proteins with de novo design (1). This framework defines the criteria for a successful artificial motor and identifies key design principles: enzymatic control of track binding, engineered asymmetry for directional movement, and allosteric regulation for internal coordination.

A milestone on this path is the development of *Tumbleweed*, a three-legged, clocked, linear walker (2). Each leg contains a distinct DNA-binding domain, enabling controlled interaction with a DNA track. *Tumbleweed* operates via a Brownian ratchet mechanism, where diffusion-driven steps are rectified by sequential ligand inputs. Using single-molecule fluorescence assays and a microfluidic device, we demonstrate directional stepping in response to ligand changes.

To transition from externally clocked to autonomous walking, enzymatic control will be essential. As enzyme design remains a challenge, I will further revisit the current state and show an example of how integration of physics- and AI- based tools enabled us to transform minimal *de novo* scaffolds into functional enzymes. For example, we engineered a lid structure to position catalytic residues and form an active-site pocket tailored for the Kemp elimination reaction (3). I will share insights from our design experience, highlighting both the opportunities and challenges encountered.

References:

[1] Nilsson, Zink, Laprevote, *et al.* (2024) Walking by design: how to build artificial molecular motors made of proteins. Nano Futures 8, 042501.
[2] Nilsson, Robertson, Gustafsson, Davies, *et al.* (2025) An artificial motor protein that walks along a DNA track. bioRxiv, doi:10.1101/2025. 03.07.641123.
[3] Beck, Smith, *et al.* (2025) Customizing the Structure of a Minimal TIM Barrel to Craft a De Novo Enzyme. bioRxiv, doi:10.1101/2025.01.28.635154.

De novo designed random walker

Liza Ulčakar[1,2], Hao Shen[3], Eva Rajh[1], Tadej Satler[1], Federico Olivieri[1], Joseph Watson[3], Yang Hsia[3], Justin Decarreau[3], Eric Lynch[4], Justin Kollman[4], David Baker[3], Ajasja Ljubetič`[1]

[1] National Institute of Chemistry, Department of Synthetic Biology and Immunology, Ljubljana, Slovenia
[2] Faculty of medicine, University of Ljubljana, Ljubljana, Slovenia
[3] Baker Lab, University of Washington, Seattle, US
[4] Kollman Lab, University of Washington, Seattle, US

liza.ulcakar@ki.si

De novo protein design has reached important milestones such as design of binders, diverse oligomeric assemblies, one- and two-dimensional finite assemblies. However, a *de novo* designed protein machine which could convert energy into mechanical work has not been demonstrated. In this work we describe a *de novo* designed protein-based molecular walker that diffuses randomly on designed protein tracks. The system will serve as a scaffold for the design of a powered molecular motor.

Our walkers are built from designed cyclic oligomers[1,2], while the tracks are based on designed protein fibers[3] (Figure 1). Walker–track interactions are mediated by coiled-coil (CC) formation, with one CC partner flexibly attached to each walker protomer (foot) and the complementary partner attached to the fibers (foothold). We prepared five walkers with varying numbers of protomers (feet) and three track types differing in CC length, which determines foothold affinity. We characterized each walker-track pair by measuring the diffusion of the walkers on tracks with total internal reflection fluorescence microscopy and determining the kinetics of the interaction by biolayer interferometry. Diffusion measurements revealed that walkers with higher valency move faster along the track. Walkers also diffused slightly faster on tracks with longer footholds, i.e. higher affinity. Kinetic measurements showed that this difference arises from a higher stepping rate on longer-foothold tracks.

These findings demonstrate that *de novo* design can produce complex systems with tunable biophysical properties. They also provide new insight into how valency, binding affinity, and stepping kinetics influence the diffusion of multivalent ligands on two-dimensional tracks. Based on the lessons learned, we are actively working on developing a powered walker.

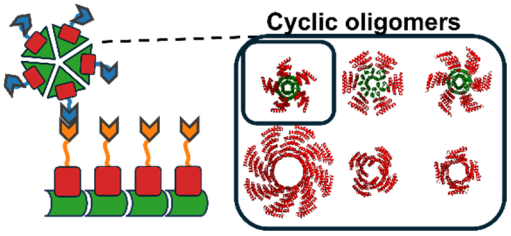


Figure 1: Schematic representation of the random walker system. *Helical bundles and designed helical repeats are shown in green and red, respectively. A coiled-coil of varying length (orange) is attached to each fibre subunit, while its orthogonal partner (blue) is attached to each subunit of the oligomers.*

References:

1. Edman, N. I. *et al.* Modulation of FGF pathway signaling and vascular differentiation using designed oligomeric assemblies. 2023.03.14.532666 Preprint at https://doi.org/10.1101/2023.03.14.532666 (2023).
2. Hsia, Y. *et al.* Design of multi-scale protein complexes by hierarchical building block fusion. *Nat. Commun.* **12**, 2294 (2021).
3. Shen, H. *et al.* De novo design of self-assembling helical protein filaments. *Science* **362**, 705–709 (2018).

From Nature to the drawing board: design of dynamic protein assemblies

Alena Khmelinskaia

Ludwig Maximilian University of Munich, Munich, Germany

akhmelin@cup.lmu.de

Recent computational methods have been developed for designing novel protein assemblies with atomic-level accuracy. Yet, when compared to their natural counterparts, the structural and functional space covered by de novo designed assemblies remains limited. I will share with you our ongoing virus-inspired efforts in diversifying the structural repertoire of protein assemblies and developing strategies to dynamically control protein assembly state. First, I will describe our approaches to the design of complex polyhedral geometries, from linked architectures assembled from rigid building blocks to the introduction of structural flexibility and the use of quasi-equivalence principles. Then, I will present our recently developed interface-seeded design approach and its use in the generation of symmetric or “sort of” symmetric oligomers responsive to a diversity of stimuli, e.g. metals, small molecules or phosphorylation states, as well as building foldamer-protein hybrid materialsp.

De novo design of multiple-geometry forming protein scaffolds

Anna Backeberg[1], Elizaveta Maltseva[1], Ho Yeung Chim[1], Alena Khmelinskaia[1]

[1] Ludwig-Maximilian-University Munich, Department of Chemistry, Munich, Germany

anna.backeberg@campus.lmu.de

Polyhedral protein assemblies are highly functional materials with applications in areas such as drug-delivery or vaccine development. While the de novo design of static protein nanoparticles with one specific target geometry is well established and exceedingly accurate [1], this study aims at the design of rigid protein monomers that can simultaneously satisfy and self-assemble into different geometries. Just like having a versatile ingredient in a kitchen pantry, our goal is to establish "off-the-shelf" programmability of protein assemblies with a set of universal building blocks, facilitating the targeted design of new and functional protein nanoparticles. Our multiple-geometry forming proteins additionally serve as a model system for biophysical studies of cooperative self-assembly in *de novo* protein assemblies.

Combining current Artificial Intelligence (AI)-based design methods with physics-based approaches, we established a pipeline for the design of multiple-geometry forming protein scaffolds. To do so, we develop approaches to tackle the challenge on two levels: (i) the generation of backbones matching multiple geometries; and (ii) the design of degenerate sequences compatible with the formation of both target architectures. I will present the first results validating our design pipeline *in silico* and *in vitro*. Taken together, our approach paves the way for a new generation of highly versatile protein building blocks.

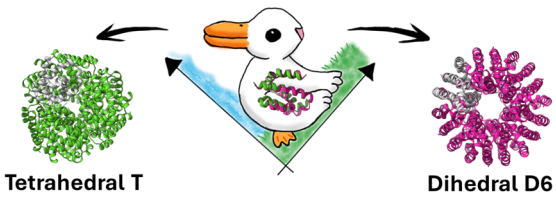


Figure 1: Designed scaffolds form multiple target geometries. *We design degenerate sequences for our rigid protein building blocks, equally compatible with forming different assembly geometries, similar to how the same set of lines can be used to draw either a duck or a rabbit.*

References:

[1] Mallik, Bhoomika Basu, et al. "De novo design of polyhedral protein assemblies: before and after the AI revolution." ChemBioChem 24.15 (2023): e202300117.

Prosculpt: An Open-Source Pipeline for Accessible Protein Design

Federico A. Olivieri[1], Alina Konstantinova[1], Nej Bizjak[1], Žan Žnidar[1], Tilman Scholz[1], Kiyan Abel[1], Liza Ulčakar[1], Eva Rajh[1], Ajasja Ljubetič[1]

[1] National Institute of Chemistry, Department of Synthetic Biology and Immunology, Ljubljana, Slovenia

Federico.alberto.olivieri@ki.si

Protein design has rapidly advanced with generative and predictive methods, yet many tools remain difficult to use without extensive computational expertise. We present **Prosculpt[1]**, an open-source framework that integrates Rfdiffusion[2], ProteinMPNN[3], AlphaFold[4], and Rosetta scoring into a single, user-friendly pipeline. Designed to run on standard high-performance clusters with minimal setup, Prosculpt requires only a single configuration file to manage all settings for structure generation, sequence design, and structure prediction. Beyond integration, Prosculpt enhances these tools with improved handling of constructs that combine designed and natural protein regions and supports automated error handling and workflow resumption. By lowering technical barriers, Prosculpt enables researchers with little or no computational experience to access advanced protein design workflows and apply them directly to their projects without sacrificing any functionality of the underlying tools. I will present two use cases with experimental results: the rigid fusion of a protein to a peptide and the design of inhibitory binder for the assembly of ferritin cages.

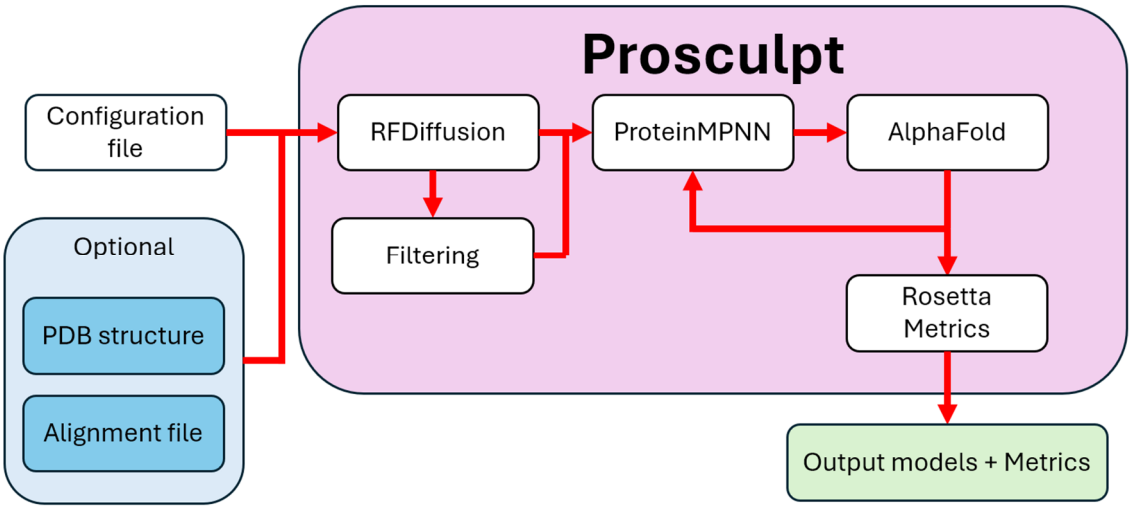


Figure 1: *Basic diagram of Prosculpt pipeline*

References:

[1] https://github.com/ajasja/prosculpt
[2] Watson, J.L., Juergens, D., Bennett, N.R. et al. De novo design of protein structure and function with RFdiffusion. Nature 620, 1089–1100 (2023). https://doi.org/10.1038/s41586-023-06415-8.
[3] J. Dauparas *et al.* Robust deep learning–based protein sequence design using ProteinMPNN.*Science*378,49-56(2022).DOI:10.1126/science.add2187
[4] J. Dauparas et al. Robust deep learning–based protein sequence design using ProteinMPNN.Science378,49-56(2022).DOI:10.1126/science.add2187

Efficient and Versatile Protein Structure Generation with SALAD

Michael A. Jendrusch[1,2,3], Jan O. Korbel[1]

[1] European Molecular Biology Laboratory (EMBL), Department of Genome Biology, Heidelberg, Germany
[2] Collaboration for joint PhD degree between EMBL and Heidelberg University, Faculty of Biosciences
[3] Institute for Pharmacy and Molecular Biotechnology (IPMB), Heidelberg University, Heidelberg, Germany

michael.jendrusch@embl.de

Protein backbone generation, sequence design and structure predictor-based filtering form the backbone of the modern machine learning-assisted protein design pipeline. However, current protein backbone generators have to contend with a three-way trade-off between efficiency, designability and versatility. While comparatively slow, hallucination-based approaches reach high designability even for large proteins and can easily be adapted to different protein design tasks merely by changing the fitness function being optimized. Conversely, protein structure diffusion models offer high generation efficiency and designability for small proteins, at the cost of versatility and designability for large proteins.

To address this issue, we introduce a family of sparse all-atom denoising models (salad) for efficient and flexible protein backbone generation [1]. salad addresses efficiency by using sparse attention without persistent amino acid pair features, outperforming previous diffusion models in terms of runtime. Simultaneously, salad bridges the designability gap to hallucination-based methods, reliably designing protein backbones with up to 1,000 amino acids. Finally, by combining salad with structure-editing – a modified generative process for protein structure diffusion models – we can adapt it to produce scaffolds for protein motifs, repeat proteins and proteins predicted to adopt multiple conformations under different conditions, bridging the gap in versatility to hallucination-based methods.

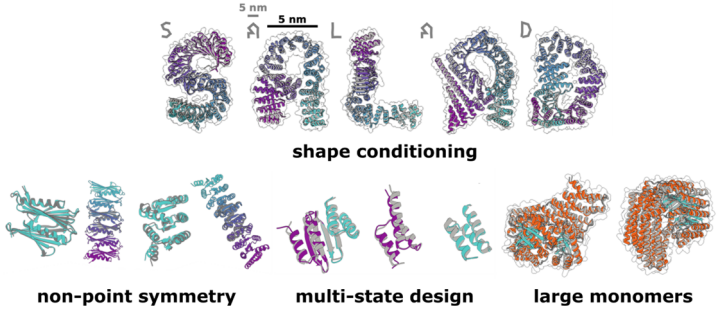


Figure 1: salad design modalities. *Examples of design modalities accessible using salad: shape conditioned design, repeat proteins with non-point symmetries, multi-state design and large monomers with up to*

References:

1. M. A. Jendrusch, J. O. Korbel. Efficient protein structure generation with sparse denoising models. (2025) *bioRxiv* 2025.01.31.635780

Session 2: Enzyme design

INVITED TALK

A hybrid machine learning and atomistic modeling approach for the design of de novo enzymes

Markus Braun[1], Adrian Tripp[1], Morakot Chakatok[1], Sigrid Kaltenbrunner[1], Celina Fischer[1], David Stoll[1], Aleksandar Bijelic[1], Wael Elaily[1], Massimo G. Totaro[1], Melanie Moser[1], Shlomo Y. Hoch[1,2], Horst Lechner[1], Federico Rossi[3], Matteo Aleotti[3], Mélanie Hall[3,4,5], Gustav Oberdorfer[1,5]*

[1]Institute of Biochemistry, Graz University of Technology; Petersgasse 12/2, 8010 Graz, Austria.
[2]Weizmann Institute of Science; Herzl St 234, Rehovot, Israel.
[3]Institute of Chemistry, University of Graz; Heinrichstrasse 28, 8010 Graz, Austria
[4]BioHealth, Field of Excellence, University of Graz, Austria
[5]BioTechMed Graz, Mozartgasse 12/2, 8010 Graz, Austria

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 activities. 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 and Morita Baylis-Hillman reaction, as well as metal cofactors of increasing complexity.

Computational design of new proteases

Sajith Kolathuparambil[1,2], Horst Lechner[1,2]

[1] Graz University of Technology, Institute of Biochemistry, Graz, Austria
[2] BioTechMed-Graz, Graz, Austria

sajith.kolathuparambil@tugraz.at

Current strategies for recombinant protein purification often rely on the use of affinity tags, which are typically removed after purification by site-specific proteases. However, most endoproteases cleave at defined sites located toward the C-terminus of their recognition sequences [1]. Consequently, when C-terminal affinity tags are used, these proteases often leave residues from the recognition sequence at the C-terminus of the recombinant protein. Such residues can adversely affect proper folding or crystallization of the target protein [2]. Achieving complete and precise removal of C-terminal affinity tags without leaving any extraneous residues remains a significant challenge.

Using state-of-the-art de novo computational design tools, we engineered sequence-specific proteases based on matrix metalloproteases (MMPs) and designed armadillo repeat protein (dArmRP) scaffolds to achieve precise cleavage of C-terminal tags without leaving additional residues. Preliminary computational results demonstrate improved binding affinity and specificity of protease-peptide pairs. Top candidates were selected via in silico protease-peptide binding affinity scoring followed by experimental screening using a sensitive green fluorogenic GFP complementation assay.

Our approach paves the way for the rapid development of proteases tailored for precise cleavage of custom tags, offering broad implications for recombinant protein production in structural and functional studies.

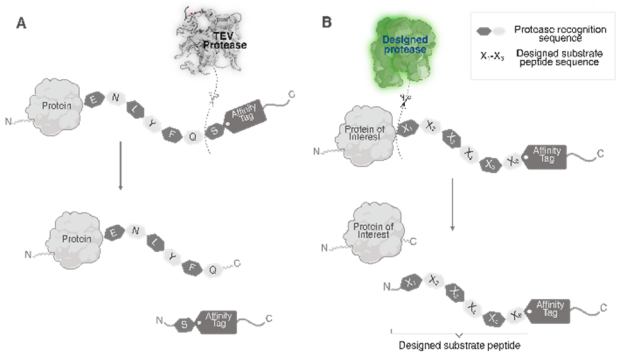


Figure 2: Schematic representation of C-terminal affinity tag removal strategies. ***A)*** *Conventional approach using TEV protease leaves residues from the recognition sequence attached to the recombinant protein.* ***B)*** *Our designed protease-peptide approach achieves complete tag removal without leaving any recognition sequence residues on the target protein.*

References:

[1] Waugh D. S. (2011). An overview of enzymatic reagents for the removal of affinity tags. *Protein expression and purification*, *80*(2), 283–293.
[2] Eschenfeldt, W. H., Maltseva, N., Stols, L., Donnelly, M. I., Gu, M., Nocek, B., Tan, K., Kim, Y., & Joachimiak, A. (2010). Cleavable C-terminal His-tag vectors for structure determination. *Journal of structural and functional genomics*, *11*(1), 31–39.

pHoptNN: Predicting Enzyme pH-Optima from Structure Using Equivariant Graph Neural Networks

Rajarshi SinhaRoy[1], Christian Clauss[1], Georg Künze[1]

[1]Institute of Drug Discovery, Leipzig University, 04103 Leipzig, Germany
sinharoy.rajarshi@uni-leipzig.de

The catalytic pH optimum is a critical parameter that plays a vital role in determining enzyme function. It significantly influences various aspects such as enzyme stability, activity, and the potential applications of enzymes in fields like biotechnology and pharmaceuticals. While there are sequence-based machine learning models available that can predict pH optima, they often overlook the essential structural context that is crucial for accurately assessing enzyme behavior.

In this study, we present an innovative structure-based deep learning framework that leverages E(n)-Equivariant Graph Neural Networks (EGNNs) to predict the catalytic pH optima of enzymes based on their three-dimensional structures. To facilitate this research, we curated a comprehensive dataset known as **EnzyBase12k**, which includes more than 11,600 enzymes. This dataset uniquely combines experimentally determined pH optima with high-confidence structural models sourced from AlphaFold3 and the Protein Data Bank, providing a robust foundation for our analysis.

Our models are designed to encode atomic-level molecular graphs, effectively integrating a variety of features, including geometric, chemical, and electrostatic properties. We conducted extensive hyperparameter optimization employing advanced techniques such as evolutionary algorithms, Bayesian optimization, and label-aware loss reweighting. As a result, the best-performing EGNN achieved a remarkable test root mean square error **(RMSE) of 0.556**. This performance surpasses that of leading sequence-based methods, including EpHod[1]. Furthermore, our model demonstrates a strong ability to generalize across different enzyme classes and exhibits robustness under both acidic and alkaline pH conditions, highlighting its potential for broad applications in enzyme research and development.

References:

[1] Gado, J.E., Knotts, M., Shaw, A.Y. *et al.* Machine learning prediction of enzyme optimum pH. *Nat Mach Intell* **7**, 716–729 (2025). https://doi.org/10.1038/s42256-025-01026-6

[2] Satorras, V. G., Hoogeboom, E., & Welling, M. (2022). E(n) Equivariant Graph Neural Networks. *arXiv [Cs.LG]*. Retrieved from http://arxiv.org/abs/2102.0984

Binding affinities and computational optimisation of fluoroacetate dehalogenase to break carbon-fluorine bonds in “forever chemicals”

Salomé Guilbert[1,2], Philippe Schwaller[1], Ursula Roethlisberger[2]

[1] Institute of Chemical Science and Engineering, Laboratory of Computational Chemistry and Biochemistry, Ecole Polytechnique Fédérale de Lausanne, Switzerland
[2] Institute of Chemical Science and Engineering, Laboratory of Chemical Artificial Intelligence, Ecole Polytechnique Fédérale de Lausanne, Switzerland

salome.guilbert@epfl.ch

Increasing concern over the environmental persistence and toxicity of per- and polyfluoroalkyl substances (PFAS), often referred to as “forever chemicals”, has pushed for the need to design technologies for their remediation. Tri-fluoroacetate (TFA) has particularly accumulated in the environment, as it is the main degradation product of longer-chain PFAS. However, few bacteria and no enzymes have yet been discovered to achieve the full degradation of TFA, while the only commercially available PFAS degradation method involves costly and unsustainable thermal deconstruction [1]. It is therefore urgent to design a biomimetic system offering a cost-effective and sustainable solution to degrade small-chain PFAS such as TFA.

To this end, we first selected a suitable biological target that had the potential to be optimised for better substrate binding, catalytic activity and thermal stability. Fluoroacetate dehalogenase (FAcD) was chosen, a dimeric enzyme capable of breaking the strong carbon-fluorine bond (~460 kJ/mol, Figure 1a) in mono-fluoroacetate (MFA) [2]. The binding free energy profiles of MFA and TFA entering the natural enzyme were determined using classical thermodynamic integration, an enhanced sampling method. Results showed that there was no barrier for MFA to enter the binding pocket compared to TFA, and highlighted the key residues that were hindering the diffusion and stabilisation of TFA in FAcD (Figure 1b). Alchemical free energy calculations further allowed to obtain the relative binding affinities of MFA, TFA and other short-chain PFAS in the enzyme.

The next step involved re-designing the protein-protein interface of FAcD to turn it into an active monomer, using a combination of RFDiffusion, ProteinMPNN, FastRelax and AlphaFold2 [3]. Starting with five monomer candidates, their reactivity towards TFA will now be optimised with the help of the genetic algorithm tool EVOLVE [4]. Catalytic activity will finally be tested using QM/MM.

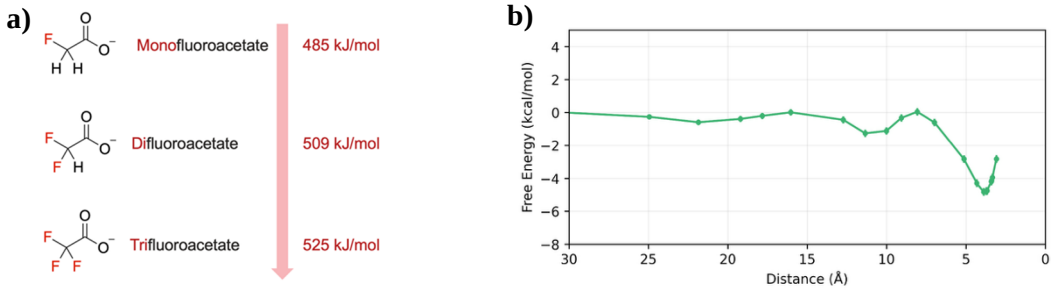


Figure 1. a) *C-F bond dissociation energy of MFA and FTA [2]* ***b)*** *Free energy profile of MFA entering FAcD using thermodynamic integration, binding energy of -4.95 kcal/mol.*

References:

[5] E.M. Suderland et al.. *J Expo Sci Environ Epidemiol*, 29(2):131-147, 2019
[6] S. Farajollahi et al., *ACS Omega*, 26(9):28546-28555, 2024
[7] R. Krishna et al., *Science*, 384(6693), 2024
[8] N. Browning et al., *ChemRxiv*, 2023

Computational design of high-efficiency PET-degrading enzymes with Rosetta and structure-guided fragment recombination

Georg Künze[1,2,3]

[1] Institute for Drug Discovery, Leipzig University, Leipzig, Germany
[2] Interdisciplinary Center for Bioinformatics, Leipzig University, Leipzig, Germany
[3] Center for Scalable Data Analytics and Artificial Intelligence, Leipzig, Germany

georg.kuenze@uni-leipizg.de

The discovery and development of **plastic-degrading enzymes** is attracting significant interest due to the potential of these enzymes to enable **eco-friendly plastic recycling** solutions. Considerable progress has been made in the study of poly(ethylene terephthalate) (PET)-degrading enzymes, which could be discovered in natural sources and optimized by means of **protein engineering**. My group has developed **high-efficiency PET-degrading enzymes** by redesigning the **polyester hydrolase Leipzig 7 (PHL7) –** a PET hydrolase that we isolated from a plant compost metagenome and that rapidly degrades low-crystalline PET at 70 °C [1]. Over the last years, we have applied different computational design strategies (Figure 1) to overcome the limited stability and activity of the wildtype PHL7 and improve its catalytic performance at industrially relevant conditions. Using **Rosetta design calculations**, we installed up to 24 mutations into PHL7, leading to variants with thermal stabilities of 88 - 95 °C and over 110-fold higher activity in 0.1 M phosphate buffer compared to the wildtype PHL7. When tested under high substrate concentrations in a bioreactor, the best variants degrade ~80% low-crystalline PET in 24 h at 65 °C, clearly outperforming the benchmark enzyme ICCG. In addition, to increase the design space of PET hydrolases, we have generated a diverse enzyme family using **structure-guided recombination** of different PET hydrolase templates, including PHL7, LCC and TurboPETase. The new chimeric enzymes show higher activity and stability than their parent enzymes in some cases, which will be further enhanced in the next rounds using **machine learning-guided directed evolution**. In conclusion, the PET hydrolases that we have engineered so far provide good candidates for the development of scalable biocatalytic PET recycling processes.

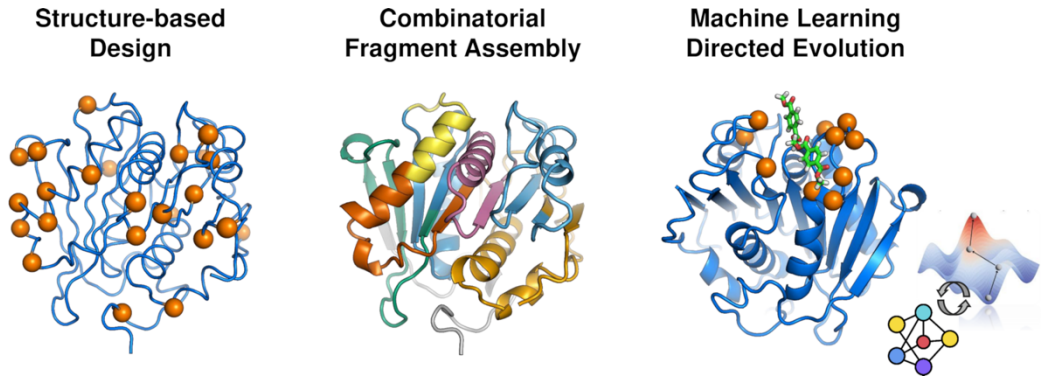


Figure 1: Design strategies used to develop high-efficiency PET hydrolases. Rosetta design of multi-point mutants (left), design of chimeric enzymes by combinatorial fragment reassembly (middle) and machine learning-guided directed evolution of the active site of PHL7 (right).

References:

[1] C. Sonnendecker et al. Low Carbon Footprint Recycling of Post-Consumer PET Plastic with a Metagenomic Polyester Hydrolase. *ChemSusChem,* 15(9):e202101062, 2022

Engineering and application of a thermostable MHETase for PET depolymerization

Natasha P. Murphy,[1,2] Japheth E. Gado,[1,2,3] Irimpan I. Mathews,[4] Evan Komp,[1,3] Elizabeth L. Bell,[1,2] Brenna Norton-Baker,[1,2,3] Matilda Clark,[5] Luisana Avilan,[5] Rebecca Garcia,[1] Hannah M. Alt,[1] Ritimukta Sarangi,[4] Andrew R. Pickford,[5] John E. McGeehan,[1,2] Nicholas P. Gauthier,[6,7] Gregg T. Beckham[1,2,3,*]

[1] Renewable Resources and Enabling Sciences Center, National Renewable Energy Laboratory, Golden, CO 80401
[2]BOTTLE Consortium, Golden, CO 80401
[3] Agile BioFoundry, Emeryville, CA 94608
[4] SLAC National Accelerator Laboratory, Stanford Synchrotron Radiation Lightsource, Menlo Park, CA 94025
[5] Centre for Enzyme Innovation, School of the Environment and Life Sciences, University of Portsmouth, Portsmouth, PO1 2DY
[6] Department of Systems Biology, Harvard Medical School, Boston, MA 02115
[7] Department of Data Sciences, Dana-Farber Cancer Institute, Boston, MA 02115

* Corresponding author: gregg.beckham@nrel.gov

Enzymatic hydrolysis of poly(ethylene terephthalate) (PET) releases mono(2-hydroxyethyl) terephthalate (MHET) as a major product, the accumulation of which can prolong reactor residence times and complicate downstream separations from the monomers. The use of an MHETase enzyme can enable MHET hydrolysis to the monomers, terephthalic acid and ethylene glycol, but since industrial PETases are used at thermophilic temperatures and the well-known MHETase from *Ideonella sakaiensis* is a mesophilic enzyme, it is important to develop thermophilic MHETases. Here, we characterize thermostable MHET-active enzymes from a natural diversity screen by applying a hidden Markov model trained on the previously reported, archaeal ferulic acid esterase, PET46.[1] We identified enzymes with higher melting temperatures than PET46 and quantified their MHETase activity in reactions at 70 °C. The crystal structure of MHT077, the homolog with the highest MHETase activity and a melting temperature (T_m) of 94.6 °C, informed site saturation mutagenesis in the active site and lid-domain interface. MHT077 exhibited a ~100-fold slower unfolding rate at 65 °C than PET46, indicating substantially greater kinetic stability. In parallel, we applied EVcouplings,[2] a machine-learning-based method that leverages co-evolutionary patterns in large multiple sequence alignments, to improve the thermostabilities of five ferulic acid esterases. One design, EV-MHT043-5 was identified with a T_m = 96.1 °C and a 3-fold improvement in its MHETase activity relative to the wildtype enzyme, MHT043. Combination variants of beneficial mutations were screened and afforded a variant, MHT077^{LFK}, with a ~4-fold improvement in MHETase activity at 70°C. Overall, this study expands the known MHET-hydrolyzing protein scaffolds available for enzymatic PET recycling.

References:

[1] Perez-Garcia, P., Chow, J., Costanzi, E. *et al.* An archaeal lid-containing feruloyl esterase degrades polyethylene terephthalate. *Commun Chem* **6**, 193 (2023).
[2] Hopf TA, Green AG, Schubert B, *et al.* The EVcouplings Python framework for coevolutionary sequence analysis. *Bioinformatics*. **9**, 1582-1584 (2019).

Session 3: AI & small molecule design

INVITED TALK

How Artificial Intelligence is reshaping Protein Structure Prediction and Therapeutic Design

Jens Meiler[1]

[1]Institute for Drug Discovery, Faculty of Medicine, Faculty of Mathematics and Informatics, Faculty of Chemistry and Mineralogy, University Leipzig, Leipzig, Germany
[2]Center for Scalable Data Analytics and Artificial Intelligence ScaDS.AI and School of Embedded Composite Artificial Intelligence SECAI, Dresden/Leipzig, Germany
[3]Department of Chemistry, Department of Pharmacology, Center for Structural Biology, Institute of Chemical Biology, Center for Applied Artificial Intelligence in Protein Dynamics, Vanderbilt University, Nashville, Tennessee, United States of America
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-large library 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 peptide therapeutic 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.

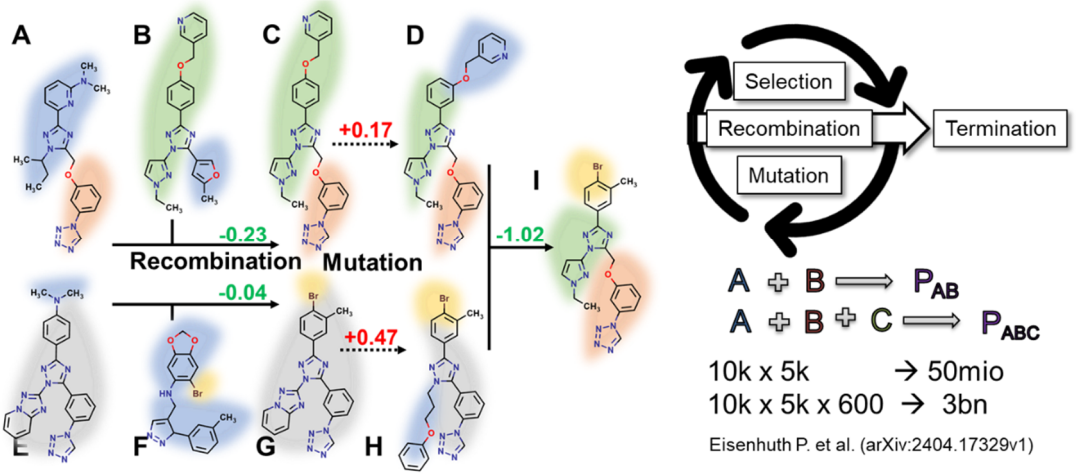


Figure 1: Rosetta Evolutionary Ligand – REvoLd: *Ultra-Large Library Screening with an Evolutionary Algorithm*

Scalable and Energy-Efficient Ultra-Large Library Screening on a SpiNNaker2 Neuromorphic Chip

Johnny Alexander Jimenez Siegert[1,4], Florian Kelber[2,5], Bernhard Vogginger[2,5], Paul Eisenhuth[1,2], Max Beining[1,4], Vivian Ehrlich[1], Johannes Partzsch[5], Christian Mayr[2,4,5], Jens Meiler[1,2,3,4]

[1] Institute for Drug Discovery, Leipzig University, Leipzig, Germany.
[2] Center for Scalable Data Analytics and AI (ScaDS.AI), Dresden/Leipzig, Germany.
[3] Center for Structural Biology (CSB), Vanderbilt University, Nashville, TN, USA.
[4] School of Embedded Composite AI (SECAI), Dresden/Leipzig, Germany.
[5] Institute of Circuits and Systems, Dresden University of Technology, Germany.
johnny.jimenez_siegert@uni-leipzig.de

Ultra-large libraries of synthetically accessible small molecules are transforming computer-aided drug discovery by making massive collections of drug candidates available for virtual screening and rapid experimental validation. As libraries grow to billions of molecules, exhaustive screening becomes prohibitively expensive, demanding large-scale cloud computing resources and incurring high energy costs. We propose to use the SpiNNaker2[1] platform, a neuromorphic system designed for massively parallel machine learning tasks, to enable energy-efficient virtual screening at scale.

Here, we present the implementation of a ligand-based ultra-large library screening approach on a SpiNNaker2 chip. We adapted neural networks trained on 2D molecular descriptors[2] as quantitative structure-activity relationship models to screen a subset of the Enamine REAL space. In comparison to a GPU-accelerated NVIDIA Jetson Orin system, the SpiNNaker2 chip achieved comparable throughput while consuming roughly 5x less energy per molecule. Although throughput was limited by memory transfer overheads, neural network inference was more than twice as fast on the SpiNNaker2 chip's accelerators. We highlight hardware features and optimizations to efficiently leverage the SpiNNaker2 system. These results demonstrate the potential of specialized hardware, such as the upcoming 5-million core SpiNNcloud system, to enable efficient billion-scale virtual screening.

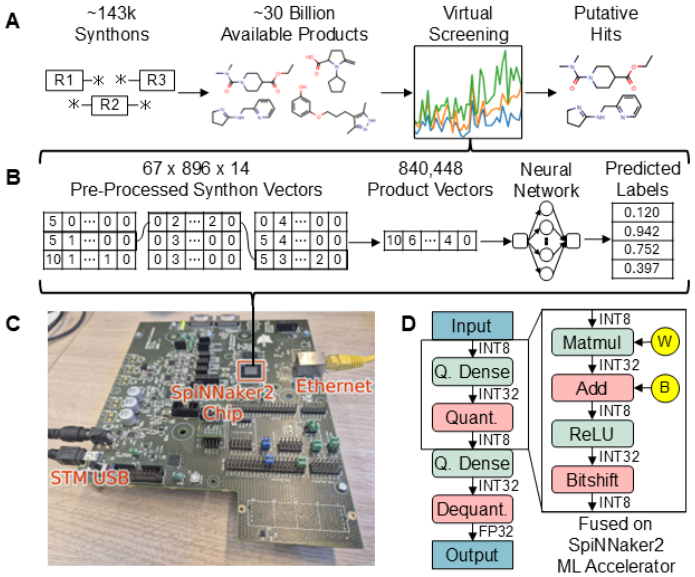


Figure 1: Ultra-large library screening on a SpiNNaker2 chip. A) The Enamine REAL space makes billions of molecules available for virtual screening and rapid experimental validation. B,C) We implemented a ligand-based screening pipeline for an Enamine REAL space reaction on a SpiNNaker2 chip with 152 processing elements. D) We quantized and optimized a neural network with 32 hidden neurons for inference on the SpiNNaker2 hardware accelerators.

References:

[1] Gonzalez, H. A. et al. SpiNNaker2: A Large-Scale Neuromorphic System for Event-Based and Asynchronous Machine Learning. *arXiv*, 2401.04491, 2024

[2] Vu, O. et al. (2019). BCL::Mol2D—a robust atom environment descriptor for QSAR modeling and lead optimization. *J. Comput.-Aided Mol. Des.*, 33(5), 477–486, 2019

Explainable AI Insights for Biologically Informed Training, Evaluation, and Interpretation of Protein Language Models

Andrea Hunklinger[1,2,3,4], Marcel Hiltscher[3], Santiago Villalba[4], Noelia Ferruz[1,5], Francesca Grisoni[3]

[1] Centre for Genomic Regulation, Barcelona, Spain
[2] Universitat de Barcelona, Barcelona, Spain
[3] Eindhoven University of Technology, Eindhoven, The Netherlands
[4] Bayer AG, Berlin, Germany
[5] Universitat Pompeu Fabra, Barcelona, Spain

andrea.hunklinger@crg.eu

Advancements in deep learning are transforming protein research. Transformer-based models have achieved remarkable success across tasks ranging from protein structure prediction to the design of functional enzymes[1]. However, these models often **operate as black boxes**, relying on mechanisms that are not directly interpretable to humans. Explainable AI (XAI) has been used to assess whether such models capture underlying scientific principles[2], yet it remains challenging to determine if the **extracted XAI results are trustworthy**, whether this information is actually used by the model to generate its output, and how we can make the most of these insights.

We propose a shift in the use of XAI from relying solely on post-hoc analysis of individual examples to extracting reliable explanations throughout the entire workflow. By incorporating **XAI into the training process**, we can guide model behavior in a more biologically informed manner. Integrating annotations such as amino acid features and functions, and comparing them with feature attribution results, supports the evaluation and **grounds it in biological context**. Depending on the model and task, whether binary classification of protein sequences, multimodal learning with enzyme–substrate representations, or novel sequence design, the analysis strategy and application of XAI insights must be adapted. As black-box models become increasingly prevalent in bioinformatics, integrating XAI methods into computational workflows and evaluating them rigorously should become standard practice.

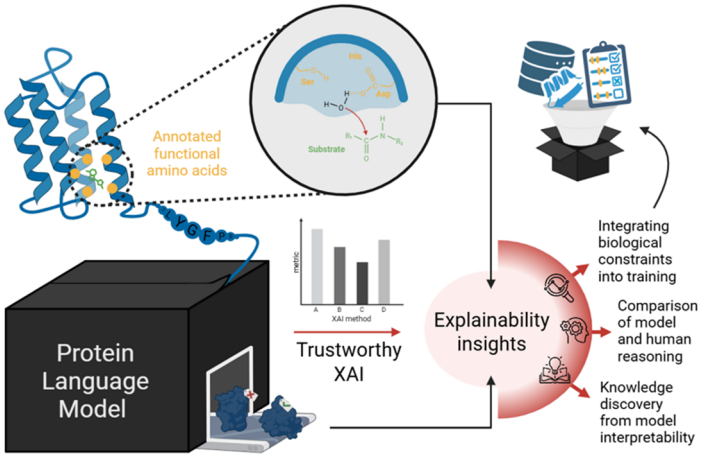


Figure 1: *Embedding trustworthy Explainable AI (XAI) in a protein language model workflow to steer training, validate outputs and judge model-internal reasoning.*

References:

[1] Zhang, Shuang, et al. Applications of transformer-based language models in bioinformatics: a survey. *Bioinformatics Advances*. 3(1); 2023

[2] Zhang, Zhidian, et al. Protein language models learn evolutionary statistics of interacting sequence motifs. *Proceedings of the National Academy of Sciences*. 121(45); 2024

Building hydrogen bond networks *in silico* with HBDesigner

Henry Dieckhaus[1,3], Nicholas Z Randolph[2,3], Brian Kuhlman[3,4]

[1] Division of Chemical Biology and Medicinal Chemistry, UNC Chapel Hill
[2] Department of Bioinformatics and Computational Biology, UNC Chapel Hill
[3] Department of Biochemistry and Biophysics, UNC Chapel Hill
[4] Lineberger Comprehensive Cancer Center, UNC Chapel Hill

dieckhau@unc.edu

Hydrogen bonds (H-bonds) involving protein sidechains can play a critical role in dictating protein structure and function [1]. Of particular interest for protein design, hydrogen bonding across protein-protein interfaces is important to achieve specific binding while avoiding aggregation [2]. However, current deep learning (DL) design tools struggle to model polar interactions within protein cores and interfaces.

To address this issue, we present a new DL-based protocol, **HBDesigner** (Figure 1). Our model, which takes a protein backbone as input, rapidly samples and packs clusters of amino acids capable of forming highly connected H-bond networks. We demonstrate through *in silico* experiments that HBDesigner samples diverse, native-like networks when applied to either native or *de novo* designed backbones and interfaces. Preliminary data indicates that HBDesigner achieves higher *in silico* success rates (i.e., refoldability) compared to Rosetta's Monte Carlo HBNet protocol [3] while offering similar or better runtimes, enabling integration into modern DL-based *de novo* design pipelines. Our method is highly flexible and can be conditioned to produce networks with a particular center of mass, burial level, or residue composition. Finally, we offer several motivating case studies demonstrating the application of HBDesigner to difficult design problems unsuited for current DL design tools. *In silico* validation efforts indicate that HBDesigner produces viable designs that can be recapitulated after refolding, and experimental characterization remains ongoing. We aim to provide the HBDesigner protocol as an open-source software package for use by the wider design community.

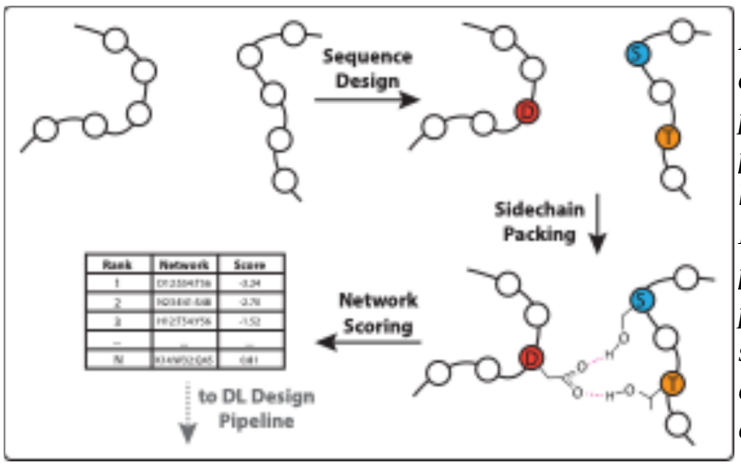


Figure 1: HBDesigner protocol overview. *Starting from an empty protein backbone, the model samples clusters of amino acids with high likelihoods of participating in H-bonding. Each sample is then packed with a specialized packing protocol and scored using the Rosetta energy function. Top-scoring designs may then used as inputs for downstream design tasks.*

References:

[1] D. Kuroda and J.J. Gray. *Bioinformatics*, 32(16):2451-2456, 2016.
[2] R.J. de Haas, et al. *Proc. Natl. Acad. Sci.* 121 (13):e2314646121, 2024.
[3] J.B. Maguire and S.E. Boyken, et al., *J. Chem. Theory Comput.* 14:2751-2760, 2018.

Leveraging Computational tools for Cost-Effective Enzyme Design in Uruguay, South America.

Gonzalo López[1, 2], Mage Teliz[1,2], Ariel Tijman[1,3], César Iglesias[1], Sonia Rodríguez Giordano[1], Pablo D. Dans Puiggròs[2,4], Paola Panizza Scasso[1]

[1] Biocatalysis and Biotransformations Laboratory and Molecular Microbiology Laboratory, DEPBIO, Faculty of Chemistry, Udelar, Uruguay, [2] Computational Biophysics Group, DCB, CENUR Litoral Norte, Udelar, Uruguay, [3] Molecular Enzymology, Faculty of Science and Engineering, University of Groningen, The Netherlands, [4] Bioinformatics Unit, Institut Pasteur de Montevideo, Uruguay

ppanizza@fq.edu.uy

In Uruguay, scientific research is constrained by limited funding, demanding an efficient use of available resources. The high cost of chemical reagents and laboratory equipment poses a major challenge for the screening of new biocatalysts. To address this, our group has adopted a rational design approach grounded in bioinformatics, applying it across three lines: ω-transaminases, imine reductases, and fluorinases. We aim to leverage structural bioinformatics, molecular modeling, and machine learning to advance biocatalyst discovery and optimization, even under resource-limited conditions.

We engineered stabilized variants of ω-transaminase TACap[1] using consensus sequence analysis, MD-based loop dynamics, and folding free energy calculations. In parallel, we successfully redesigned a native *Streptomyces* sp. imine reductase (Ss-IRED-R)[2] for the reductive amination of phenylacetone—a key step in PET radiotracer synthesis. Comparative analysis of the Ss-IRED-R and *Aspergillus oryzae* AspRedAm[3] active sites using homology models, molecular dynamics simulations, docking, and free energy calculations pinpointed key spots for mutation. Construction of single, double, and triple mutants revealed substitutions that conferred new activity. Iterative feedback from experiments continues to refine the computational approach in successive design cycles.

Additionally, we are tackling the challenge of enzymatic C–F bond formation by designing novel fluorinases. As nature has a low diversity of enzymes capable of performing this reaction, we are working on the application of *de novo* design using ProteinMPNN to explore the sequence space. Multidimensional physicochemical descriptors—electrostatic potentials (APBS), cavity mapping (CRAFT), and fluoride-probe interaction profiles (CMIP)— are used to select those sequences with highest chances of being an active fluorinase. Molecular dynamics are also used to assess candidate stability and substrate accessibility. Top hits from this pipeline are being prioritized for expression and catalytic evaluation, offering a route to expand the limited repertoire of natural fluorinases.

This work reflects how computational tools can democratize protein design, enabling meaningful advances in enzyme engineering from under-resourced settings.

[1] C. Iglesias, P. Panizza, S. Rodriguez Giordano, Applied Microbiology and Biotechnology 101 (2017) 5677–5687.
[2] C. Iglesias, A. Tijman, G. López, M.I. Lapaz, M.J. Pianzzola, P. Panizza, S. Rodríguez Giordano, Frontiers in Catalysis 1 (2021) 785963.
[3] G.A. Aleku, S.P. France, H. Man, J. Mangas-Sanchez, S.L. Montgomery, M. Sharma, F. Leipold, S. Hussain, G. Grogan, N.J. Turner, Nature Chemistry 9 (2017) 961–969.

The Open Datasets Initiative and Protein Engineering Tournament — Scaling Open Science for Protein Function Prediction

Bijoy Desai[1]

[1]The Align Foundation

bijoy@alignbio.org

The Align Foundation is a nonprofit organization committed to advancing biological science through reproducible, scalable, and shareable research. Central to this mission is our flagship GROQ-seq project—**Gro**wth-based **Q**uantitative **Seq**uencing—a high-throughput experimental platform designed to enable the quantitative measurement of protein function at scale. GROQ-seq empowers the systematic collection of large, standardized, and open-access datasets linking protein sequences to their biological functions, addressing a critical bottleneck in the development of accurate sequence-to-function predictive models.

The GROQ-seq platform is built around pooled growth-based assays, wherein a gene's function is coupled to cellular growth under selective pressure. Libraries of up to 500,000 barcoded protein variants are transformed, grown, and sequenced to quantify variant fitness. Function scores are calibrated using internal variant ladders—reference sets with known activities—yielding quantitative, reproducible measures of function at a cost of approximately $0.05 per sequence. The system is designed to be extensible, supporting diverse protein functions such as transcription factor binding, protease specificity, aminoacyl tRNA synthetase activity, RNA polymerase function, and more.

Our **Open Datasets Initiative** supports the GROQ-seq work, partnering with academic institutions and automation labs—including NIST and Boston University's DAMP Lab—to coordinate data generation across multiple sites. Seven distinct protein functions are currently in the GROQ-seq pipeline, with early data demonstrating robustness and reproducibility across independent labs. These datasets are shared through a public data portal, with provisions for community engagement, data reuse, and collaborative expansion.

Align also facilitates **benchmarking competitions** and open challenges that invite the broader scientific and machine learning communities to evaluate and refine predictive models using our datasets. Our upcoming **2025 Protein Engineering Tournament** is a global community challenge designed to benchmark both predictive and generative models in protein engineering. The Tournament will center on improving the functional design of **PETase**, an enzyme with critical applications in plastic degradation.

Through its emphasis on open collaboration, standardized measurements, and transparent data practices, Align is pioneering resources for the next generation of AI-driven protein science. Together, GROQ-seq and The Tournament exemplify Align's vision for a future in which biological discovery is accelerated by open data, community challenges, and collaborative innovation. The Align Foundation invites researchers to contribute new protein functions, develop modeling approaches, and participate in collaborative benchmarking efforts that drive forward the sequence-to-function frontier.

Tuesday, Oct 21

Session 4: Synthetic Biology

KEYNOTE LECTURE

Modular protein design for protein cage assembly and mammalian cell regulation

Roman Jerala

National Institute of Chemistry, Department of Synthetic Biology and Immunology, Ljubljana, Slovenia

roman.jerala@ki.si

Modular protein design enables the design of novel protein folds and the regulation of biological processes in mammalian cells. Coiled-coil dimers represent a module, frequently used in natural proteins, that could be designed de novo and repurposed for diverse tasks. Coiled-coil protein origami (CCPO) design enables the construction of protein assemblies unknown in nature, such as polyhedral scaffolds, based on abstract mathematical principles. Extension to larger CC oligomers increases the design space. We determined the structures of CCPO-based trigon and tetrahedron, along with chemically regulated assembly and preorganization strategies such as protein cyclization. Modular and tunable coiled-coil interactions enable the formation of liquid-liquid protein condensates from single or multiple polypeptide chains. In mammalian cells, coiled-coil dimers were combined with split proteases to allow rapid control of cellular processes, including protein secretion and membrane localization. The INSRTR system introduces a coiled-coil peptide into target proteins, disrupting function upon heterodimer formation. This approach enables ON/OFF protein switches, small-molecule regulation, and logic functions with fast cellular responses. INSRTR has been demonstrated in ten different proteins, including enzymes, signaling mediators, transcriptional regulators, fluorescent proteins, and antibodies. It has also been integrated into chimeric antigen receptors on T cells, highlighting its therapeutic potential. Overall, designable coiled-coil dimers offer powerful tools for protein engineering, synthetic biology, and biomedicine, enabling precise control of biological systems.

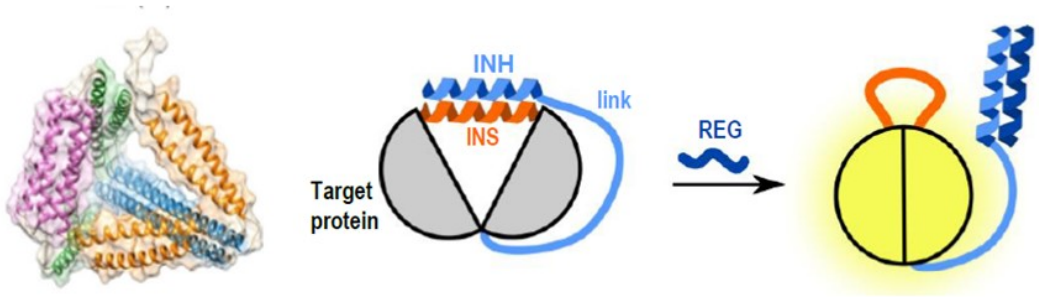


Figure 1: Coiled-coil modules for protein cage design and regulation. *Concatenation of coiled coil modules in a defined order (CCPO) facilitates selfassembly into a defined protein cage (left). Introduction of coiled-coil modules into the functional proteins represents a mechanism of allosteric regulation (INSRTR) of their function and cellular processes (right).*

Design and Engineering of Artificial Membrane Cytochromes for Synthetic Bioenergetics

Benjamin J. Hardy[1], Laila-Marie Latin Vicente[1], Ross Anderson[1], Paul Curnow[1]

[1] University of Bristol, School of Biochemistry, Bristol, UK
ben.hardy@bristol.ac.uk

The *de novo* design of bioenergetic proteins provides insight into the underlying principles of protein folding and assembly, permits the study of electron transport in minimal biomimetic systems, and produces new components for synthetic biology. Natural bioenergetic proteins function within the membrane and bind redox-active cofactors to perform electron transport (ET) reactions. Many respiratory and photosynthetic complexes share an ancestral four-helix bundle architecture at their core that positions two b-type hemes for transmembrane (TM) electron transport. This fundamental scaffold is an ideal basis for creating man-made bioenergetic membrane proteins. Our core module based on this fold is cytochrome bX (CytbX), a de novo di-heme membrane protein capable of transmembrane ET [1]. CytbX is compatible with multiple microbial hosts and binds both heme cofactors *in vivo*. We now seek to incorporate CytbX (as a minimal 'wire') into larger complexes with diversified electron transport activities. To do this, we must first establish robust design principles and methodologies. I will share recent progress towards various elements of *de novo* bioenergetic complex design, including the design of homo-oligomeric CytbX assemblies through lateral association in the membrane, photosensitization of CytbX for light-activated electron transport, incorporation of CytbX into proteoliposomes, fusion of soluble redox domains for NADPH-powered electron transport, and alteration of the heme chain length (Figure 1). With these and further developments, the field of synthetic bioenergetics will unlock user-defined control of energy flow within living cells, with far reaching implications for fundamental science and biotechnology.

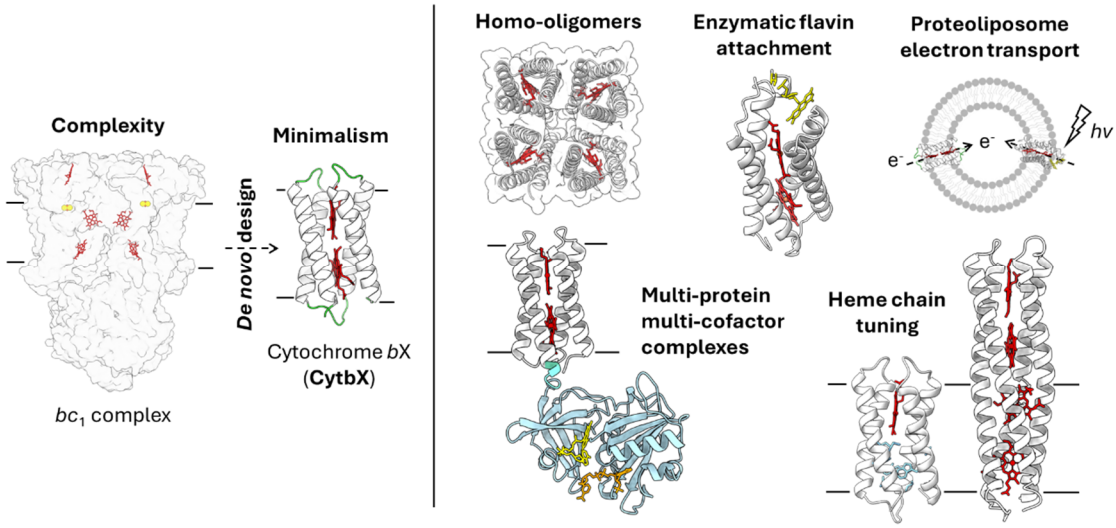


Figure 1. Membrane protein design for synthetic bioenergetics. (left) The design of CytbX is inspired by the core diheme four-helix bundle from the bc_1 complex. (right) Efforts towards building larger bioenergetic complexes incorporating CytbX.

References:

[1]. Hardy BJ, et al. (2023) Cellular production of a de novo membrane cytochrome. Proc. Natl. Acad. Sci. 120:2022.12.06.519282.

Towards meta-stable de novo protein assemblies

Miguel Atienza [1,2], Alena Khmelinskaia[1,2]

[1] Department of Physical Chemistry, LMU, Munich, Germany
[2] Center for NanoScience (CeNS), LMU, Munich, Germany

miguel.atienza@cup.lmu.de

Self-assembling polyhedral protein assemblies offer powerful platforms for engineering the targeted encapsulation and transport of biomolecules. While recent de novo designs have achieved remarkable structural precision, their extreme stability limits their utility in dynamic applications such as controlled cargo delivery. Functional protein nanoparticles will require not only precise structure but also tunable stability and assembly dynamics.

I will present our ongoing efforts to design meta-stable protein assemblies, deliberately moving away from the prevailing paradigm of maximizing thermodynamic stability. I will share initial results from our computational design pipeline, which combines ProteinMPNN with Rosetta-based scoring to explore variants with reduced stability and potentially dynamic behavior. Although the early designs highlight the difficulty of tuning stability without compromising assembly, they also provide valuable insights for guiding future iterations.

Additionally, I will present the first experimental results from our developing pipeline of a scalable pipeline to systematically characterize nanoparticle assembly. Our approach combines FRET and subunit mixing/exchange experiments to determine apparent constants and assembly kinetics. Ultimately, we aim to generate a consistent, high-quality dataset that can help bridge the current gap in modeling and predicting assembly stability.

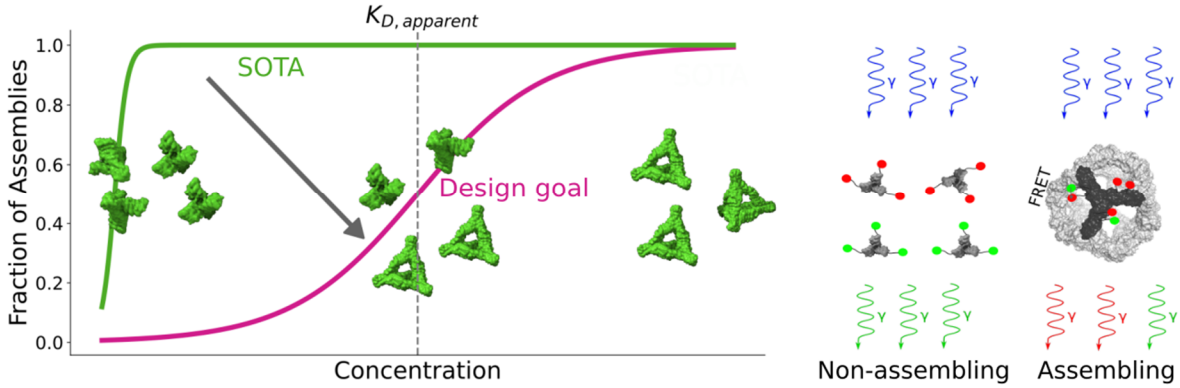


Figure 1: Towards meta-stable de novo protein assemblies. Left: Schematic of an apparent binding curve ($K_{D,\,apparent}$). Right: Schematic of the FRET-based subunit exchange assay used to probe assembly dynamics.

References:

[1] J. Dauparas et al. ,Robust deep learning–based protein sequence design using ProteinMPNN.Science378,49-56(2022).DOI:10.1126/science.add2187
[2] Wargacki, A.J., Wörner, T.P., van de Waterbeemd, M. et al. Complete and cooperative in vitro assembly of computationally designed self-assembling protein nanomaterials. Nat Commun 12, 883 (2021). https://doi.org/10.1038/s41467-021-21251-y
[3] Bhoomika Basu Mallik∗, Jenna Stanislaw∗, Tharindu Madhusankha Alawathurage∗, Alena Khmelinskaia. De novo design of polyhedral protein assemblies: before and after the AI revolution. ChemBioChem, doi: 10.1002/cbic.202300117

Unexpected modes of cooperativity in the enzyme glutathione reductase

Alexandra Dashevsky[1], Anna Vanyushkina[1], Michal Sharon[1], Sarel J. Fleishman[1] and Amnon Horovitz[2]

Departments of Biomolecular Sciences[1] and Chemical and Structural Biology[2], Weizmann Institute of Science, Rehovot 7610001, Israel

Alexandra.Dashevsky@weizmann.ac.il

Binding cooperativity is a fundamental property of many multimeric proteins, yet the ability to design and control it remains limited. My research focuses on understanding the mechanisms underlying cooperativity in homomultimers with the goal of developing nature-inspired approaches for design of cooperativity. As a model system, I focused on a mutant of the glutathione reductase (GR) enzyme from *E. coli*, due to an unusually high reported Hill coefficient for this system (n=1.96) [1,2].

I used native mass spectrometry to analyse cofactor and substrate binding cooperativity. Surprisingly, I found that this mutant of GR is subject to two distinct and overlapping mechanisms that contribute synergistically to cooperativity, explaining its very high cooperativity. To integrate these findings, we developed a novel mathematical framework that accounts for two distinct cooperativity mechanisms, demonstrating that it may open the way to dramatic levels of cooperativity that have not been previously reported. For example, under some parameter regimes, the Hill coefficient of a homodimer subject to two allosteric mechanisms may exceed 2 — a property that, to the best of our knowledge, has not been suggested in decades of intense study of allostery.

I am now using these insights to develop a new strategy for modulating allostery and cooperativity in a homodimer. The intended outcome is to generate subtle backbone motions that are restrained upon cofactor binding, thereby coupling protein stability and activity directly to cofactor occupancy.

Collectively, these efforts advance both our mechanistic understanding of cooperativity and our ability to computationally design it in multimeric proteins.

References:

[1] Scrutton NS, Deonarain MP, Berry A, Perham RN. Cooperativity induced by a single mutation at the subunit interface of a dimeric enzyme: glutathione reductase. Science. 1992;258: 1140–1143.

[2] A Bashir, R N Perham, N S Scrutton, A Berry; Altering kinetic mechanism and enzyme stability by mutagenesis of the dimer interface of glutathione reductase. *Biochem J* 1 December 1995; 312 (2): 527–533.

INVITED TALK

Stepwise design of multifunctional proteins and assemblies

Florian Pretorious[1]

[1]Institute of Science and Technology Austria

In the age of deep learning and generative models it has become somewhat straightforward to design large proteins and assemblies with arbitrary shapes. Even the design of protein-protein interfaces works with astonishing success rates, as long as we define "success" mostly as "any type of interaction with the target". Specifically designing other aspects of protein-binding proteins, such as affinity, rate constants, off-target interactions/specificity remains challenging and often is easier achieved by (wetlab-) screening than by explicit design. When building large assemblies from proteins (and potentially non-protein components) it might thus be more successful to combine multiple pre-validated and characterized interfaces into single protein chains. In this presentation, I will talk about some of the building blocks my group is designing, as well as some of our attempts to combine these building blocks.

Designing for exclusivity: Can we engineer orthogonal heterodimers *de novo* using sequence-based design approaches?

Leoni Abendstein[1], Florian Praetorius[1]

[1]Institute of Science and Technology (ISTA), 3400 Klosterneuburg, Austria

Leoni.abendstein@ista.ac.at

Orthogonal heterodimers - protein pairs that bind exclusively to their intended partners without cross-reactivity to other heterodimers sharing the same backbone but differing in sequence - are commonly found in natural systems. These natural systems often exhibit high specificity and conserved interface features, raising the question: can such **sequence-specific orthogonality** be achieved synthetically, using *de novo* protein design?

Previously, as variety of heterodimers with distinct backbone geometries were designed. Here, we use these as starting templates to investigate whether orthogonal interaction specificity can be encoded at the sequence level. We employ a multi-step pipeline combining **PyRosetta** for interface selection, **ProteinMPNN** for sequence design, and **AlphaFold2** for structure prediction and interaction validation. Following redesign, heterodimeric pairs were computationally screened for folding stability, interface complementarity, and specificity toward their intended partner.

To evaluate orthogonality, candidate sets were also screened for *negative interactions*: only those where "off-target" heterodimer pairs showed inter-chain PAE values above 20, and designed pairs showed inter-chain PAE values below 6, were considered for ordering. Selected designs were then expressed and characterized to assess binding specificity and orthogonality using biochemical assays.

This integrated ***in silico–in vitro*** approach allows us to systematically probe the sequence determinants of orthogonality and explore design principles that promote exclusivity within a given set of interface sequences. Our findings offer insights into the sequence–structure–function relationship underlying orthogonal protein interactions and inform future strategies for building modular, programmable systems in synthetic biology. This work also contributes toward the broader goal of constructing large-scale, composable interaction networks where precise molecular pairing is essential.

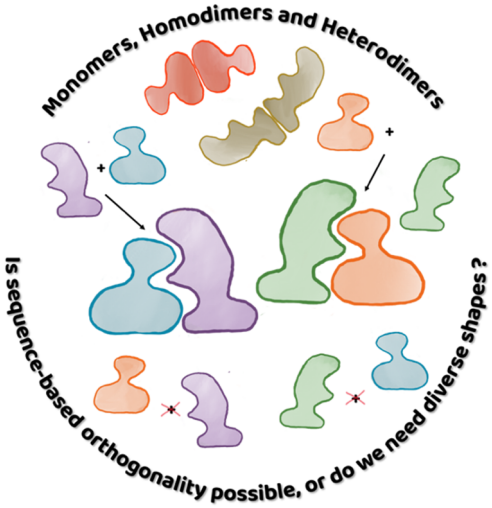


Figure 1: Schematic overview illustrating monomers, homodimers, heterodimers and orthogonality.

Session 5: Antibodies

INVITED TALK

Can we extract physics-like potentials from diffusion models?

Jeffrey J. Gray

Professor of Chemical and Biomolecular Engineering, Johns Hopkins University

Although our community collectively refined energy functions for the last 25 years (50 if you include the MD community!), somehow AlphaFold is able to identify much more accurate structures. Why is that? Does AlphaFold and related methods encapsulate, in some way, a function that has minima at native conformations? If so, could you extract that function? I will briefly describe several approaches in my lab that strive to pull physics out of AI or put physics into AI. A focus case will be on denoising diffusion models, where the score can be equated to the gradient of a learned potential. We explore multiple formulations to extract energy of a generated structure from its diffusion or flow trajectory. Preliminary explorations with our docking model DFMDock show that the learned energy function funnels reflect the docking accuracy of the model, but in this case, funnels are not as reliable or deep as with Rosetta.

Computational design of monobody binders to target viral epitopes

Pedro Moreira[1,2], Shuhao Xiao[3], Wenjun Gao[4], Joseph Schmidt[3], Mariana Parada[1], Carlota J.B. Conceição[1], Madalena C. Marques[1], Diogo H. Silva[1], Carolina C. Buga[1,5,7], Christine Cruz-Oliveira[5,7], Marta Miranda[5,7], Beatriz Vieira-da-Silva[1,5,7], Sandrine Georgeon[3], Marta Alenquer[6], Maria João Amorim[6], Manuel N. Melo[1], Miguel Rocha[2], Ana S. Veiga[5,7], Miguel A.R.B. Castanho[5,7], Isabel Abreu[1], João B. Vicente[1], Ye Wei[3,4], Cláudio M. Soares[1], Bruno E. Correia[3], Diana Lousa[1]

[1] ITQB NOVA, Oeiras, Portugal. [2] Center of Biological Engeneering, School of Engineering, University of Minho, Braga, Portugal. [3] School of Life Sciences, EPFL, Lausanne, Switzerland. [4] City University of Hong Kong, Hong Kong, China. [5] GIMM, Lisboa, Portugal. [6] Católica Biomedical Research Centre, Lisboa, Portugal. [7] Lisbon School of Medicine, University of Lisboa, Lisboa, Portugal.

pmoreira@itqb.unl.pt

As demonstrated by the SARS-CoV-2 pandemic, developing a platform for a rapid and reliable discovery of antiviral therapeutics is of critical importance. For this purpose, monobodies (Mobs) have shown particular promise. These synthetic, antibody-like proteins are derived from the Human Fibronectin Type III domain. Mobs are well-suited for bacterial expression and are thought to be less immunogenic than *de novo*-designed mini binders due to their human origin. They also have shown potential as therapeutic agents against cancer and viral targets. However, the process of identifying high-affinity Mob binders largely depends on screening large *in vitro* libraries, which is costly and time-consuming. Similar to antibody or VHH computational design, Mob binder design is challenging due to the unstructured nature of the CDR-equivalent loops.

In this work, multiple strategies have been developed to establish a computational platform for generating Mob binder sequences with high affinity for specific viral protein targets. The first strategy envisioned a generative pipeline incorporating RFdiffusion[1], AFdesign[2], and ProteinMPNN[3] to create Mob binder sequences. Mob designs selected based on the most promising AlphaFold2[4] metrics were tested experimentally. The second strategy relies on generating an *in silico* library of Mob loop sequences using ESM3[5], so to introduce paratope diversity. Protein complex models were then generated using AlphaFold2 co-folding of the Mob and target sequences. The most confident predictions were refined using a Monte Carlo tree optimization algorithm[6], which applies iterative rounds of ProteinMPNN mutations to improve the confidence of protein complex predictions.

As a proof-of-concept, we applied both strategies to generate Mob binders against the SARS-CoV-2 receptor binding domain. Several Mobs from each protocol exhibited significant affinity, with some displaying K_D below 100 nM and a few demonstrating viral neutralization activity. These results highlight the potential of both computational frameworks for the rapid development of antibody-like binders targeting viral epitopes.

References:

[1] Watson, J.L., et al. *Nature* 620.7976 (2023): 1089-1100.
[2] Frank, C., et al. *Science* 386.6720 (2024): 439-445.
[3] Dauparas, J., et al. *Science* 378.6615 (2022): 49-56.
[4] Jumper, J., et al. *Nature* 596.7873 (2021): 583-589.
[5] Hayes, T., et al. *Science* 387.6736 (2025): 850-858.
[6] Wei, Y., et al. *arXiv preprint arXiv:2404.04062* (2024).

Epitope-specific antibody design using diffusion models on the latent space of ESM embeddings

Tomer Cohen[1], Dina Schneidman-Duhovny[1]

[1] The Hebrew University of Jerusalem, School of Computer Science and Engineering, Jerusalem, Israel
tomer.cohen13@mail.huji.ac.il

Significant progress has been made in computational protein design using deep learning approaches, with most methods predicting sequences for predefined protein backbone structures. Recently, diffusion-based generative methods have enabled the generation of protein backbones. However, de novo design of epitope-specific antibody binders remains an unsolved problem due to the challenge of simultaneous optimization of the antibody sequence, variable loop conformations, and antigen-binding specificity.

Here, we introduce **EAGLE (Epitope-specific Antibody Generation using Language-model Embeddings)**, a diffusion-based deep-learning model that designs full-length antibody sequences (both constant and variable regions) entirely in continuous embedding space, without requiring predefined backbone structures. Analogous to image-generation models conditioned on textual prompts, EAGLE conditions the antibody sequence generation directly on the antigen structure and epitope residues.

We demonstrate the practical utility of our method with a comprehensive pipeline: for a given antigen target, EAGLE generates over 100,000 candidate antibody sequences, which are then structurally modeled using AlphaFold3 (AF3) to prioritize designs exhibiting high structural confidence (ipTM scores). Applying this pipeline to the therapeutically relevant antigen FGFR4, we experimentally validated 31 computationally prioritized nanobody candidates. Of these, two sequences were confirmed experimentally to specifically bind FGFR4, validating our computational strategy.

We are now employing computational affinity maturation strategies to further enhance binding affinity and optimize these novel antibody candidates. Overall, EAGLE establishes a robust and experimentally validated approach for rapidly designing diverse, epitope-specific antibody binders using straightforward epitope-level information.

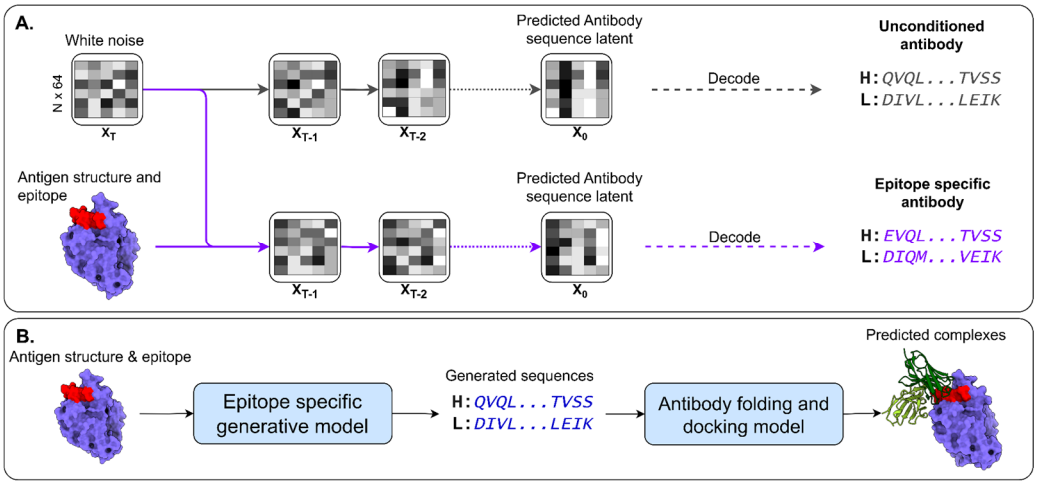


Figure 1: A. *Sampling antibody sequences without (top) and with epitope conditioning (bottom).* ***B.*** *A pipeline for epitope-specific sequence and structure generation.*

Multi-Objective Active Learning for Nanobody Development

Klara Kropivsek[1], Katharina Dost[2], Christian Leonardo Camacho-Villalón[2], Saso Dzeroski[2], Ario de Marco[1]

[1] Laboratory for Environmental and Life Sciences, University of Nova Gorica, Nova Gorica, Slovenia
[2] Department of Knowledge Technologies, Jozef Stefan Institute, Ljubljana, Slovenia

klara.kropivsek@ung.si

Nanobodies—compact, single-domain antibody fragments—are increasingly used in therapeutics and diagnostics due to their high specificity and stability. However, optimizing multiple properties such as expression yield and binding affinity remains experimentally costly. While machine learning can accelerate candidate selection, its effectiveness depends on the quality and diversity of labeled data. Standard active learning (AL) approaches address this by prioritizing informative samples, but typically ignore practical constraints critical to nanobody development.

We present a multi-objective active learning (MOAL) framework tailored to nanobody discovery. The framework integrates predictive models for binding affinity and expression yield with uncertainty estimation from ensemble learning. Candidate selection is guided by three objectives: informativeness (model improvement), feasibility (predicted expression), and performance (binding affinity). To balance trade-offs among these objectives, we apply evolutionary multi-objective optimization algorithms, specifically NSGA-II and IBEA. This enables exploration of diverse, high-potential regions of nanobody sequence space.

We evaluate our framework on a curated dataset of characterized nanobody sequences and a large-scale nanobody repertoire comprising over 10 million candidates. The curated data enable supervised learning, while the repertoire supports broad exploration. Our approach identifies nanobody candidates that are both experimentally viable and model-informative, improving generalization while reducing experimental costs. By avoiding redundant queries and favoring biologically diverse selections, the method supports efficient discovery.

Our domain-aware MOAL approach provides an effective strategy for guiding nanobody selection under multiple constraints. It enables iterative refinement of predictive models while maintaining experimental feasibility. Though developed for nanobody engineering, the framework generalizes to other biological domains requiring data-efficient, multi-objective decision-making.

Evaluating the Biological Realism of IgLM-Generated Antibody Sequences

Natalia Rincon[1], Donovan Vincent Jr.[2], Sravani Venkatayogi[3], Kevin Wiehe[3], Jeffrey J. Gray[2]

[1] Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD, USA

[2] Department of Chemical and Biomolecular Engineering, Johns Hopkins University, Baltimore, MD, USA

[3]Duke Human Vaccine Institute, Duke School of Medicine, Durham, NC, USA

nrincon1@jhmi.edu, dvincen9@jh.edu, sravani.venkatayogi@duke.edu, kevin.wiehe@duke.edu, jgray@jhu.edu

In this work, we evaluate the realism of antibody sequences generated by the Immunoglobulin Language Model (IgLM)[1]. We focus on their alignment with known immunological mechanisms such as V(D)J recombination, somatic hypermutation (SHM) and CDRH3 diversification. For each mechanism, we compare natural and synthetically generated sequences, and evaluate IgLM likelihoods against those from mechanistic models, to identify the limitations of LM–generated sequences. To assess V(D)J recombination fidelity, we compare gene usage distributions (ANARCI[2]) and CDRH3 likelihoods (OLGA[3]). In an initial comparison, we find that while IgLM replicates V gene distributions well, its D and J gene usage diverges from the natural repertoire. This may stem from the autoregressive nature of the model, which loses coherence toward the end of a sequence. To evaluate how well IgLM captures SHM, we compare per-residue likelihoods to ARMADiLLO[4] mutation probabilities. Although IgLM-generated sequences exhibit fewer mutations than natural antibodies, common limitation of most antibody language models (AbLMs), comparisons with ARMADiLLO and IMGT reference sequences indicate partial learning of V gene family features. Our final analysis examines the CDRH3 region, where accurate evaluation requires nucleotide-level analysis because the D gene is short and highly sensitive to splicing. Most AbLMs operate at the amino acid level, so we employ three reverse-translation methods to reconstruct nucleotide sequences: codon table–based reverse translation (EMBOSS[5]), Ig-specific reverse translation using ARMADiLLO-derived statistics, and BANANA, a nucleotide-level AbLM. We then compare CDR3s generated from randomly sampled amino acids (uniform 1/20 probability per residue) with those from natural and synthetic repertoires to assess biological realism. These findings reveal both the strengths and limitations of IgLM in capturing immunological patterns, offering guidance for improving AbLMs and designing evaluation metrics that better reflect biological realism. While wet-lab validation remains essential, our analysis can help explain model failures and highlight missing biological mechanisms.

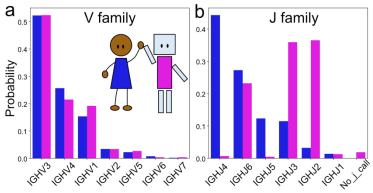


Figure 1: Gene usage differs between natural (blue) and synthetic (magenta) repertoires. Preliminary gene usage analysis shows a) concordant V gene usage but b) discordant J gene usage.

References:

[1] R. W. Shuai, J. A. Ruffolo, and J. J. Gray, "IgLM: Infilling language modeling for antibody sequence design," *Cell Syst*, vol. 14, no. 11, pp. 979-989.e4, Nov. 2023.

[2] J. Dunbar and C. M. Deane, "ANARCI: antigen receptor numbering and receptor classification," *Bioinformatics*, vol. 32, no. 2, pp. 298–300, Jan. 2016.

[3] Z. Sethna, Y. Elhanati, C. G. Callan, A. M. Walczak, and T. Mora, "OLGA: fast computation of generation probabilities of B- and T-cell receptor amino acid sequences and motifs," *Bioinformatics*, vol. 35, no. 17, pp. 2974–2981, Sep. 2019.

[4] K. Wiehe *et al.*, "Functional relevance of improbable antibody mutations for HIV broadly neutralizing antibody development," *Cell Host Microbe*, vol. 23, no. 6, pp. 759-765.e6, Jun. 2018.

[5] F. Madeira *et al.*, "The EMBL-EBI Job Dispatcher sequence analysis tools framework in 2024," *Nucleic Acids Res.*, vol. 52, no. W1, pp. W521–W525, Jul. 2024.

Analysis of Pareto optimality at the antibody-antigen interface

Tadej Medved[1,2], Goran Miličić[1], San Hadži[2]

[1] Biologics Drug Product Development, Novartis LLC, Slovenia
[2] Faculty of Chemistry and Chemical Technology, University of Ljubljana, Slovenia

tadej.medved@fkkt.uni-lj.si

A constant challenge in biologics drug product development is the multi-objective optimization of a molecule's various developability properties, which should not compromise its function, such as antigen binding in the case of antibody-based therapeutics [1]. The binding process can be considered as a trade-off between two underlying thermodynamic contributions in the form of enthalpy and entropy; however, the structural characteristics of the binding interface that define these contributions are still not completely understood. Understanding the structural features that define the thermodynamic driving force of high-affinity binding could help guide the optimization of developability properties without detrimental effect to an antibody's primary function. To this end, we expanded the Pareto Task Inference method [2].

Under the Pareto theory of optimality, entities that have withstood selective pressure by optimizing necessary trade-offs between intrinsic features are distributed inside low-dimensional polytopes (**"Pareto fronts"**) in the space defined by those features, and the vertices of these polytopes correspond to specialist "**archetypes**" [2]. We determined Pareto-optimal trade-offs between crystal structures of camelid single-chain antibodies (nanobodies) bound to their respective antigens and correlated those with the antigen-binding enthalpy and entropy as measured by ITC, while also assessing the distributions of structural features that could pose a higher developability risk. Our approach identified a statistically significant fit (p = 0.01) of a triangular Pareto front in the interface feature space (Fig. 1). Each of the putative archetypes was significantly enriched in a different thermodynamic property, namely the normalized binding enthalpy ΔH_b /ΔG_b (co-enriched with specific polar contacts and antigen 2° structure), the binding entropy ΔS_b (co-enriched with interface loop flexibility), and the theoretical desolvation entropy ΔS_{desolv} (co-enriched with hydrophobic interface patches and nonpolar interactions). ΔG_b was not enriched at any archetype, leading us to postulate that the observed Pareto front reflects three different strategies of achieving high binding affinity, though they differ in other aspects of developability, such as hydrophobic surfaces. Such a model could see potential use in candidate selection or redesign, serving as both a cutoff for binding-optimal interface features and as a guide for selection of candidates closer to an archetype with better developability criteria in order to reduce the attrition rate.

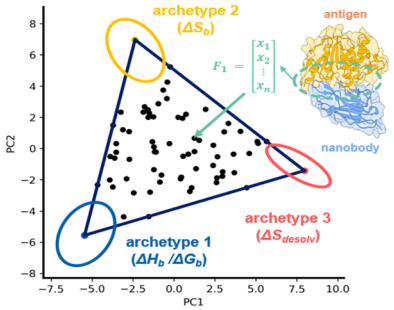


Figure 1: Pareto front in principal component space of nanobody interface features. *Datapoints (in black) correspond to projections of individual nanobody-antigen interfaces based on interface features derived from their crystal structures. The vertices correspond to the archetypes of the Pareto front, with the ellipses depicting their positional errors estimated by bootstrapping (95 % CI).*

References:

[1] W. Zhang, et al. Developability assessment at early-stage discovery to enable development of antibody-derived therapeutics. *Antibody therapeutics*, 6(1):13-29, 2023

[2] Y. Hart, et al. Inferring biological tasks using Pareto analysis of high-dimensional data. *Nature Methods*, 12(3): 233-235, 2015

Session 6: Non-canonicals & AI

INVITED TALK

RareFold: Structure prediction and design of proteins with noncanonical amino acids

Patrick Bryant[1]

[1] Science for Life Laboratory, The Department of Molecular Biosciences, The Wenner-Gren Institute, Stockholm University, Solna 171 65, Sweden
Patrick.bryant@scilifelab.se

Protein structure prediction and design have historically been constrained to the 20 canonical amino acids, limiting the functional and chemical diversity of designed biomolecules. Expanding this space to include non-canonical amino acids (NCAAs) provides new opportunities to access novel interaction geometries, enhance proteolytic stability, and reduce immunogenicity, critical features for therapeutic applications. We present **RareFold**, a deep learning framework capable of accurate structure prediction for proteins and peptides incorporating both canonical residues and 29 additional NCAAs. RareFold tokenises each amino acid as a chemically unique entity, enabling the model to learn residue-specific atomic interactions across a vastly expanded sequence space. This representation supports both forward prediction and inverse design. Building on RareFold, we introduce **EvoBindRare**, a design framework for generating linear and cyclic peptide binders with NCAAs. As a proof of concept, we designed and experimentally validated binders targeting a human ribonuclease. Both linear and cyclic formats achieved micromolar binding affinity similar to the WT binder, demonstrating the practical utility of this expanded design space. RareFold represents a step-change in peptide design, unlocking access to novel chemistries and binding modes inaccessible to standard tools. It enables the creation of next-generation peptide therapeutics with enhanced properties and opens new avenues for functional design unconstrained by canonical biochemistry.

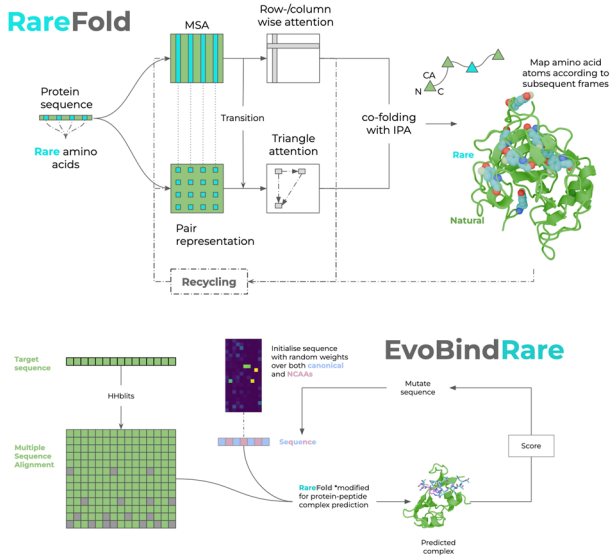


Figure 1. Structure prediction and design with RareFold

RareFold *is a structure prediction network based on the EvoFormer architecture from AlphaFold2, trained to model both the 20 canonical amino acids and 29 additional noncanonical amino acids (NCAAs).*

EvoBindRare *is a peptide binder design framework based a modified RareFold model for protein-peptide complex prediction. The predicted complex is scored, a mutation is introduced, and the process is repeated. EBR performs joint structure and sequence optimisation over 1000 mutation steps to design both linear or cyclic binders of 10-20 residues using up to 49 amino acid types.*

All-atom protein design via SE(3) flow-matching with ProteinZen

Alex J. Li[1,2], Tanja Kortemme[1,2]

[1] UC Berkeley-UCSF Bioengineering PhD Program, CA, USA
[2] UCSF, CA, USA

alexjli@berkeley.edu

De novo protein design has been greatly accelerated by the advent of generative models of protein structure. While more coarse-grain tasks such as backbone generation [1] and binder design [2] are increasingly possible, existing models still struggle on problems that require precise placements at the atomic scale, motivating the development of all-atom generative models. However, all-atom generative modeling remains challenging due to the complex interplay between discrete and continuous aspects of protein structure. In this work we propose a framework to capture this interplay by decomposing residues into collections of rigid chemical groups, allowing us to utilize SE(3) flow-matching [3][4] for all-atom protein structure generation. Our method, ProteinZen, generates designs with high sequence-structure consistency while retaining competitive diversity and novelty on both unconditional and conditional generation tasks. We demonstrate strong performance for both unconditional monomer design as well as various forms of motif scaffolding, including full-atom motif scaffolding, sidechain atom scaffolding, and motif scaffolding without specifying motif segment spacing or relative primary sequence order.

References:

[1] J. L. Watson, D. Juergens, N. R. Bennett, et al. De novo design of protein structure and function with RFDiffusion. Nature, 620(7976):1089–1100, July 2023.
[2] M. Pacesa, L. Nickel, C. Schellhaas, et al. Bindcraft: one-shot design of functional protein binders. *bioRxiv preprint bioRxiv:2024.09.30.615802*, 2024.
[3] J. Yim, A. Campbell, A. Y. K. Foong, et al. Fast protein backbone generation with SE(3) flow matching. *arXiv preprint arXiv:2310.05297*, 2023.
[4] J. Bose, T. Akhound-Sadegh, G. Huguet, et al. SE(3)-stochastic flow matching for protein backbone generation. In *The Twelfth International Conference on Learning Representations* (ICLR), 2024.

Prediction of Amino Acid Dispensability in Model Organisms

Tereza Neuwirthova[1], Jakub Mišťúrik[1] & Klára Hlouchová[1,2]

[1]Department of Cell Biology, Charles University, Prague, Czech Republic

[2]Institute of Organic Chemistry and Biochemistry, Czech Academy of Sciences, Prague, Czech Republic

tereza.neuwirthova@natur.cuni.cz

Does current life need all 20 amino acids? Let's assume that the answer is no. We think that organisms with fewer amino acids likely existed in the past and life's protein alphabet evolved to its current 20-amino-acid form only shortly before the Last Universal Common Ancestor (LUCA). Building on this assumption, we ask two more questions. Which current amino acid would be easiest to replace? What types of replacements would work best? These questions are important for two reasons. First, we don't understand why all life uses exactly the same 20 amino acids despite the wild variety of life forms. Second, the answers could guide future protein design and artificial life projects. We addressed these questions using computational methods with two different approaches: physics-based and phylogeny-based. The physics-based approach uses Rosetta molecular modeling suite calculations of mutations favorability [1]. The phylogeny-based approach uses an epistatic model called GEMME predicting mutational effects [2]. We selected two model organisms: E. coli and Mycoplasma JCVI-Syn3a. For each organism, we predicted the most favorable replacements for all occurrences of selected amino acids in the proteome. We then calculated an overall dispensability score for each selected amino acid.

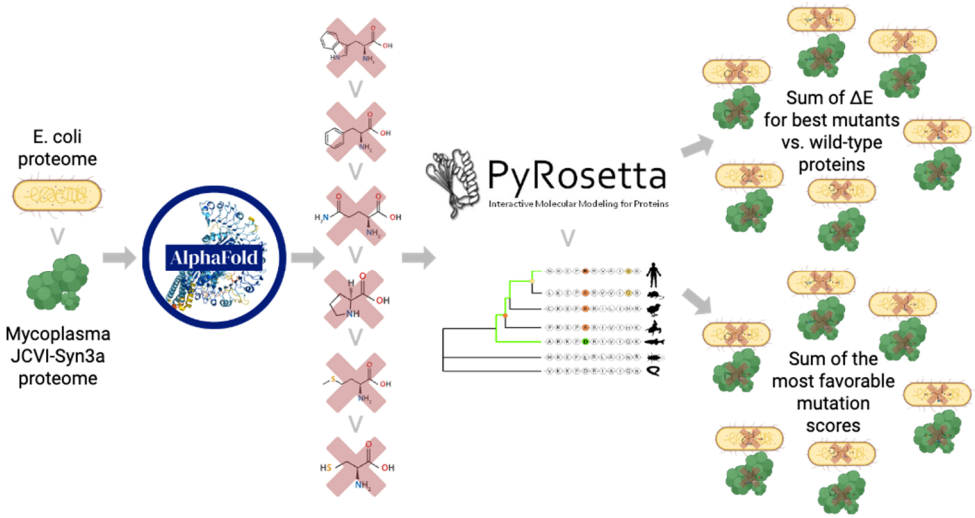


Figure 3: Workflow of amino acid dispensability simulation *AlphaFold predictions of model organism's proteomes were taken as an input and energy change or mutations scores were calculated with PyRosetta or GEMME, respectively. As a last step, overall dispensability score for each selected amino acid was calculated. Meaning of the v symbol in the diagram is logical operator OR.*

References:

[1] Thieker et al., *Protein Science*, 31:e4428, 2022

[2] Laine et al., GEMME: A Simple and Fast Global Epistatic Model Predicting Mutational Effects. *Mol. Biol. Evol*, 36(11):2604–2619, 2019

Amino Acid Alphabet (R)Evolution: Changing the Molecular Basis of Life

Tadeáš Kalvoda[1], Sean Brown[2], Robin Kryštůfek[3], Ján Michael Kormaník[3], Mikhail Makarov[1], Erik Andris[3], Jan Konvalinka[3], Václav Veverka[3], Lubomír Rulíšek[3], Stephen Freeland[2], Klára Hlouchová[1]

[1]Department of Cell Biology, Faculty of Science, Charles University, Prague, 128 00, Czech Republic
[2]Department of Biological Sciences, University of Maryland Baltimore County, Baltimore, MD 21250, USA
[3]Institute of Organic Chemistry and Biochemistry of the Czech Academy of Sciences, Czech Republic

kalvodat@natur.cuni.cz

Life as we know it constructs protein-based metabolism using one, genetically encoded "alphabet" of 20 amino acids. This alphabet exhibits an unusual physicochemical profile: Its 20 amino acids are distributed very evenly over a broad range of volume and a narrow range of hydrophobicity. This is consistent with natural selection of a set of monomeric units which are capable to construct diverse, stable protein structures.

Our goal is to design an alternative (xeno) amino acid alphabet that shares a similar biophysical properties as canonical amino acids and is capable to form peptides with defined secondary structure. By the combination of heuristic search, quantum mechanical calculations with implicit solvation models, PCA, and biochemical principles, we designed two xeno amino acid alphabets to mimic the canonical early 10 amino acids.

Using these xeno amino acids, we synthesized a mixture of random peptides of 25 residues for each xeno alphabet. Experiments on these revealed that there was indeed a fraction that formed secondary structure. We then designed and synthesized specific sequences to describe this phenomenon in more detail. At least one of these peptides contain α-helical elements according to the NMR spectroscopy, thus confirming the potential of xeno amino acids to form structured peptides.

Such a result represents the first ever study of proteins with secondary structure composed of amino acids that have never been seen in proteins in Nature. Consequently, it may serve as proof of concept that life can be built in an alternative way, using a potentially much larger and more diverse pools of amino acids.

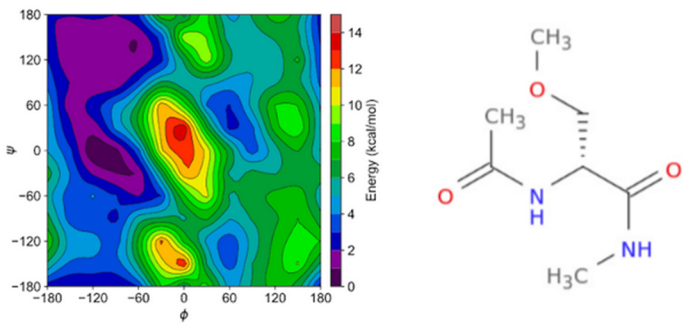


Figure 4: An example of potential energy surface of a xeno amino acid. *The potential energy surface as function of backbone dihedral angles (left) of the C- and N-termini capped xeno amino acid (right) was obtained via quantum mechanical (DFT) calculations.*

Combinatorial assembly and design of novel enzyme specificities

Tal Levin[1], Hermen S. Overkleeft[2], Gideon J. Davies[3], Sarel J. Fleishman[1]

[1]Department of Biomolecular Sciences, Weizmann Institute of Science, Rehovot, Israel
[2]Leiden Institute of Chemistry, Leiden University, Leiden, Netherlands.
[3]Department of Chemistry, The University of York, York, UK.

tal.levin@weizmann.ac.il

Our lab has developed a new strategy for designing dramatically different enzyme active sites based on natural protein backbones. This atomistic design and ML-based strategy exerts unprecedented control over protein activity and has yielded thousands of functional glycoside hydrolases with 80-160 mutations from any natural one [1] and the first demonstration of completely designed, high-efficiency *de novo* enzymes [2]. Here, we develop a new strategy for co-folding modular assembly and design based on AlphaFold3. The new approach generates designed and substrate-docked enzyme models with the aim of exhibiting novel substrate specificities. We targeted two decorated xylans for which efficient glycoside hydrolases (GHs) are much sought after and computationally screened thousands of designed GH backbones to select the few that are predicted to react with these bulky substrates. Experiments with activity-based probes are ongoing. This research will provide a proof-of-concept that modular backbone assembly and design provides a path towards dramatically expanding the functional scope of natural enzyme families to substrates of interest for basic and applied research.

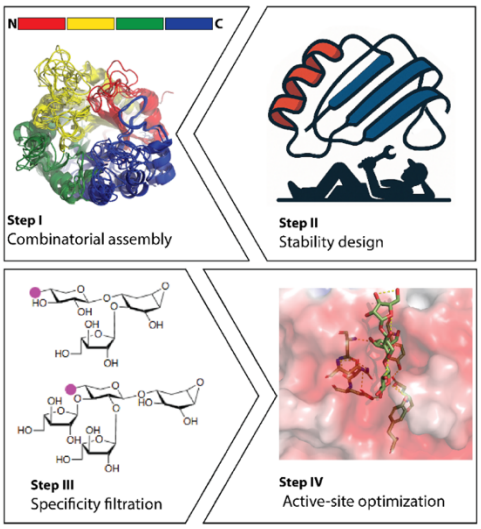


Figure 1: Overview of the combinatorial design pipeline. *Step 1: combinatorial assembly – fragmentation of a protein to several fragments followed by retrival of alternative structurally similar fragments. combinatorial assembly of the alternate fragments into novel assemblies Step 2: Modeling each protein assembly and applying stabilizing atomistic design tools [3]. Step 3: Filtering the assemblies based on statistical and atomistic filters following docking with the desired substrate [4]. Step 4: Further stabilization of the design active site.*

References:

[1] R. Lipsh-Sokolik et al. Combinatorial assembly and design of enzymes. Science, Vol 379, Issue 6628 Jan 12, 2023.

[2] D. Listov et al. Complete computational design of high-efficiency Kemp elimination enzymes. Nature, 643, June 18, 2025.

[3] A. Goldenzweig et al. Automated Structure- and Sequence-Based Design of Proteins for High Bacterial Expression and Stability. Molecular Cell, Volume 63, Issue 2 p337-346 July 21, 2016.

[4] T. L. R. Correa et al. A Chemoproteomic Biotechnological Toolkit for Resolving Xylanase Specificity in Decorated Xylan. ChemRxiv (2025)

Heme3D: A Heme Location Prediction Tool with 3dCNN/Docking Procedure

Qihao Zhang[1], Ute F. Röhrig[2], Simon L. Dürr[3], Ursula Röthlisberger[1]

[1] Laboratory of Computational Chemistry and Biochemistry, Institute of Chemical Sciences and Engineering, School of Basic Sciences, École polytechnique fédérale de Lausanne, Lausanne, Switzerland
[2] Swiss Institute of Bioinformatics, Lausanne, Switzerland
[3] Institute of Life Sciences, HES-SO Valais-Wallis (University of Applied Sciences of Western Switzerland), Sion, Switzerland

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 ligands as the input. In this work, we develop a 3dCNN-based method Heme3d combining with force-field based docking by Swissdock[1] for blind heme prediction. Given a single protein structure as input, Heme3d can generate an accurate full-atom heme location prediction. Now, in our test set of 1140 proteins, RMSD values in general are either <2.5 A (good vinyl positions) or in the range of 5.5-7 A (flip by 180 deg). This performance is comparable to that of AlphaFold 3 in predicting the ring position and full atom position, even though AlphaFold 3 was given the number of hemes.

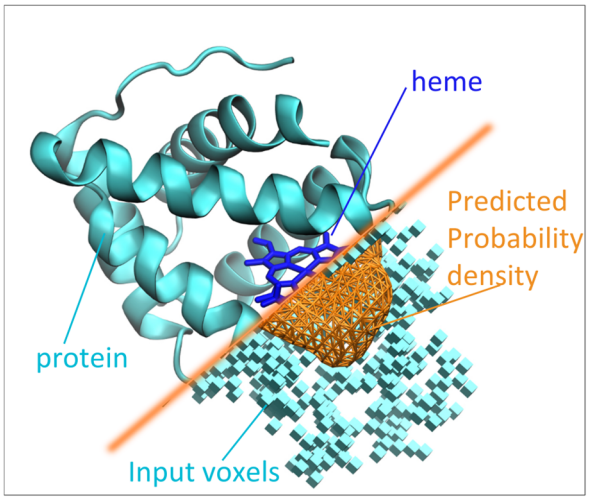


Figure 5: Scheme of Heme3D CNN workflow. *Input protein structures are voxelized and processed by a 3D convolutional neural network (3D-CNN), which outputs the predicted probability density of heme. This density is then transformed into predicted heme center positions, which serve as input for SwissDock to perform a traversal search and, if a suitable site is identified, provide the full atomic coordinates of the bound heme.*

References:

[1] M. Bugnon, et al. SwissDock 2024: major enhancements for small-molecule docking with Attracting Cavities and AutoDock Vina. *Nucleic Acids Research*, 52:324–332, 2024.

KEYNOTE LECTURE

Design of new protein functions using deep learning

David Baker[1,2,3]

[1]Department of Biochemistry, University of Washington, Seattle, WA, USA
[2]Institute for Protein Design, University of Washington, Seattle, WA, USA
[3]Howard Hughes Medical Institute, University of Washington, Seattle, WA, USA

dabaker@uw.edu

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

Wednesday, Oct 22

Session 7: Multistate proteins

KEYNOTE LECTURE

De novo design of programmable protein shapes, dynamics, and cellular functions

Tanja Kortemme[1,2]

[1] Department of Bioengineering & Therapeutic Sciences, University of California San Francisco (UCSF), San Francisco, CA, USA
[2] Quantitative Biosciences Institute, UCSF, San Francisco, CA, USA

tanjakortemme@gmail.com

Methods from artificial intelligence (AI) trained on large datasets of sequences and structures can now "write" proteins with new shapes and molecular functions de novo, without starting from proteins found in nature [1]. I will discuss our recent progress with developing deep learning models for de novo protein design [2, 3] and applying them to difficult problems, including control over complex protein shapes and atomically accurate design of dynamic proteins [4]. Emerging approaches incorporate engineering principles – tunability, controllability, modularity – into the design process from the beginning [1]. Exciting frontiers lie in constructing synthetic cellular signaling from the ground up using de novo proteins.

References:

[1] Kortemme T. De novo protein design - From new structures to programmable functions. *Cell* 187, 526-544, 2024.
[2] Akpinaroglu D, Seki K, Guo A, Zhu E, Kelly MJS & Kortemme T. Structure-conditioned masked language models for protein sequence design generalize beyond the native sequence space. *bioRxiv*, 2023.2012.2015.571823, 2023.
[3] Li AJ & Kortemme T. ProteinZen: combining latent and SE(3) flow matching for all-atom protein generation. Machine Learning in Structural Biology: Workshop at the 38th *NeurIPS Conference* Vancouver 2024. https://www.mlsb.io/papers_2024/ProteinZen:_combining_latent_and_SE(3)_flow_matching_for_all-atom_protein_generation.pdf, 2024.
[4] Guo AB, Akpinaroglu D, Stephens C, Grabe M, Smith CA, Kelly MJS & Kortemme T. Deep learning guided design of dynamic proteins. *Science* 388(6749):eadr7094. 2025.

Design of facilitated dissociation enables timing of cytokine signaling

Adam J. Broerman[1], Christoph Pollmann[2], Yang Zhao[3], K. Christopher Garcia[3], Florian Praetorius[1], Jacob Piehler[2], David Baker[1]

[1] Institute for Protein Design, University of Washington, Seattle, WA, USA
[2] Department of Biology/Chemistry, Osnabrück University, Osnabrück, Germany
[3] Department of Molecular and Cellular Physiology, Stanford University, CA, USA

broerman@uw.edu

Protein design has focused primarily on the design of ground states, ensuring they are sufficiently low energy to be highly populated. Designing the kinetics and dynamics of a system requires, in addition, the design of excited states that are traversed in transitions from one low-lying state to another [1]. This is a challenging task as such states must be sufficiently strained to be poorly populated, but not so strained that they are not populated at all, and because protein design methods have generally focused on creating near-ideal structures.

We developed a general approach for designing systems which use an induced-fit forcing motion [2] to generate a strained excited state, allosterically driving protein complex dissociation in a process termed "facilitated dissociation" [3]. **Kinetic binding measurements demonstrate that incorporating excited states enables design of effector-induced acceleration of dissociation as high as 6000-fold.** Crystal structures throughout the facilitated dissociation process demonstrate accurate design of such excited states. Within these excited states, we find that strain distributes nonuniformly depending on local topology, leading to kinetic asymmetry. Force spectroscopy and kinetic binding measurements with multiple effectors show that flexibility facilitates the force-generating conformational transition used to rapidly form the strained excited state.

We further investigated whether facilitated dissociation could enable faster-timescale control over cellular processes [4]. Many cytokines form tight complexes with their receptors which dissociate or degrade on timescales of hours, so controlling the temporal dynamics of signaling is difficult: there is no off switch. **We designed an off-switchable IL-2 mimic, enabling seconds-timescale control over its signaling**, and used this tool to investigate early events in IL-2 signaling.

Overall, this work provides a route to designing the rates and pathways of protein motion and change, which should ultimately enable construction of complex lifelike protein machines demanding precise timing and rapid functional change.

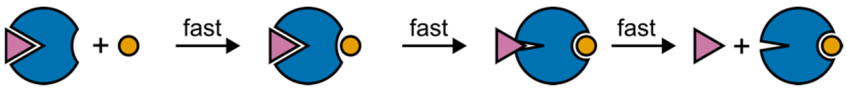


Figure 6: Schematic of induced-fit allosteric facilitated dissociation.

References:

[1] Alberstein, R. G., Guo, A. B. & Kortemme, T. Design principles of protein switches. *Curr. Opin. Struct. Biol.* **72**, 71–78 (2022).
[2] Howard, J. Protein power strokes. *Curr. Biol.* **16**, R517–R519 (2006).
[3] Pennington, L. F. *et al.* Directed evolution of and structural insights into antibody-mediated disruption of a stable receptor-ligand complex. *Nat. Commun.* **12**, 7069 (2021).
[4] Moraga, I. *et al.* Instructive roles for cytokine-receptor binding parameters in determining signaling and functional potency. *Sci. Signal.* **8**, ra114–ra114 (2015).
[5] Broerman, A. J. *et al.* Design of facilitated dissociation enables timing of cytokine signaling. *Nature* (in press) (2025).
Preprint: https://www.biorxiv.org/content/10.1101/2024.11.15.623900v1

Toehold mediated strand displacement as a tool for kinetic control of protein dissociation

Eva Rajh[1,2], Roman Jerala[1], Ajasja Ljubetič[1]

[1] National Institute of Chemistry, Department of Synthetic Biology and Immunology, Ljubljana, Slovenia
[2] University of Ljubljana, Medical faculty, Ljubljana, Slovenia

Eva.rajh@ki.si

Toehold-mediated strand displacement (TMDS) is the corner stone of many dynamic DNA origami machines. In order to transfer the principle from the DNA to the protein world we need tightly bound protein heterodimers. We designed two highly stable proteins that form a complete four-helix bundle with a low dissociation rate constant. One of the subunits contains an unpaired, coiled-coil segment that serves as a toehold (substrate strand), the other one does not have a toehold (incumbent strand).

By introducing an 'invading' protein with a matching toehold to the substrate strand, we successfully induced the displacement of the original protein from the complex up to 50 times faster than the normal complex half-life and demonstrated the formation of a short-lived triple protein complex.

Experimental results show concentration dependent rate of displacement, matching well with the theoretical modelling of the TMDS system. By varying the length and strength of interactions in the overhang and further understanding the displacement mechanism, we aim to develop dynamic protein machines mirrored on their DNA origami counterparts. Potential applications for such kinetically controlled systems include protein switches, logic circuits, and tools for powered walking of molecular machines.

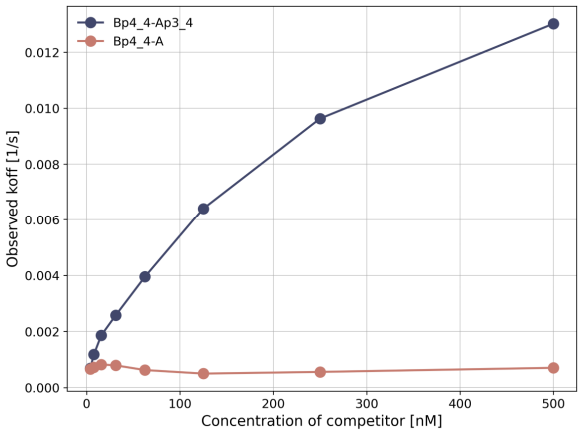


Figure 7: speed of displacement reaction is dependent on concentration of competitor strand. *Observed speed of incumbent protein exchange for an invading protein with a full matched overhang (blue) or no overhang (orange). Introduction of overhang can speed up reaction speed up to 50 times.*

Bound or not bound - annotating conformational state of protein structure models

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

[1] Institute of Organic Chemistry and Biochemistry of the Czech Academy of Sciences, Prague, 160 00
[2] Department of Cell Biology, Faculty of Science, Charles University, Prague, 128 00
[3] Faculty of Engineering Sciences, Heidelberg University, Heidelberg, 691 15
[4] Department of Software Engineering, Faculty of Mathematics and Physics, Charles University, Prague, 121 16

marian@natur.cuni.cz

Thanks to advances in protein generative modelling, most proteins in UniProtKB now have a predicted structure in the AlphaFold Protein Structure Database (AF-DB). Un-like sequence, structural data capture the conformational diversity that a protein can exhibit as a response to physiological perturbations, such as protein-ligand interactions. A single model can therefore reflect any one of these native states – or perhaps none – but not all of them. A question that emerges is “which native state does a protein model repre-sent?”

Here, we classify a selection of predicted structures from AF-DB, based on their structural similarity to two categories of native states with regard to ligand binding: the ligand-bound (holo) and the ligand-unbound state (apo) not on the chain level, but on the binding site level. After categorizing the alternative states of biologically relevant small-ligand binding sites of known protein structures from the Protein Data Bank into these two catego-ries[1]we use them as a reference to classify the same ligand-binding regions of AF-DB models according to their relative “proximity” to each category, and map the conformatio-nal diversity of these regions throughout alternative binding states in known experimental structures. We show that the apo and holo states are not always distinct and that indivi-dual binding sites differ in their conformational diversit. The conformational diversity of each binding site might affect the difficulty and performance of machine-learning prediction tasks and shall therefore be considered in planning these experiments.

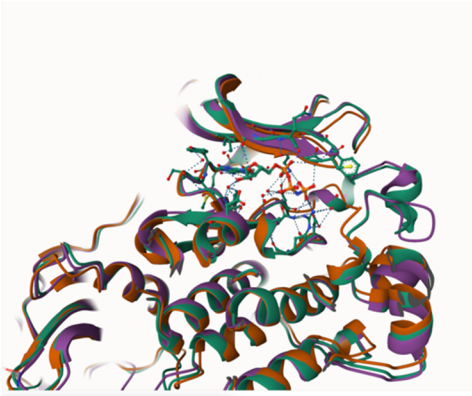


Figure 8: A comparison. Of alphafold model of human Src kinase *(violet) with its experimental apo structure(orange) and holo structure (green).*

References:

[1] Feidakis, C.P., Krivak, R., Hoksza, D., and Novotny, M. (2024). AHoJ-DB: A PDB-wide Assignment of apo & holo Relationships Based on Individual Protein–Ligand Interactions. J. Mol. Biol. 436, 168545.

De novo design of light-regulated dynamic proteins using deep learning

Lorenzo Scutteri[1,2], L.A. Abriata[1,3], S. Zhang[1], A.C. Souleiman[1,4], K. Lau[3], F. Poyer[3], S.J. Rahi[2], Patrick Barth[1]

[1] Institute of Bioengineering, EPFL, Switzerland
[2] Institute of Physics, EPFL, Switzerland
[3] Protein Production and Structure Core Facility, EPFL, Switzerland
[4] Department of Computational Biology, UNIL, Switzerland

lorenzo.scutteri@epfl.ch

Recent advances in deep learning have enabled accurate design of static protein structures, but the *de novo* design of protein functions controlled by programmable, intramolecular conformational changes remains an unsolved challenge. Previous advances include Ca^{2+}-responsive proteins[1] and stimulus-responsive hinge proteins[2]. Yet, these examples rely on rigid-body movements driven by intermolecular interactions or localized intramolecular rearrangements. In contrast, natural protein switches frequently integrate stimulus sensing and response at distant sites, enabling long-range signal transmission and the evolution of modular, multi-domain architectures. To date, the rational design of optogenetic switches with tunable, allosterically-controlled conformational dynamics and function has not been achieved.

Here, we present a **general deep learning–guided framework for designing dynamic, multi-domain proteins allosterically regulated by light**. Our strategy integrates deep learning–based multistate sequence design and structure prediction with physics-based energy calculations and molecular dynamics simulations. By integrating photoresponsive domains into *de novo* scaffolds, we engineered conformational switches that exhibit precise, reversible structural transitions upon illumination. Structural, spectroscopic, and functional analyses validated our designs and demonstrated precise spatiotemporal optogenetic control of diverse cellular processes, including subcellular localization, intercellular signaling, and population-level behaviors.

This work represents the first demonstration of **optogenetically controlled *de novo* protein switches**, and opens new avenues for programming dynamic protein functions from first principles, with implications for basic research, synthetic biology, and therapeutic development.

Link to the preprint publication: https://doi.org/10.1101/2025.08.12.669910

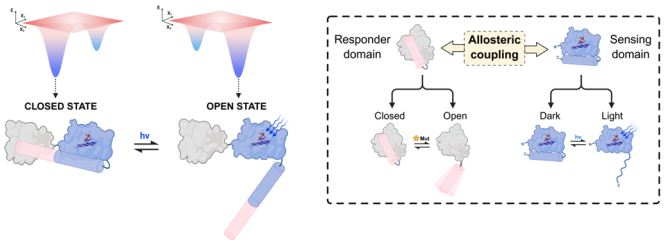


Figure 9: Deep learning-guided de novo design of dynamic protein optoswitches. *Schematic of design goal to engineer multi-domain dynamic protein optoswitches in a two-state equilibrium, where the responder domain can be allosterically controlled by the stimulus-sensing domain.*

References:

[1] A. B. Guo et al. Deep learning-guided design of dynamic proteins. Science 388, eadr7094, 2025.
[2] F. Praetorius et al. Design of stimulus-responsive two-state hinge proteins. Science, 381, 754–760, 2023.

INVITED TALK

De Novo Design of Light-Driven Protein Motor Domains

Kathryn Shelley[1,2], Cullen Demakis[1,2] & David Baker[1,2]

[1] Department of Biochemistry, University of Washington, Seattle, WA, USA
[2] Institute for Protein Design, University of Washington, Seattle, WA, USA

ks427@uw.edu

In nature, the most complex tasks performed by the cell—such as nucleic acid and protein synthesis, cargo transport, and ATP production—are accomplished by protein nanomachines. These nanomachines comprise asymmetric assemblies of proteins acting in concert to execute repeated cycles of an ordered series of steps.

A component all nanomachines require is a motor that converts biochemical energy into the mechanical motion needed to drive these ordered steps. Previously the Baker lab has published a series of modular α-helical "hinge" proteins that irreversibly switch between two structurally defined conformations upon addition of a binding partner[1]. Here I will present our research to convert these designs into light-driven protein motor domains that can rapidly cycle back and forth between their two conformational states by crosslinking them to the photoswitchable small molecule azobenzene. I will describe how we can modulate the cycling behaviour of our motor domains by tuning the properties of binding partners designed to selectively interact with different domain conformations. I will also discuss how we can incorporate these motor domains into larger protein assemblies to reversibly cycle between alternate conformations.

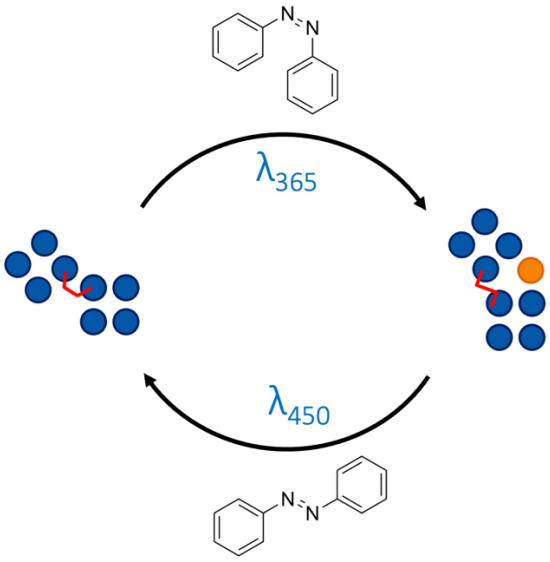


Figure 1: Schematic of light-driven motor domain. *An α-helical "hinge" protein (blue circles) is crosslinked to azobenzene (red lines) to drive cycles of conformational change plus association/dissociation with one or more binding partner(s) (orange circle).*

References

[1] Praetorius F, Leung PJY, *et al.*, (2023) Design of stimulus-responsive two-state hinge proteins. Science, 381(6659): 754-760.

Session 8: PPI prediction & binders

INVITED TALK

Variant effects on protein-protein interactions

Sven Larsen-Ledet[1], Aleksandra Panfilova[1], Amelie Stein[1]

[1] Department of Biology, University of Copenhagen, Copenhagen, Denmark

amelie.stein@bio.ku.dk

Advances in sequencing have revealed that each individual carries about 10,000 missense variants. For the vast majority, we do not know what the functional consequences – if any – will be. Further, mechanistic insight, such as structural details, would be immensely helpful in development of therapeutic approaches. We have systematically screened the human MLH1 protein, in which >150 variants are known to cause early-onset colon cancer, for stability and changes in interaction with PMS2, identifying both loss and gain of binding as well as a hitherto unknown interaction interface [1]. Further, I will present benchmarking efforts for Rosetta protocols in assessing both qualitative and quantitative changes in interaction upon mutation.

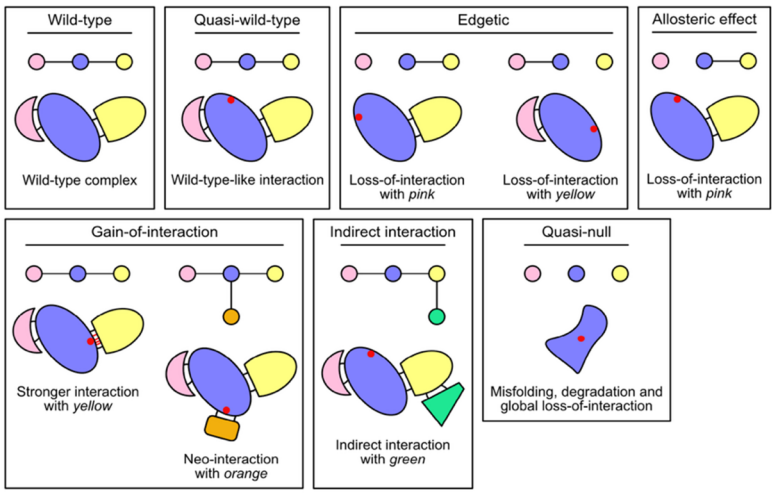


Figure 10: An overview of molecular mechanisms in mutational effects on protein-protein interactions. *For further details see [2].*

References:

[1] S Larsen-Ledet, A Panfilova, A Stein. Disentangling the mutational effects on protein stability and interaction of human MLH1. PLOS Genetics 1011681, 2025

[2] S Larsen-Ledet, A Panfilova, A Stein. Variant effects on protein-protein interactions: methods, models and diseases. *arXiv*:2507.17446, 2025

Versatile In-Vitro Validation of AI-Designed Protein Binders

Gustavo Avelar Molina[1], Diego Ruiz Espi[1], Esperanza Rivera de Torre[1], Andreas Hald[2], Anne Ljungars[1], Timothy Patrick Jenkins[1]

[1] Section for Biologics Engineering, Department of Biotechnology and Biomedicine, Technical University of Denmark, Kongens Lyngby, Denmark
[2] Research & Development, Novo Nordisk A/S, Måløv, Denmark

gumol@dtu.dk

Artificial intelligence (AI) models are increasingly enabling *de novo* protein binder design, offering enormous potential for biotechnology. Applications include the development of molecules that inhibit toxins and disease-related targets, control and stabilize industrial enzymes, serve as diagnostic reagents, and regulate pathways in metabolic engineering. Despite advances in AI-driven pipelines, achieving binders with the desired affinity, stability, and developability still requires extensive experimental validation and iterative optimization, which remain labor-intensive, expensive, time-consuming, and error-prone.

As part of an academic–industrial initiative between DTU and Novo Nordisk, we are in the process of developing a **fully autonomous closed-loop platform** for protein binder design and validation. The platform will integrate modular dry- and wet-lab components to generate and optimize effective binders against virtually any target of interest. A key component will be a technology that permits validation of binding affinity and kinetics in a rapid, cheap, and automatable fashion. For this, we are exploring **Flow Induced Dispersion Analysis (FIDA)**, a first-principle method capable of accurately determining protein size, complex formation, binding strength, and quality-control parameters. Importantly, FIDA achieves these measurements using only nano- to microliters of complex samples, including plasma and unpurified cell lysates.

Our preliminary results show that FIDA enables robust in-solution binding measurements across diverse and challenging systems, from intrinsically disordered peptides and snake venom toxins to a wide range of binders such as *de novo* designed proteins, antibodies, and VHHs. Successful measurements were also obtained in complex mixtures, including cell-free expression systems. These findings highlight the versatility of FIDA and support its integration into our platform for automated, scalable validation of AI-designed binders, paving the way toward an open-source solution that improves data quality, reduces costs, and minimizes manual labor.

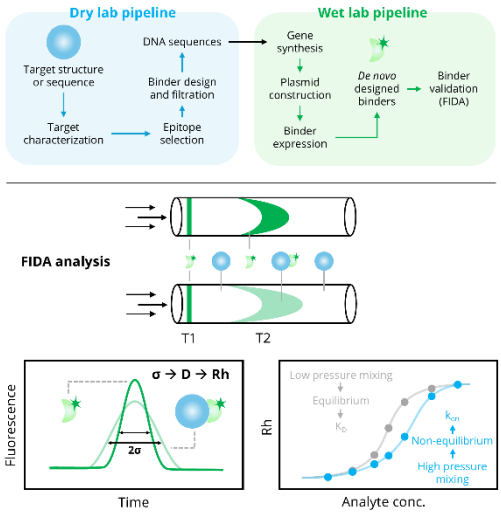


Figure 11: Closed-loop scheme and FIDA principle. *In our closed-loop platform, the dry lab branch automatically designs and filters binders based on input target sequence or structure. The wet lab branch autonomously clones, expresses, and validates the binders. FIDA measures the hydrodynamic radius (Rh) of fluorescently labelled binder or target. When mixed with binding partner, protein-protein complex is formed, and measured Rh increases. Binding affinity is measured via a titration series, and kinetics is obtained by controlling pressure to ensure out-of-equilibrium measurements.*

De novo design principles for stabilizing β-sheet frameworks to enable novel binding functions

Marta Nadal[1], Jorge-Roel Touris[1], Lourdes Carcelén[1], Joana-Albi Puig[1], Adrián Velázquez-Campoy[2], Enrique Marcos[1]

[1]Protein Design and Modeling Lab, Department of Structural and Molecular Biology, Molecular Biology Institute of Barcelona (IBMB), CSIC, Baldiri Reixac 10, 08028 Barcelona, Spain
[2]Institute for Biocomputation and Physics of Complex Systems (BIFI), University of Zaragoza, Zaragoza, Spain

embcri@ibmb.csic.es

β-sheet–based proteins are attractive scaffolds for the *de novo* design of ligand-binding pockets and loop-anchoring frameworks, yet they often face a fundamental tradeoff between stability and function. To address this challenge, we have developed design principles that enhance the stability and function of *de novo* β-sheet–containing protein folds by combining physics- and deep learning–based structure prediction and design methods. These principles are demonstrated in two distinct fold families. First, we introduce stabilization strategies for NTF2-like domains by incorporating buttressing elements on the convex face of curved β-sheets, preserving pocket geometry and accessibility [1,2]. Second, we describe the design of novel immunoglobulin-like (Ig) scaffolds—single- [3,4] and two-domain architectures [5] not found in nature—that display hyperstability, structural accuracy and robustness to anchor multiple binding loops; surpassing limitations of natural Ig frameworks. These approaches demonstrate that structural stabilization and functional site optimization need not be mutually exclusive, paving the way for more efficient computational design of functional β-sheet–based proteins.

References:

[1] E. Marcos, B. Basanta, T.M. Chidyausiku, Y. Tang, G. Oberdorfer, G. Liu, G.V.T. Swapna, R. Guan, D.-A. Silva, J. Dou, J. H. Pereira, R. Xiao, B. Sankaran, P. H. Zwart, G. T. Montelione, D. Baker. Principles for designing proteins with cavities formed by curved β-sheets. *Science*, 355: 201-206, 2017.

[2] M. Nadal, J. Albi-Puig, A. Castellví, P. Garcia-Franco, S. Vega, A. Velázquez-Campoy, E. Marcos. Buttressing ligand-binding pockets in de novo designed NTF2-like domains. *Under review*.

[3] T. M. Chidyausiku, S. Mendes, J. Klima, M. Nadal, U. Eckhard, J. Roel-Touris, S. Houliston, T. Guevara, H. K. Haddox, A. Moyer, C. H. Arrowsmith, F. X. Gomis-Rüth, D. Baker D, E. Marcos. De novo design of immunoglobulin-like domains. *Nat. Commun.*, 13: 5661, 2022

[4] J. Roel-Touris, L. Carcelén, E. Marcos. The structural landscape of the immunoglobulin fold by large-scale de novo design. *Protein Science*, 33(4): e4936, 2024

[5] J. Roel-Touris J, M. Nadal, E. Marcos. Single-chain dimers from de novo immunoglobulins as robust scaffolds for multiple binding loops. *Nat. Commun.*, 14: 5939, 2023.

Generating All-Atom De Novo Protein Binder Designs with High Lab Success Rates

Alex Bridgland, Jonathan Crabbé, Henry Kenlay, Daniella Pretorius, Sebastian M. Schmon, Agrin Hilmkil, Rebecca Bartke-Croughan, Robin Rombach, Michael Flashman*, Tomas Matteson, Simon Mathis, Alexander W. R. Nelson**, David Yuan, Annette Obika, Simon A. A. Kohl*****

*** Corresponding author. E-mail: simon@latentlabs.com.
** Work performed as an advisor to Latent Labs.
* Work performed while at Latent Labs.

We are introducing a new state of the art frontier AI model for generating lab-functional protein binders with all-atom resolution. The model generalizes and can go beyond the repertoire of nature by generating all-atom binder structures that obey the biochemical rules of binding at the atomic level, such as the direct generation of hydrogen bonds or the pi-stacking of aromatic rings. We present experimental validation results for de novo protein binders across two therapeutically relevant modalities: macrocycles and mini-binders. The frontier AI model is made publicly available via web UI, allowing to reproduce the lab-validated results. Generating and scoring binder designs for lab-feasible batch sizes completes in the order of minutes and does not require coding skills.

Macrocycle Validation: We experimentally tested de novo macrocycle designs against multiple therapeutically relevant protein targets. Our approach achieved high binding success rates across all targets, with functional binders identified for every tested target.

Mini-binder Validation: Across multiple therapeutically relevant protein targets, we validated computationally designed mini-binders. Functional binders were identified for all tested targets with high success rates. Generated designs span diverse folds and extend to substantial protein lengths suitable for therapeutic applications.

Further Validation: We will present additional lab results obtained both internally and externally.

Methodological Impact: Our experimental validation demonstrates consistent success across diverse target classes, with functional binding hits identified for every target tested. This reliability enables predictable workflows for therapeutic binder development, potentially transforming the drug discovery pipeline from screening-based to design-based approaches.

References:

[1] Latent Labs Team. (2025, July 22). *Latent-X: An Atom-level Frontier Model for De Novo Protein Binder Design*. Latent Labs. https://www.latentlabs.com/wp-content/uploads/2025/07/Latent-X-Technical-Report.pdf

De novo design of mini-protein binders broadly neutralizing Clostridioides difficile toxin B variants

Xinchen Lyu[1], Ke Sun[1] , Peilong Lu[1]

[1] Research Center for Industries of the Future, Westlake University, Hangzhou, Zhejiang 310024, China

lyuxinchen@westlake.edu.cn

Clostridioides difficile toxin B (TcdB) is the key virulence factor accounting for C. difficile infection-associated symptoms. Effectively neutralizing different TcdB variants with a universal solution poses a significant challenge. Here we present the de novo design and characterization of pan-specific mini-protein binders against major TcdB subtypes. Our design successfully binds to the first receptor binding interface (RBI-1) of the varied TcdB subtypes, exhibiting affinities ranging from 20 pM to 10 nM. The cryo-electron microscopy (cryo- EM) structures of the mini protein binder in complex with TcdB1 and TcdB4 are consistent with the computational design models. The engineered and evolved variants of the mini-protein binder and chondroitin sulfate pro- teoglycan 4 (CSPG4), another natural receptor that binds to the second RBI (RBI-2) of TcdB, better neutralize major TcdB variants both in cells and in vivo, as demonstrated by the colon-loop assay using female mice. Our findings provide valuable starting points for the development of therapeutics targeting C. difficile infections (CDI)[1].

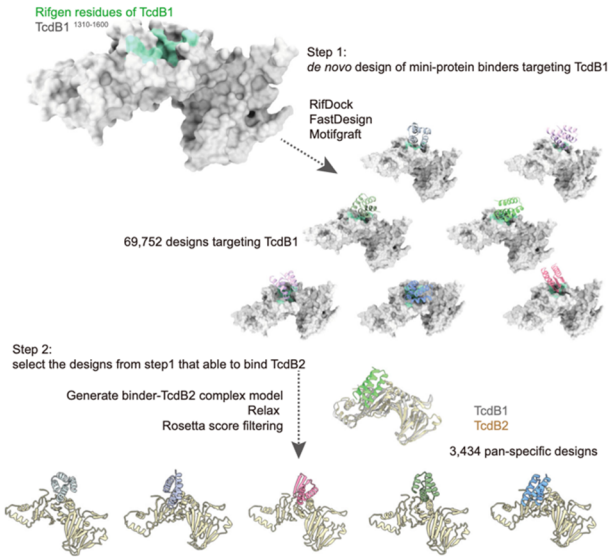


Figure 12: Overview of the computational design of broad-spectrum mini-protein binders against TcdB variants. Our computational design approach consists of two steps. *In step 1, we performed the de novo design of mini-protein binders targeting TcdB1; in step 2, we selected binders with the capability to bind the RBI-1 of TcdB2 from the previously designed pool of mini-protein binders targeting TcdB1.*

References:

[1] **Lv, Xinchen#** ; Zhang, Yuanyuan# ; Sun, Ke;Tao ,Liang; Lu,Peilong. De novo design of mini-protein binders broadly neutralizing *Clostridioides difficile* toxin B variants, ***Nature Communications, 2024***, 15, 8521

A strategy for the de novo design of highly polyspecific proteins targeting coral snake three-finger toxins

Kasper H Björnsson[1], **Cecilie Knudsen**[1], **Emil Berntsen**[1], **Melisa Benard-Valle**[1], **Timothy P Jenkins**[1]

[1] Technical University of Denmark, Department of Bioengineering, Lyngby, Denmark

khabj@dtu.dk

Snakebite envenomings are responsible for up to 130,000 deaths and 400,000 disabilities every year. Current treatments rely on century-old technology - serum IgG isolated from immunized animals - which present side-effects, batch-to-batch variation, and limited cross-reactivity [1]. Recent advances in diffusion-based protein design have made it possible to design proteins that recognize and neutralize the highly toxic and medically relevant snake venom three-finger toxins (3FTx), but the designs show limited polyspecificity[2], [3]. Given the vast diversity of snake venoms, polyspecificity is crucial for developing effective therapies.

Here, we present a strategy for the design of highly polyspecific proteins against a set of targets, which we applied on coral snake 3FTx. Initially, protein binders were designed against all AlphaFold2 models of snake 3FTx available from Uniprot using RFdiffusion and ProteinMPNN. Designs were filtered using AlphaFold2 and cross-panned in silico against all coral snake 3FTx using Boltz-2 [4]. The top polybinders, identified by predicted alignment error derived metrics, were then diversified through partial diffusion and ProteinMPNN with all coral snake 3FTx as targets, and subsequently cross-panned again. Twenty designs, selected for target breadth and diversity, are now undergoing in vitro testing for binding to coral snake toxins, followed by in vivo neutralization assays. As several designs are predicted to bind over 20 toxins, this work could lead to novel polyspecific antivenoms and pave the way for designing binders against other diverse protein sets.

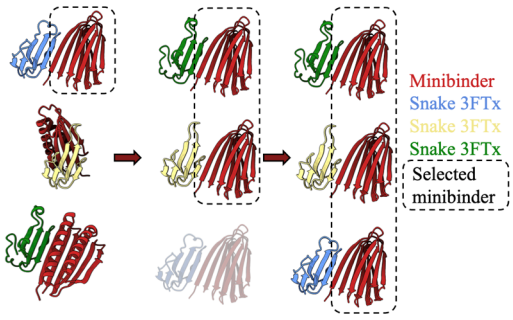


Figure 1: Design of polyspecific pro-teins. *Binding proteins were designed against all targets, followed by an in silico cross-panning. The most polyspecific are subjected to further diversification, resulting in increased polyspecificity.*

References:

[1] S. K. Menzies, R. N. Patel, and S. Ainsworth, “Practical progress towards the development of recombinant antivenoms for snakebite envenoming,” 2025, *Taylor and Francis Ltd.* doi: 10.1080/17460441.2025.2495943.

[2] S. Vázquez Torres *et al.*, “De novo designed proteins neutralize lethal snake venom toxins,” *Nature*, vol. 639, no. 8053, pp. 225–231, Mar. 2025, doi: 10.1038/S41586-024-08393-X.

[3] J. L. Watson *et al.*, “De novo design of protein structure and function with RFdiffusion,” *Nature*, vol. 620, no. 7976, pp. 1089–1100, Aug. 2023, doi: 10.1038/S41586-023-06415-8.

[4] S. Passaro *et al.*, “Boltz-2: Towards Accurate and Efficient Binding Affinity Prediction,” *bioRxiv*, p. 2025.06.14.659707, Jun. 2025, doi: 10.1101/2025.06.14.659707.

Session 9: Binders & membrane proteins

Practically useful computational liability analysis for protein-based binders

Moritz Ertelt[1,2], Leonie Reetz[1], Peter Braun[1,2]

[1] Fraunhofer Institute for Translational Medicine and Pharmacology ITMP - Immunology, Infection and Pandemic Research IIP, Penzberg, Germany
[2] Institute of Infectious Diseases and Tropical Medicine, University Hospital Ludwig-Maximilian University Munich, Munich, Germany

moritz.ertelt@itmp.fraunhofer.de

Protein-based binders have vast applications in medicine, diagnostics and biotechnology. While antibodies derived from immunization of animals still dominate, new technologies aim to either speed up the development or fill the gaps where antibody-based approaches have failed. This includes computational generation of de novo proteins or antibody formats, as well as utilization of natural proteins. There are already many potential pitfalls, such as post-translational modifications altering the stability or affinity over time in classical immunization-based approaches. For protein-based binders derived from alternative, computational approaches, this could be substantially worse as these proteins have not undergone classical selection processes like clonal deletion, which purges self-interacting B cells and therefore restricts interaction with common structures. Therefore, we have created a computational screening in Rosetta that checks a myriad of potential liabilities, ranging from simple sequence-based to machine learning approaches. These checks include common receptor-binding sequence motifs, hydrophobic and charged patches, degradation motifs and post-translational modifications. We then use this new tool to analyze the rate of potential liabilities in computationally generated binders and highlight how it can be used to de-risk candidates.

Generative Design of Functional Proteins by Fusing Sequence and Structural Modalities

Chentong Wang[1], Shuohua Chen[1], Yannan Qu[1], Zhangzhi Peng[2], Ze Zhang[1]

[1] School of Life Sciences, Westlake University, Hangzhou, Zhejiang, China
[2] Duke University, Durham, North Carolina, USA

wangchentong@westlake.edu.cn

The predominant paradigm for current generative models focuses on generating only the protein backbone. This backbone-centric approach fundamentally limits their ability to leverage the vast information within sequence databases and, consequently, to model the sequence-dependent, all-atom structures essential for function. To overcome this, we have developed a novel, unified platform that introduces a sequence generation model—pre-trained on tens of millions of sequences in a highly compressed latent space—directly into the structural generation process. This enables a seamless co-generation of both sequence and all-atom structure.

Our multi-modality joint mask strategy unifies all major design paradigms into a single framework without task-specific fine-tuning. The model has been robustly validated through the successful de novo design of pico- to nanomolar binders against challenging targets (e.g., Amyloid-beta 42, α-synuclein, TNF family, IL3Ra, IL13Ra2) and several structurally confirmed, atomically accurate monomers up to 1254 residues.

In conclusion, by seamlessly integrating pre-trained sequence knowledge into the structural generation process, our model establishes a powerful new paradigm for computational protein design. This unified framework's demonstrated ability to co-generate both the sequence and all-atom structure for a wide array of functional proteins, from large monomers to high-affinity binders, paves the way for accelerated innovation in therapeutic and biotechnological applications.

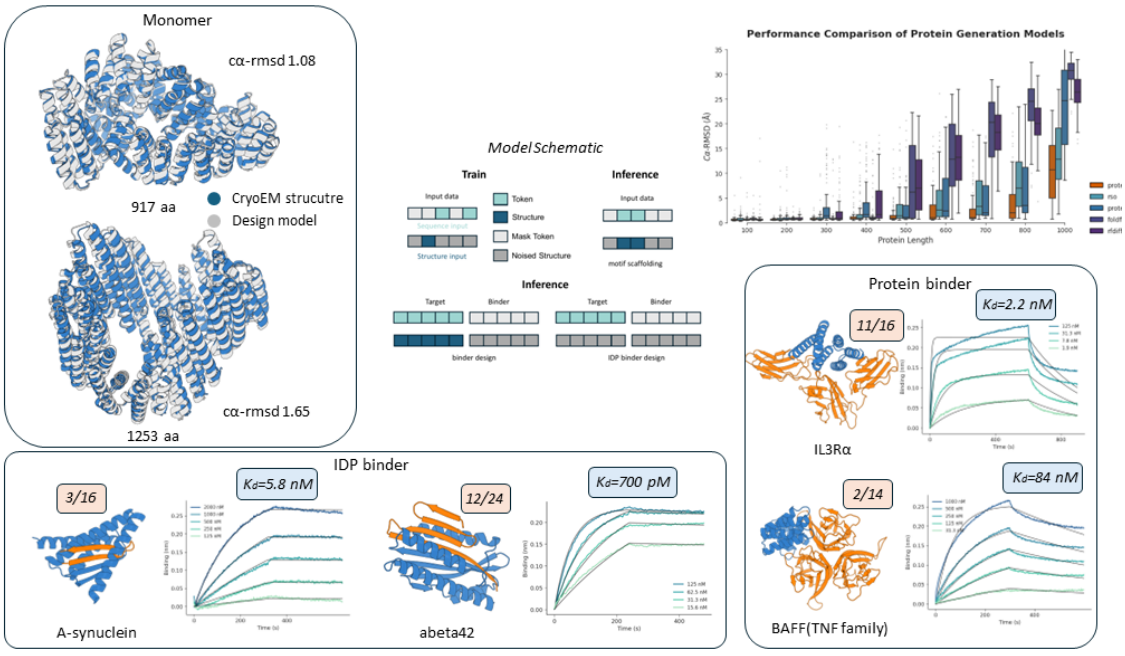


Figure 1: Model Schematic and Performance Validation. *The figure displays the model's schematic and performance benchmarks alongside key experimental results: the successful de novo design of atomically accurate monomers and pico- to nanomolar binders for challenging targets.*

Targeting the RANK-TRAF6 Axis: A Molecular Approach

Rita Melo[1,*], Margarida Carreira[1], Francisco Fernandes[2], João D. G. Correia[1]

[1] Centro de Ciências e Tecnologias Nucleares and Departamento de Engenharia e Ciências Nucleares, Instituto Superior Técnico, Universidade de Lisboa, CTN, Estrada Nacional 10 (km 139,7), 2695 – 066 Bobadela LRS, Portugal
[2] Instituto de Engenharia de Sistemas e Computadores: Investigação e Desenvolvimento em Lisboa (INESC-ID), Lisbon, Portugal

ritamelo@ctn.tecnico.ulisboa.pt

This work explores the computational design of peptide-based inhibitors targeting the RANK-TRAF6 signaling axis, an essential pathway driving osteoclasteogenesis and the development of bone metastases in prostate and breast cancers [1]. This study employed an integrative in silico design pipeline using state-of-the-art protein modeling and generative tools, which includes RFdiffusion, AlphaFold, and ProteinMPNN, to engineer peptides capable of disrupting the RANK–TRAF6 protein-protein interaction. Peptide candidates were prioritized based on predicted binding affinity and interface complementarity using docking and scoring tools such as AutoDock Vina, GNINA, PPI-Affinity, HADDOCK, and molecular dynamics simulations to assess interaction stability and conformational dynamics.

This study presents a structure-guided, computation-first strategy for peptide inhibitor development against challenging protein-protein interaction targets. The integration of deep learning-based design with classical Rosetta-compatible evaluation methods underscores the promise of hybrid computational-experimental pipelines for accelerating targeted therapy discovery.

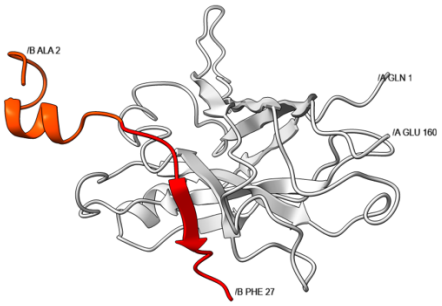


Figure 1: *The selected binder (orange) complexed with TRAF6 (light gray).*

References:

[1] J. Ming, S.J.F. Cronin, J.M. Penninger. Targeting the RANKL/RANK/OPG Axis for Cancer Therapy. *Front Oncol*. 7(10) 1283, 2020.

AI protein design at ACE2 binding interface

Yang Yang[1], Changsheng Zhang[1], Lu hualai[1,2]

[1] College of Chemistry and Molecular Engineering, Peking University, Beijing, China
[2] Center for Quantitative Biology, Peking University, Beijing, China

2001110427@stu.pku.edu.cn

The ab-initio protein design harnesses the power of artificial intelligence to create novel protein structures from first principles, offering an unprecedented level of precision and control over protein function. Unlike traditional de novo design methods that rely on static force fields or limited backbone generation1, our approach integrates diffusion-based scaffold design (RFD-iffusion) with structure-aware sequence optimization (ProteinMPNN) and multi-tiered validation (AlphaFold2 + MD), enabling efficient generation of high-affinity binders with experimentally verifiable stability.

In the context of the COVID-19 pandemic, we designed a trimeric bundle of alpha-helix minibinders with high affinity to the spike-binding interface of human angiotensin-converting enzyme 2 (ACE2). A computational protein design pipeline was applied with the aim of impairing SARS-CoV-2cell entry via competitive ACE2 binding, supporting the development of broad-spectrum antiviral agents. Surface Plasmon Resonance (SPR) assays determined a Kd of 2 nM for ACE2 and confirmed competitive binding against the SARS-CoV-2 RBD. Confocal imaging revealed specific fluorescence signals localized to ACE2-expressing cell membranes, validating binder localization. This work demonstrates a generalizable AI-assisted framework for targeting protein-protein interfaces, with implications for rational drug design—particularly in optimizing binding affinity and specificity through computational screening, a key focus of chemical informatics.

References:

[1] Chen, Y. et al. ACE2-targeting monoclonal antibody as potent and broad-spectrum coronavirus blocker. *Signal Transduct. Tar.* 6, 1–9 (2021).
[2] Watson, J. L. et al. De novo design of protein structure and function with RFdiffusion. *Nature* 620, 1089–1100 (2023).
[3] Dauparas, J. et al. Robust deep learning–based protein sequence design using ProteinMPNN. *Science* 10.1126 (2022).

INVITED TALK

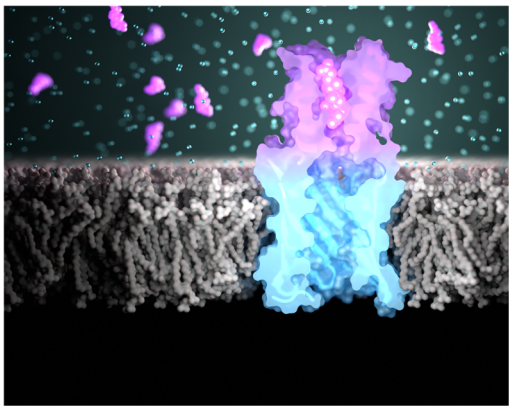
De novo design of β-barrels nanopores with defined size and function

Anastassia A. Vorobieva[1,2]

[1] Structural Biology Brussel, Vrije Universiteit Brussels, Brussels, Belgium.
[2] VUB-VIB Center for Structural Biology, Brussels. Belgium.

Anastassia.andreevna.vorobieva@vub.be

The field of de novo protein design has undergone major transformations in recent years, largely driven by advances in artificial intelligence. However, most progress has focused on proteins that fold in aqueous environments, while de novo design of membrane proteins (MPs) is only beginning to emerge. In this talk, I will discuss how both classical and modern design approaches can be applied to the design of transmembrane β-barrels (TMBs)—a class of multi-pass β-sheet MPs that form water-accessible pores across lipid membranes. Using de novo design, we generated TMBs with varying sizes, strand numbers, and lumen shapes [1,2]. Structural characterization by X-ray crystallography and NMR confirmed that these designs folded into the intended architectures. The most stable designs were shown to integrate into planar lipid bilayers and form stable nanopores. By fusing a de novo–designed binder to the cholic acid small molecule, we demonstrated the first proof-of-principle for the fully de novo design of nanopore sensors [3]. Finally, I will discuss current limitations in de novo design methods—particularly their difficulty in modeling unfolded states (which are critical for TMB design [4]) and non-globular architectures—and present our ongoing efforts to address these challenges through the development of new design models.



References:

[2] Vorobieva A.A. *et al.* (2021), De novo design of transmembrane β-barrels, Science, 371: eabc8182.
[3] Berhanu S., Majumder S., *et al.* (2024), Sculpting conducting nanopore size and shape through de novo protein design, Science, 385: 282-288.
[4] An L., *et al.* 2024), Binding and sensing diverse small molecules using shape-complementary pseudocycles, Science, 385:276-282.
[5] Pedrelli G., *et al.* (2025) Negative design enables cell-free expression and folding of designed transmembrane β-barrels, bioRxiv.

Expression of a *de novo* fluorescence-activating transmembrane β-barrel in mitochondrial membranes

Alvaro Martin[1,2], Hahnbeom Park[3], Julie Van Duyse[4], Gert Van Isterdael[4], Anastassia A. Vorobieva[1,2,5]

[1] Structural Biology Brussels, Vrije Universiteit Brussel, Brussels, Belgium
[2] VIB-VUB Center for Structural Biology, Brussels, Belgium
[3] Korea instiute of Science and Technology, Seoul, South Korea
[4] VIB Flow Core, VIB Center for Inflammation Research, Ghent, Belgium
[5] VIB Center for AI and Computational Biology, Belgium
Alvaro.martin@vub.be, Anastassia.Andreevna.Vorobieva@vub.be

Membrane protein design poses a significant challenge due to its difficulty to computationally desing models and test them experimentally [1]. In addition, there is a great interest in the functionalisation of membrane proteins for their use as biosensors, selective transporters and therapeutics, although their success has been limited.
Here, we attempt to tackle both of these challenges: we employ computational design tools to generate a ***de novo* transmembrane β-barrel** (TMB) with a binding site that can accomodate a **fluorogenic ligand**. We developed a high-throughput wrokflow to express and screen fourty TMBs with varying pocket conformations, out of which 7 were capable of binding and activating three different fluorogens with distinct spectral profiles [2]. Further characterisation of the best designs verfied the predicted binding mode and explained the observed differences in ligand specifity between designs. Using Rosetta ddG calculations and Molecular Dynamics simulations of the TMB/ligand complex, we identified surface mutations which improved protein stability and fluorogen activation simulataneously, yielding a variant with a 2-fold increase in fluorescence.
Next, the brightest variant was succesuflly **expressed in mitochondrial membranes** of yeast, retaining its fluorescent properties without impacting cellular growth or mitochondrial function. Finally, we constructed a library of mutants and identified designs with optimised *in vivo* expression, folding and binding characteristics by fluorescence activating cell sorting (FACS).
The TMB and methods described in this research present an ideal framework to elucidate the biophysical principles involving membrane protein biogeneis and mitochondrial biology *in vivo*

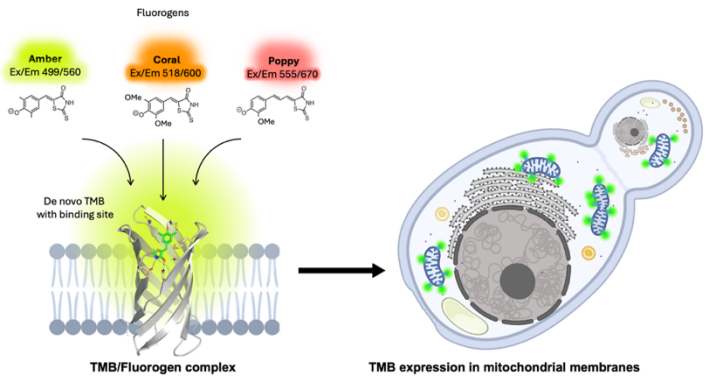


Figure1: Overview of the functional TMB expressed into mitochondrial membranes.

(Right) The de novo TMB is capable of binding and activating three different fluorogens with the same pocket.

(Left) This TMB is expressed into mitochondrial membranes whilst retaining its fluorescent properties

References:

[1] Vorobieva A. A. (2021). Principles and Methods in Computational Membrane Protein Design. *Journal of molecular biology*, *433*(20), 167154.
[2] Benaissa, H., et al. (2021). Engineering of a fluorescent chemogenetic reporter with tunable color for advanced live-cell imaging. *Nature communications*, *12*(1), 6989.

POSTERS

(1) In-silico Design of Foldamer – Protein Interactions

Hubert Klein[1], Yajuan Ma[2], Zitong Chen[2], Florian Sanchez[2], Johannes Sigl[2], Prof. Dr. Huc[2], Prof. Dr. Khmelinskaia[1]

[1] Department of Physical Chemistry, LMU, Munich, Germany
[2] Department of Pharmaceutical Chemistry, LMU, Munich, Germany
hubklph@cup.uni-muenchen.de

Foldamers are a chemist's dream of proteins. Unlike natural peptides – where sidechain identity and backbone conformation are intrinsically coupled – aromatic delta-peptide foldamers adopt a helical scaffold regardless of sidechain variation. The decoupling is made possible by pi-pi interactions and hydrogen bonding at the backbone level, leading to a very rigid scaffold that can be modularly modified on the sidechain level. As a result, foldamers are perfect candidate molecules to be used as versatile building blocks in macromolecular assemblies and modulators of protein-protein interactions.

First, I will present to you our approach to the design of **foldamer-protein hybrid assemblies** (Figure 1, left), combining foldamer modularity with de-novo protein design. By adapting the PPI-seeded design pipeline established in our lab (see Ho Yeung Chim's poster), I demonstrate how foldamer-binding proteins previously identified through display selection techniques, can be modularly used to generate cages of predetermined geometry. Such cages could exhibit physiochemical properties unattainable with natural amino acids.

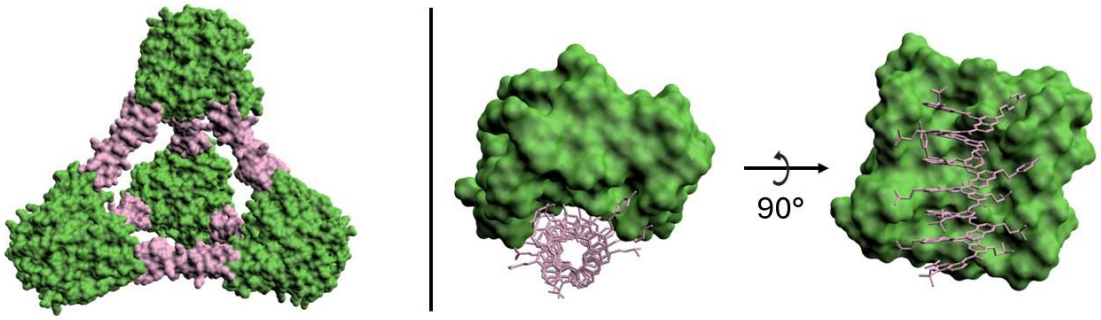


Figure 1: Protein-foldamer constructs. Left:*Tetrahedral assembly formed by a C3-symmetric protein (green) and a C2-symmetric foldamer (pink).* ***Right:*** *Designed foldamer-protein interaction with a concave miniprotein binder.*

Second, I will present our efforts to establish a pipeline for the de novo design of protein-foldamer interactions. While previously characterized interactions rely on the hydrophobicity of the foldamer cross-section, our goal is to achieve direct engagement of the foldamer side-chains upon binding. To pursue this, we adapt the well-established Rosetta-based RifDock/RifGen protein binder design pipeline[1] to generate foldamer-binding proteins (Figure 1, right) and assess the accuracy of the Rosetta Scoring function for these non-natural interfaces. I will share the first *in vitro* validation experiments that have already identified one weak candidate binder. Ultimately we aim to invert this workflow to design foldamers targeting natural protein interfaces.

References:

[1] Cao, L., Coventry, B., Goreshnik, I. *et al.* Design of protein-binding proteins from the target structure alone. *Nature* **605**, 551–560 (2022)

(2) Computational design of functional cyclic oligomers through interface-seeded design

Ho Yeung Chim[1], Mubarak Olanrewaju Idris[2], Sandrine Georgeon[2], Dominic Rieger[3], Phillip Schlegel[3], Prof. Clara T. Schoeder[3], Prof. Bruno Correia[2], Prof. Alena Khmelinskaia[1]

[1] Ludwig Maximilian University of Munich, Department of Chemistry, Munich, Germany
[2] Laboratory of Protein Design and Immunoengineering, Institute of Bioengineering, Ecole polytechnique fédérale de Lausanne, Lausanne, Switzerland
[3] Institute for Drug Discovery, Faculty of Medicine, Leipzig University, Leipzig, Germany

h.chim@lmu.de

Protein design enables the creation of novel biomaterials, but programming proteins with controlled and dynamic assembly remains a key challenge [1]. Natural proteins often respond to environmental cues by changing their oligomeric states, a feature difficult to replicate in designed systems. Here, we introduce a new RFdiffusion [2] module for the design of cyclic oligomers using pre-existing protein interfaces. Our method preserves interface properties while scaffolding them into diverse symmetric target geometries. Our experimental results confirm **successful design based solely on interface inputs**, including environmentally responsive assemblies. To expand the design space of symmetric proteins, we introduce the concept of **"sort of symmetries"**—a strategy that enforces symmetry at the scaffold level while allowing diverse validated interfaces. This approach mimics pseudo-symmetry, reducing design complexity while maintaining the potential for functional diversity, opening new possibilities to multiplexed sensors. Taken together, this work demonstrates a generalizable approach for creating symmetric protein assemblies with dynamic and controllable properties, advancing their integration into complex biological contexts, e.g. signaling hubs.

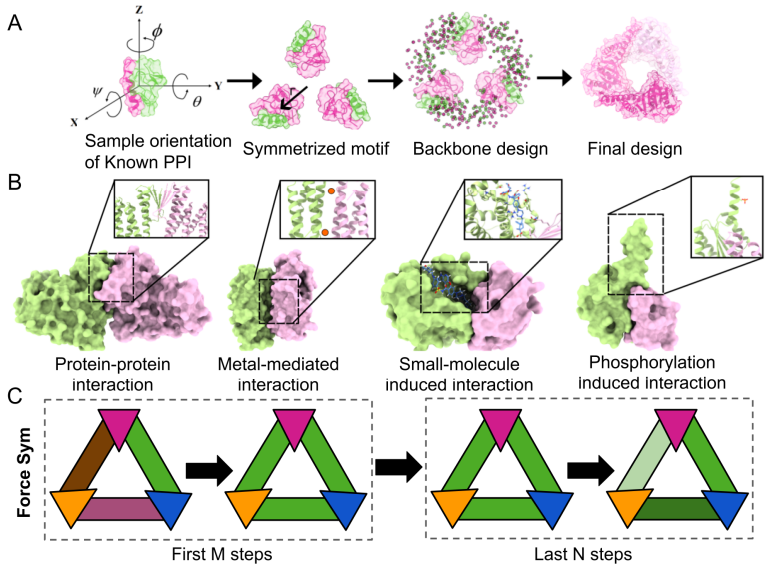


Figure 1: Interface seeded design approach for cyclic protein. ***(A)*** *Computational pipeline for interface-seeded design.* ***(B)*** *A large variety of functional interfaces are benchmarked with the new pipeline.* ***(C)*** *Designing cyclic oligomers with "sort of symmetry" from different validated interfaces by enforcing symmetry at early steps of the diffusion process.*

References:

[1] Mallik, Bhoomika Basu, et al. "De novo design of polyhedral protein assemblies: before and after the AI revolution." ChemBioChem 24.15 (2023): e202300117.
[2] Watson, Joseph L., et al. "De novo design of protein structure and function with RFdiffusion." Nature 620.7976 (2023): 1089-1100.

(3) Towards predicting oligomorphism in de novo designed protein assemblies

Bhoomika Basu Mallik[1], Leyi Liu[2], Kevin O' Brien[1], Neville Bethel[2], Alena Khmelinskaia[1]*

[1] Ludwig Maximilian University of Munich, Germany
[2] University of California, San Diego, USA

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 [1]. 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. We are generating test cases by creating pairs of oligomorphic vs. monomorphic assemblies, by fixing or diversifying individual degrees of freedom. Preliminary results show that single shot template-based prediction over a diverse set of templates allows distinguishing between flexible and rigid sequences. We are expanding the pipeline to generate an assembly phase diagram, allowing us to controllably incorporate flexibility as a design principle and predict assembly outcome.

Figure 1:. Can we control scaffold flexibility to give us predictable assembly outcomes? *we are developing a computational pipeline that combines physics-based and AI-driven methods to rapidly predict structural flexibility and consequently control assembly out-come.*

References:

[1] Khmelinskaia, A., Bethel, N.P., Fatehi, F. *et al.* Local structural flexibility drives oligomorphism in computationally designed protein assemblies. *Nat Struct Mol Biol* **32**, 1050–1060 (2025). https://doi.org/10.1038/s41594-025-01490-z

(4) Computational Investigation of Phenolic Compounds from Hazelnut Green Shell on *S. Aureus* Antibacterial and Antibiofilm Targets

Görkem Güngör[1], Mehmet Özbil[1]

[1]Gebze Technical University, Biotechnology Institute, Kocaeli, Türkiye

gorkem.gungor2017@gtu.edu.tr

Every year, thousands of tons of hazelnut green shells are discarded as agricultural waste in Türkiye, despite their rich content of bioactive compounds. Efficiently utilizing this waste could significantly contribute to sustainability and the circular bioeconomy by transforming it into valuable resources for therapeutic applications. In this study, the structural and functional effects of hazelnut green shells, which are widely produced in Türkiye and evaluated as agricultural waste in our country, against multidrug-resistant *Staphylococcus aureus* bacteria were comprehensively investigated using computational methods. Two different strains, strong biofilm-forming (ATCC 25923) and weak biofilm-forming (ATCC 29213) [1], were selected to evaluate strain-specific molecular interactions and binding preferences in detail.

The antibacterial target was identified as the DNA Gyrase A-B protein complex, while the antibiofilm target proteins included icaA, icaD, and fnbA proteins. The three-dimensional structures of these target proteins were modeled with high confidence using the AI-based AlphaFold algorithm [2], and molecular dynamics (MD) simulations were performed to obtain stable and biologically relevant protein conformations. Phenolic compounds identified from the hazelnut green shells were screened and their binding affinities on the protein targets were calculated via molecular docking simulations. Ligand-protein complexes with the highest binding affinities were further evaluated for interaction stability and dynamic behavior within the binding pocket through MD simulations. Binding free energies were calculated using the MM/PBSA method [3].

Early computational findings indicated that quercetin revealed favorable interactions with the antibacterial target and quercetin and catechin with the antibiofilm target proteins, suggesting their potential as potential inhibitor molecules. These results provided early insights that could guide future detailed simulations and experimental validation.

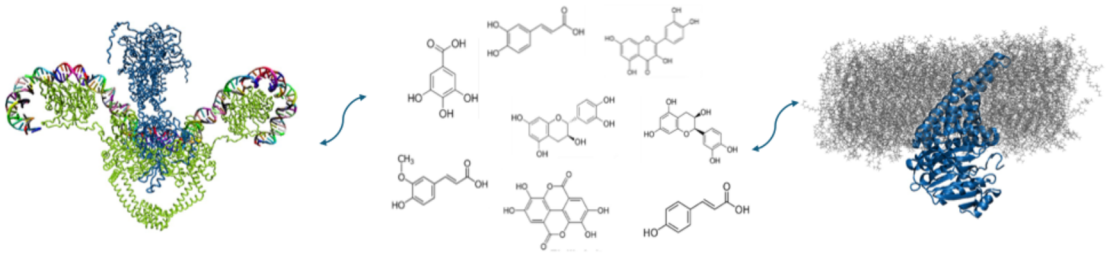


Figure 1: *Antibiotic and antibiofilm targets obtained from Alphafold and phenolic compounds obtained from hazelnut green shell extracts.*

References:

[1] Balaban, M., Koc, C., Sar, T., & Akbas, M. Y. (2021). Antibiofilm effects of pomegranate peel extracts against B. cereus, B. subtilis, and E. faecalis. International Journal of Food Science and Technology, 56(10), 4915–4924.

[2] Jumper, J., Evans, R., Pritzel, A. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

[3] Wang, E., Sun, H., Wang, J., Wang, Z., Liu, H., Zhang, J., and Hou T., *Chemical Reviews* **2019** *119* (16), 9478-9508

(5) Make structure generative prediction model an all-atom scaffold optimizer and designer

Minchao Fang[1], Chentong Wang[1], Qihan Jin[1], Longxing Cao[1]

[1] Westlake University, School of Life Science, Hangzhou, China

fangminchao@westlake.edu.cn

The advent of AlphaFold3 has initiated a new paradigm by leveraging generative frameworks, such as diffusion models, for structure prediction. Given the architectural similarity between these predictive models and generative structure diffusion models, we explored the possibility of directly harnessing the original, pretrained network for optimization and design tasks. This approach seeks to bypass the need for fine-tuning or conventional backpropagation-based hallucination methods, proposing a more direct route from prediction to design [1].

We demonstrate this concept through an all-atom sequence and structure co-optimization framework. Its core leverages an intrinsic property of diffusion-based predictors like AlphaFold3: they are trained to maximize a learned probability distribution of structures, often by reversing a stochastic process. We found this principle transforms the model itself into a powerful optimizer. This spontaneous optimization is broadly applicable across diverse systems, from protein monomers to complexes involving ligands and nucleic acids.

By leveraging the optimization power of hallucination in this novel framework, we successfully demonstrate a wide range of applications. We designed ligand-binding proteins, cyclic peptides with Unnatural Amino Acids (UAAs), antibodies, and large monomers with over 1000 amino acids. In experimental validations, we engineered several small molecule binders with high success rates, showcasing the potential of this approach as a general and effective platform for precision protein design.

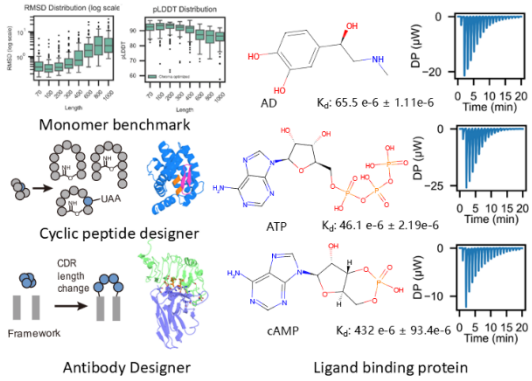


Figure 1: A universal protein designer. *This figure illustrates the application of our framework to monomers and cyclic-peptide antibodies, as well as its practical use in designing highly hydrophilic, multi-torsional small molecules.*

References:

[1] Cho Y, Pacesa M, Zhang Z, Correia B, Ovchinnikov S. BoltzDesign1: Inverting All-Atom Structure Prediction Model for Generalized Biomolecular Binder Design. bioRxiv. 2025:2025.04. 06.647261.

(6) BinderFlow: A batch-based pipeline for protein binder design

Carlos Chacón-Sánchez[1$], Nayim González-Rodríguez[1$], Oscar Llorca[1], Rafael Fernández-Leiro[1*]

[1] Spanish National Cancer Research Center, Structural Biology Programme, Madrid, Spain
$ Equal contributon * Corresponding Author

cchacon@cnio.es

Generative models have transformed de novo protein design, enabling the creation of novel protein structures with remarkable accuracy [1]. However, current workflows—typically involving large, sequential jobs for backbone generation with RosettaFold Diffusion [1], sequence assignment with ProteinMPNN [2], and design scoring with AlphaFold2-InitialGuess [3]—**remain computationally intensive, limiting accessibility for many labs**. Moreover, because in silico success cannot be predicted in advance, campaigns often over- or under- produce designs, wasting computational resources or requiring additional runs. Finally, design evaluation and selection for experimental validation remain complex and expert-driven.

We present BinderFlow, a modular, parallelized pipeline for protein binder design on SLURM-based computing systems. **BinderFlow enables real-time monitoring, simplifies design extraction and evaluation**, and streamlines preparation for downstream applications. By lowering the entry barrier for non-specialists and offering flexibility for experts, BinderFlow makes generative protein design more accessible, efficient, and scalable for both exploratory and production-level research.

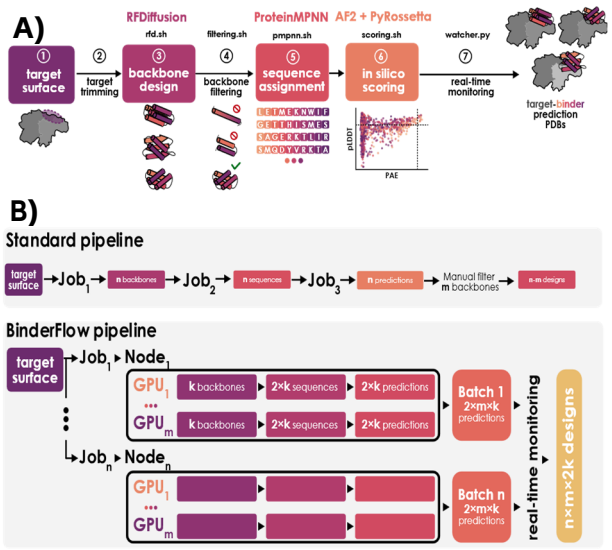


***Figure 1: (A) Overview of the BinderFlow pipeline**. After manual target selection and trimming (steps 1–2), the rest of the process—backbone design, filtering, sequence assignment, and in-silico scoring (steps 3–6)—is automated in small batches. Real-time campaign monitoring is also supported (step 7). **(B) Comparison of the standard workflow and BinderFlow**. The standard approach requires manual input/output handling and separate job execution at each step. BinderFlow automates and parallelizes end-to-end predictions, enabling efficient distribution across GPUs and HPC nodes.*

References:

[1] Watson, J.L., Juergens, D., Bennett, N.R. et al. De novo design of protein structure and function with RFdiffusion. Nature 620, 1089–1100 (2023).
[2] J. Dauparas et al., Robust deep learning–based protein sequence design using ProteinMPNN.Science378,49-56(2022).
[3] Bennett, N.R., Coventry, B., Goreshnik, I. et al. Improving de novo protein binder design with deep learning. Nat Commun 14, 2625 (2023)

(7) *In silico* design of polyspecific minibinders against snake venom phospholipase A_2 toxins

Cecilie Knudsen[1], Kasper H. Björnsson[1], Timothy P. Jenkins[1]

[1] Technical University of Denmark, Department of Bioengineering, Lyngby, Denmark

Cecknu@dtu.dk

Roughly 72% percent of the world's population lives in close proximity to at least one highly venomous snake species, and snakebites result in an estimated six million disability-adjusted life years per year [1]. This metric accounts for years of life lost due to either premature death, or years lived with some kind of morbidity, such as amputations, blindness, or extensive scarring due to the snakebite. Currently, the primary treatment of snakebite envenoming is antivenom based on antibodies that have been extracted from hyperimmunized animals [1]. These antivenoms save lives but have drawbacks including high production costs and the risk of inducing adverse reactions [1].

Here, we attempt to de novo design minibinders capable of neutralizing some of the most medically relevant snake venom toxins to help overcome some of the limitations in current treatment options. Specifically, we targeted snake venom toxin phospholipase A_2s (PLA_2s), which can be subdivided into enzymatically active and inactive forms based on point mutations in Asp49. We designed minibinders against all 378 available PLA_2 structures using RFdiffusion and ProteinMPNN and subsequently investigated their polyspecific potential via *in silico* crosspanning using Boltz-2 [2]. The broader cross-neutralization can be achieved, the fewer therapeutic proteins will be necessary, and the more affordable the treatment will be. Once designs with the required *in silico* polyspecificity have been generated, the proteins will be validated *in vitro* and *in vivo*. We envision that polyspecific proteins capable of neutralizing snake venom toxins can constitute a part of future snakebite treatments.

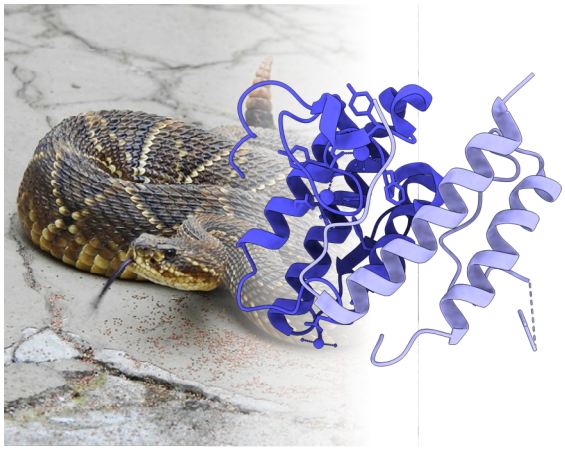


Figure 1: The South American rattlesnake and crotoxin. *Some rattlesnake venoms contain the potent neurotoxin crotoxin. Crotoxin consists of a heterodimer made up of PLA_2 homologues, and it's one of the 378 PLA_2-based toxins that are the targets of this campaign to design broadly neutralising proteins in silico. Image from [3] and PDB structure 3R0L.*

References:

[1] Gutiérrez J., et al. Snakebite envenoming. Nature Reviews: Disease primers 3, 17063, 2017. https://doi.org/10.1038/nrdp.2017.63

[2] S. Passaro et al. Boltz-2: Towards Accurate and Efficient Binding Affinity Prediction. *bioRxiv*, 2025. https://doi.org/10.1101/2025.06.14.659707

[3] https://commons.wikimedia.org/wiki/File:Crotalus_durissus.jpg

(8) Modeling Disease-Associate T-cells Distribution

Gal Passi[1*], Amir Weinfeld[1*], Dina Schneidman-Duhovny[1]

[1] The Hebrew University of Jerusalem, School of Computer Science and Engineering, Jerusalem, Israel
gal.passi@mail.huji.ac.il, dina.schneidman@mail.huji.ac.il, [*]Equal Contribution

T-cells play a central role in adaptive immunity and are implicated in numerous human pathologies. Their function relies primarily on the hypervariable complementarity-determining region 3 (CDR3), a short sequence (~12-18 amino acids) crucial for peptide recognition. Identifying and generating disease-associated T-cells holds significant potential in personalized medicine, yet despite advances in sequencing technologies, known disease-associated TCR sequences remain limited.

Here, we propose an unsupervised generative approach to model the distribution of CDR3 sequences associated with multiple sclerosis (MS), curating a dataset of 73 patients and 152 healthy subjects from TCRdb. Pathological CDR3 sequences are isolated from a repertoire of patients' blood samples by identifying sequences enriched in patients relative to a healthy background model. Sequences commonly found in healthy individuals are subsequently removed. To enhance the training dataset, we incorporate closely related sequences that differ by a single amino acid substitution from the naive set (Fig. 1A).

We demonstrate that a transformer model, trained on our curated dataset, effectively discriminates between patient and healthy controls (Fig. 1B), achieving good predictive capabilities (TPR=0.90, TNR=0.81, FPR=0.19, FNR=0.10) While a recent work by Zaslavsky et al (2025) demonstrated a diagnostic potential of using genetic data from B and T cells in other diseases, our sequence only approach employs a fully generative method. Confident positive sequences identified (Fig. 1B) are used to train a variational autoencoder (VAE), effectively capturing the MS-associated CDR3 sequence distribution. Sequence representations distinctly separate MS-associated sequences from naive circulating T-cells (Fig. 1C), capturing a disease-specific sub-distribution. Generative approaches have broad applications in designing personalized immunotherapies such as CAR-T cells. Our proposed pipeline can be readily adapted to other diseases.

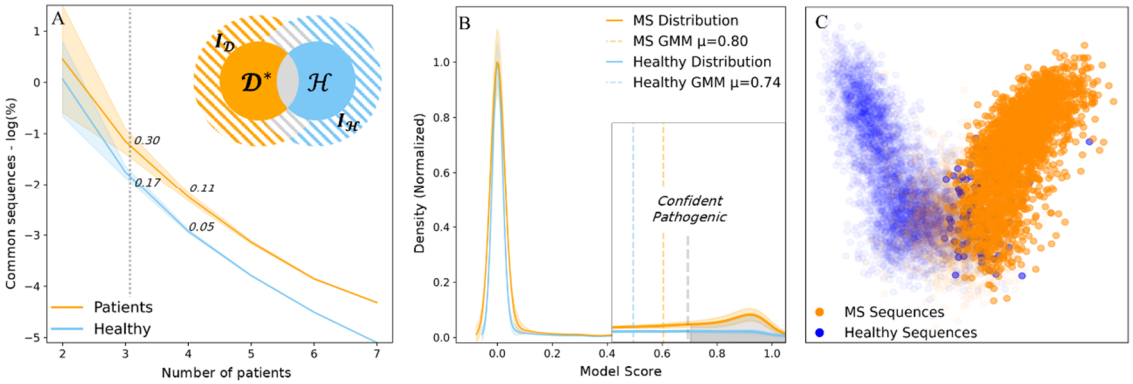


Figure 1*. A.** *Graphs represent the ratio of repeating sequences (y-axis log scale) in increasing numbers of patients and healthy subjects. Shades represent a 95% confidence interval. Dashed line - significant sequence enrichment difference (k*=3, p-value < 0.0001).* ***Top right *schematic representation of sequence processing pipeline. D* - sequences repeating in ≥k* patients. H sequences repeating in ≥3 healthy subjects. ID , IH - inflation process to sequences 1-edit distance away. Gray areas represent sequences intersecting with the healthy set which are removed during preprocessing.* ***B.*** *Aggregated probabilities of blood-drawn CDR3 sequences as predicted by our trained transformer on a test set consisting of eight MS patients and eleven healthy subjects. Scores closer to 1 are predicted to be related to MS. Orange and blue dotted lines represent the mean values of a Gaussian Mixture Model used for patient classification - trained on the train-set sequence distribution. Gray area represents confident MS-related sequences used to train the downstream VAE* ***C.*** *Dimensionality reduction (PCA) of test-set CDR3 sequences. Transparent dots represent low-confidence sequences (scores < 0.75) as predicted by our trained CDR3-transformer.*

(9) Language-Model guided Multi-Objective Lead Optimization

Franziska Geiger [1],

[1] Cradle

franzi@cradle.bio

Traditional protein lead optimization often resorts to optimizing one property at a time, resulting in many costly iterative cycles of lab testing. We present an automated framework to simultaneously optimize multiple properties, significantly reducing time and cost requirements for reaching desired profiles. We employ protein language models, specifically fine-tuned for each campaign with an active learning strategy, to generate optimized sequence libraries, while incorporating user-defined preferences and constraints. We demonstrate the success of the method in a three-round optimization campaign of an anti-SARS-CoV-2 VHH where thermostability, expression and binding affinity to two antigen variants were simultaneously improved. The top hit after three rounds of optimization showed 5.9 and 2 fold improvement on the two antigen variants respectively, 8.3ºC increase in thermostability and a 2.2 fold improvement in expression.

(10) Incorporating Biophysics into ML for Antibody Structure Prediction

Britnie J. Carpentier[1], Sen Wei[2], Laurent Ludwig[1], Jeffrey J. Gray[1]

[1] Johns Hopkins University, Department of Chemical and Biomolecular Engineering, Baltimore, USA

[2] Johns Hopkins University, Department of Biomedical Engineering, Baltimore, USA

bcarpen6@jh.edu

Alzheimer's disease (AD) remains a major therapeutic challenge due to its complex pathology and the restrictive blood-brain barrier (BBB), which limits the delivery of effective treatments to the central nervous system [1]. Multi-specific antibodies offer a promising strategy by targeting multiple pathological features while utilizing receptor-mediated transport mechanisms to cross the BBB [2]. However, the design of such antibodies with optimal binding affinity and specificity is hindered by limitations in current computational modeling techniques, due to limited experimental data, complexity of protein-protein interactions, as well as large full complex size. Other limitations to current protein structure prediction models, such as AlphaFold-2 and -3 [3,4] include reliance on multiple sequence alignment (MSA) which leads to failures in important cases such as antibodies and nanobodies [5].

To address these challenges, this project will develop a physics-informed deep learning framework that integrates molecular energy calculations into state-of-the-art protein modeling. First, an equivariant graph neural network will be constructed to learn individual intra- and intermolecular force terms from the Rosetta energy function (REF15) [6], named the Rosetta Energy Approximation (REA) network [7]. This energy-aware model will then be incorporated into established antibody structure prediction (IgFold) [8] and docking (DFMDock) [9] frameworks to assess improvements in structural accuracy, energy consistency, and interface prediction. Finally, this integrated approach will be extended to design and model multi-specific antibodies, predicting their architectures and binding partners with enhanced precision. The outcome will be a validated computational platform that advances the rational design of multi-specific antibody therapeutics for effective AD treatment.

References:

[1] Ibrahim MM, Gabr MT. Multitarget therapeutic strategies for alzheimer's disease. *Neural regeneration research*, 14(3):437-440, 2019.

[2] Edavettal S, Cejudo-Martin P, *et al*. Enhanced delivery of antibodies across the blood-brain barrier via TEMs with inherent receptor-mediated phagocytosis. *Med*, 3(12):860–882, 2022.

[3] Jumper J, Evans R, *et al*. Highly accurate protein structure prediction with alphafold. *Nature*, 596(7873):583–589,2021

[4] Abramson J, Adler J, *et al*. Accurate structure prediction of biomolecular interactions with alphafold 3. *Nature*, 630(8016):493–500, 2024.

[5] Hitawala FN, Gray JJ. What does alphafold3 learn about antigen and nanobody docking, and what remains unsolved? *bioRxiv*, pages 2024–09, 2025

[6] Alford RF, Leaver-Fay A, *et al*. The rosetta all-atom energy function for macromolecular modeling and design. *JCTC*, 13(6):3031–3048, 2017

[7] Wei S. Rosetta Energy Approximation Using a Machine Learning Approach. Master's thesis, Johns Hopkins University, 2024

[8] Ruffolo JA, Chu LS, Mahajan SP, Gray JJ. Fast, accurate antibody structure prediction from deep learning on massive set of natural antibodies. *Nature communications*, 14(1):2389, 2023.

[9] Chu LS, Sarma S, Gray JJ. Unified sampling and ranking for protein docking with dfmdock. *bioRxiv*, 2024.

(11) De novo design of protein assemblies with interlocked geometries

Liurong Chen, Bhoomika Basu Mallik, Seif Elmallawany, Ekaterina Prilevskaya, Aleksander Antanasjevich, Prof Dr. Alena Khmelinskaia

Ludwig Maximilian University of Munich, Munich, Germany

Liurong.Chen@lmu.de

Symmetry is widely explored by biological systems, enabling the formation of large complexes from simpler, smaller components, [1]. The integration of artificial intelligence with physical modeling has emerged as a leading strategy for designing such assemblies in the field of biomaterials engineering, [2]. Yet several aspects of current methods limit the structural and functional space that can be explored. For instance, designed protein nanoparticles are often more porous than their natural counterparts, and thus offer limited protection of and mechanisms for controlled release of small cargo clients. Here, I present a novel hierarchical symmetry-driven design approach of a new class of protein assemblies - interlocked two-compound polyhedra (Figure 1). This type of assembly architecture is expected to give rise to compartments with reduced porosity. Our experimental results demonstrate the successful assembly of an interlocked tetrahedron from a single protein building block and its unexpected responsiveness to pH changes. Overall, interlocked assemblies are a promising next-generation biomaterial with diverse yet unexplored potentials.

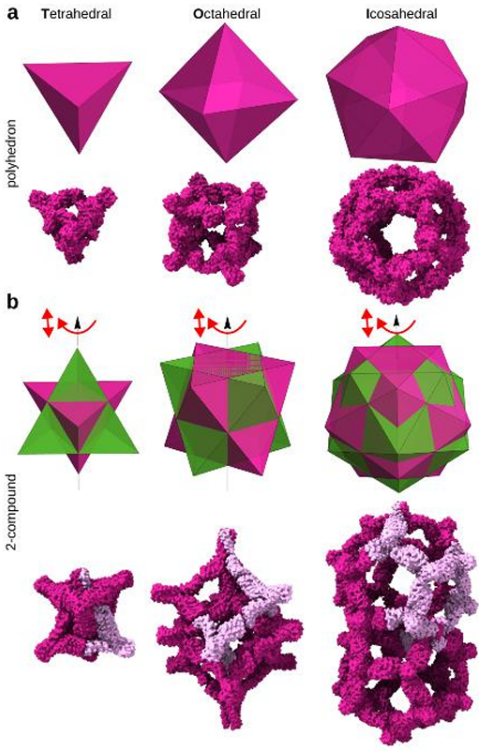


Figure 1: Interlocked protein assemblies. *(a) The Interlock structures are derived from T3, O3 and I3 symmetrical structures. (b) Two basic structures interlock with each other to form the final Interlock assembly. The lighter-colored regions are the asymmetric subunits.*

Reference:

[1] Goodsell, D.S., Olson, A.J.. Structural symmetry and protein function[J]. *Annu. Rev. Biophys. Biomol. Struct.* **29**, 105–153 (2000).

[2] Mallik, B. B., Stanislaw, J., Alawathurage, T. M. & Khmelinskaia, A. De Novo Design of Polyhedral Protein Assemblies: Before and After the AI Revolution. *ChemBioChem* **24**, e202300117 (2023).

(12) Protein Language Models for Peptide Binder Identification via Generation and Interaction Prediction

Bryan Renard[1,2], Jean-Yves Parfait[1], Catherine Michaux[2]

[1] Multitel, AI Department, Mons, Belgium
[2] University of Namur, Laboratoire de Chimie Physique des Biomolécules, UCPTS; Namur Institute of Structured Matter (NISM); Namur Research Institute for Life Sciences (NARILIS), Namur, Belgium

renard@multitel.be, bryan.renard@unamur.be

Peptide-based therapeutics offer high specificity and low toxicity, making them promising candidates for targeting proteins. In this work, we explore large language models (LLMs) pretrained on protein sequences to develop computational approaches for peptide binder identification.

We focus on two complementary tasks: **target-conditioned peptide generation** and **peptide-protein interaction prediction** (Fig.1). For the first, we fine-tune autoregressive and encoder-decoder models, specifically ProtGPT2 [1] and Ankh [2], to generate peptide sequences conditioned on a target protein sequence, learning binding-relevant sequence patterns. This approach formulates peptide design as a sequence-to-sequence generation problem. In parallel, we employ ESM2 [3], an encoder model, to predict peptide-protein interaction. By leveraging rich contextual representations, the model captures functional cues relevant to binding and enables rapid in silico screening and scoring of candidate binders. Additionally, we extend this approach to **binding site prediction** on the target protein, identifying potential interaction regions at the residue level. This dual prediction — peptide-protein interaction and binding site — provides insights into binding mechanisms.

This study highlights the potential of protein LLMs in accelerating peptide binder identification. The complementary approaches allow both exploratory design and rational screening. Importantly, this framework is scalable, requiring only sequence information, and generalizable across diverse protein targets without reliance on structural data or handcrafted features. As such, it offers a powerful approach for peptide-based drug design, with applications in therapeutic development, functional peptide engineering, and systematic exploration of protein–peptide interaction space.

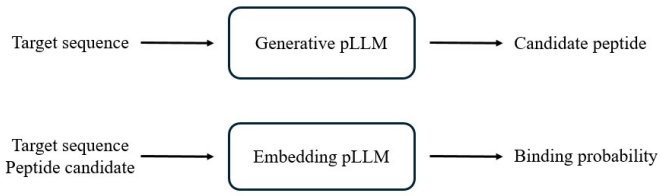


Figure 1: *Schematic of the peptide binder identification models combining target-conditioned generation and interaction prediction using protein-pretrained language models.*

References:

[1] N. Ferruz, S. Schmidt, and B. Höcker. ProtGPT2 is a deep unsupervised language model for protein design. *Nature Communications*, 13(1):4348, 2022.
[2] A. Elnaggar, H. Essam, et al. Ankh: Optimized protein language model unlocks general-purpose modelling. *bioRxiv*, https://doi.org/10.1101/2023.01.16.524265, 2023.
[3] Z. Lin, H. Akin, et al. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science*, 379(6637):1123–1130, 2023.

(13) Antiviral proteins by design: A broad spectrum strategy for pandemic preparedness

Diogo H. Silva[1], André Bagão[1], Pedro Moreira[1,2], Shuhao Xiao[3], Carlota Conceição[1], Carolina C. Buga[1,4], Mariana Parada[1], Rita Teixiera[1], Marta Miranda[4,5], Christine Cruz-Oliveira[4], Manuel N. Melo[1], Isabel Abreu[1], César Santiago[6], José M. Valpuesta[6], M. Alenquer[5], M. J. Amorim[5], Ana S. Veiga[4,7], Miguel A. R. B. Castanho[4,7], Iebe Rossey[8], Xavier Saelens[8], Bruno Correia[3], Cláudio M. Soares[1], João B. Vicente[1] and Diana Lousa[1]

[1]ITQB NOVA, Oeiras, Portugal. [2]Center of Biological Engeneering, School of Engineering, University of Minho, Braga, Portugal. [3]School of Life Sciences, EPFL, Lausanne, Switzerland. [4]GIMM, Lisboa, Portugal. [5]Católica Biomedical Research Centre, Lisboa, Portugal,[6]Spanish National Research Council (CSIC), Madrid, Spain, [7]Lisbon School of Medicine, University of Lisboa, Lisboa, Portugal, VIB-UGent Center for Medical Biotechnology

dlousa@itqb.unl.pt

As it is not possible to accurately predict which virus(es) will cause the next pandemic(s), it is of utmost importance to find new solutions that allow the effective targeting of a broad range of viruses and have easy and cost-effective discovery, development, production and validation.

We are using an innovative solution to tackle this challenge: a platform for the discovery, development, production and validation of evolvable and rapidly adaptable antivirals based on designed proteins of different types: helical miniproteins, nanobodies and monobodies (which are derived from the human fibronectin domain 3). To demonstrate the capacity of the platform to address a broad range of viruses, we have already successfully developed protein binders against two major pathogenic viruses as test cases: SARS-CoV-2 and respiratory syncytial virus (RSV). Using a protocol based on RF diffusion, Protein MPNN and biophysical based methods, including MD simulations, we were able to achieve >50% success rate as evaluated by the ability of the binders to bind to target and neutralize the virus at sub-micromolar concentrations. The best designed proteins for the two viruses have neutralization IC50 < 100 nM without any experimental refinement. Additionally, Cryo-EM structures and large-scale atomistic MD simulations enable a detailed characterization of how the designed binders interact with target, providing important clues for future design rounds. This work shows that protein design based strategies are a key asset in preparing for future pandemics.

(14) Validating Structural ML Predictors for TCR–pMHC Recognition

Laura Machado[1], Sebastian N. Deleuran[2], Morten Nielsen[2], Timothy P. Jenkins[1].
[1] Dept. of Bioengineering, Section for Biologics Engineering
[2] Dept. of Health Technology
Technical University of Denmark, Kongens Lyngby, Denmark

laugon@dtu.dk

Adaptive immunity relies on the ability of T cell receptors (TCRs) to recognize peptide–MHC (pMHC) complexes with high precision. This interaction functions as the ignition switch of adaptive immune responses, underpinning processes from pathogen clearance to cancer immunotherapy. However, predicting TCR–pMHC specificity remains one of the hardest open problems in computational immunology. Sequence-based approaches such as NetTCR-2.0/2.1 [1] have provided proof of concept but fail to generalize to unseen targets, while structural extensions (e.g. NetTCR-struct [2]) have not yet achieved reliable accuracy. These limitations stem from TCR diversity, scarce and noisy datasets, and the absence of frameworks tailored to immune receptor interactions.

While structural ML methods have succeeded in other receptor–ligand systems, TCR–pMHC recognition is uniquely difficult: interfaces are dominated by TCR–MHC contacts, often involving only a few peptide side chains, yet cross-reactivity is widespread. This makes it nearly impossible to distinguish productive from non-productive interfaces with current predictors (Fig. 1). To address this, we validated the highest-quality curated dataset of TCR–pMHC complexes (TCRvdb) [3] against the latest composite evaluation metrics recently introduced for protein–protein structural prediction [4]. These analyses yielded insightful results that provide proof of potential for applying ML-based structural prediction to TCR–pMHC interactions.

Embedding these structural metrics into domain-aware deep learning is emerging as a compelling strategy to achieve more interpretable and generalizable predictions for TCR–pMHC interactions. Applied to the uniquely difficult challenge of TCR recognition, this approach outlines a trajectory where even incremental advances could reshape biomedical research and revolutionize therapeutic development across disease areas.

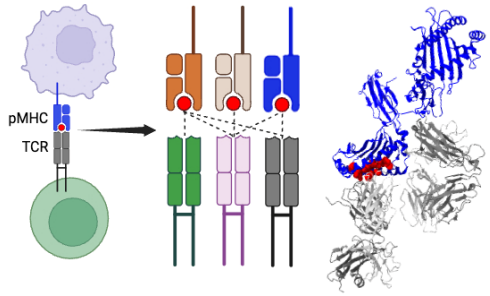


Figure 1: The challenge of TCR–pMHC prediction. *TCRs often engage primarily with the MHC surface while contacting only a few peptide side chains, yet exhibit broad cross-reactivity across diverse ligands. This structural complexity underlies the difficulty of distinguishing productive from non-productive interfaces and highlights the limits of current predictive models. (PDB ID: 5YXU).*

References:

[1] Montemurro A et al. NetTCR-2.0 enables accurate prediction of TCR-peptide binding by using paired TCRα and β sequence data. *Commun Biol*. 2021;4(1):1060.
[2] Deleuran SN et al. NetTCR-struc, a structure driven approach for prediction of TCR-pMHC interactions. *Front Immunol*. 2025;16:1616328.
[3] Messemaker M et al. A functionally validated TCR-pMHC database for TCR specificity model development. *bioRxiv*. 2025.
[4] Overath MD et al. Predicting experimental success in de novo binder design: A meta-analysis of 3,766 experimentally characterised binders. *bioRxiv*. 2025.

(15) Protein design by fragment-based chimeragenesis

Katharina Bachschwöller, Birte Höcker

University of Bayreuth, Department of Biochemistry, Bayreuth, Germany

katharina.bachschwoeller@uni-bayreuth.de

Protein domains are considered as the fundamental units of protein structure, function and evolution. They act as building blocks that play an important role in shaping the diversity observed in the protein landscape [1]. These domains themselves evolved from **ancestral fragments** that have been repurposed within protein folds across different structural and functional contexts, leading to differently folded protein domains that share conserved fragments. As such, the recombination of these fragments holds the potential to generate proteins with novel functions, presenting a starting point for rational protein design [2].

Therefore, we utilize the database Fuzzle (Fold Puzzle) [3] to identify evolutionary relevant and potentially engineerable conserved fragments to then combine them to chimeric proteins. To optimize interfaces and improve soluble expression, the sequences are redesigned utilizing different methods, including PROSS [4], ProteinMPNN [5] and LigandMPNN [6].

Such newly designed chimeras expressed solubly in *E.coli* and in high yields. Biochemical characterization indicates correct folding and high thermal stability.

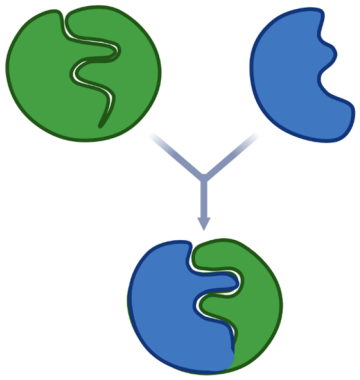


Figure 1: Principle of fragment-based chimeragenesis. *New protein chimeras can be built by recombining fragments from different folds. The chimera comprises parts of different proteins that can fold and retain function.*

References:

[1] Alva V, Remmert M, Biegert A, *et al.* A galaxy of folds. Protein Sci. 2010 Jan;19(1):124-30
[2] Höcker B. Engineering chimaeric proteins from fold fragments: 'hopeful monsters' in protein design. Biochem Soc Trans. 2013 Oct;41(5):1137-40
[3] Ferruz N, Michel F, Lobos F, *et al*. Fuzzle 2.0: Ligand Binding in Natural Protein Building Blocks. Front Mol Biosci. 2021 Aug 18;8:715972
[4] Goldenzweig A, Goldsmith M, Hill SE, *et al*. Automated Structure- and Sequence-Based Design of Proteins for High Bacterial Expression and Stability. Mol Cell. 2016 Jul 21;63(2):337-346.
[5] Dauparas J, Anishchenko I, Bennett N, *et al.* Robust deep learning-based protein sequence design using ProteinMPNN. Science. 2022 Oct 7;378(6615):49-56.
[6] Dauparas J, Lee GR, Pecoraro R, *et al.* Atomic context-conditioned protein sequence design using LigandMPNN. Nat Methods. 2025 Apr;22(4):717-723.

(16) Computational Design of Pseudosymmetric Protein Assemblies

Siddhant Sharma[1], Florian Praetorius[1]

[1] Institute of Science and Technology Austria, Klosterneuburg, 3400 Austria

siddhant.sharma@ista.ac.at

The de novo design of modular protein assemblies with strict interaction specificity is an interesting goal for building reconfigurable multi-subunit protein networks. My research focuses on constructing bivalent and trivalent protein connectors derived from β sheet-mediated heterodimers which were engineered through implicit negative design principles, and validated for spontaneous assembly from independently expressed monomers [1].

I have established a computational pipeline centered on generating C2 and C3 pseudosymmetric complexes from existing heterodimers through RPXdock [2], which provide a sequence-independent rigid-body protein docking across a wide range of symmetric architectures [3]. The heterodimeric C2 and C3 docked complexes are decomposed into individual chains, and the resulting internal components are fused into a single-chain using motif-scaffolding and sequence inpainting with RFDiffusion. Custom pyrosetta scripts are utilized for structure manipulation and filtering the glycine-only backbones. Then, the original heterodimer chain sequence is threaded onto the generated single chain assembly, and the designed linkers and their interacting contacts are subjected to a SolubleMPNN sequence design. Structure prediction is carried through AlphaFold2-multimer via ColabFold-1.5.5-v3. The resulting models are evaluated using per-residue pLDDT scores and predicted aligned error (PAE) between chains, and further filtered for homodimerization propensity and sequence diversity. Experimental screening including IMAC protein purification are performed via E. coli expression and SEC-based validation for the designed constructs. Furthermore, we are implementing SALAD (sparse all-atom denoising models for protein backbone generation) [4] for symmetric multi-motif scaffolding to alternatively generate symmetrical trivalent protein connectors.

I am working towards demonstrating a successful generation of bivalent and trivalent connectors capable of mediating specific protein-protein interactions while maintaining monomeric stability. This systematic approach establishes a robust pipeline for enabling construction of linearly arranged hetero-oligomers, and branched complexes resulting in enabling the construction of reconfigurable networks for designer nanomaterials.

References:

[1] Sahtoe, D. D., Praetorius, F., Courbet, A., Hsia, Y., Wicky, B. I. M., Edman, N. I., … Baker, D. (2022). Reconfigurable asymmetric protein assemblies through implicit negative design. *Science*, 375(6578), eabj7662.

[2] Sheffler, W., Yang, E. C., Dowling, Q., Hsia, Y., Fries, C. N., Stanislaw, J., Langowski, M. D., Brandys, M., Li, Z., Skotheim, R., Borst, A. J., Khmelinskaia, A., King, N. P., & Baker, D. (2023). Fast and versatile sequence-independent protein docking for nanomaterials design using RPXDock. *PLoS Comput Biol*, 19(5), e1010680.

[3] Mallik, B. B., Stanislaw, J., Alawathurage, T. M., & Khmelinskaia, A. (2023). De Novo Design of Polyhedral Protein Assemblies: Before and After the AI Revolution. *ChemBioChem*, 24(15).

[4] Jendrusch, M., & Korbel, J. O. (2025). Efficient protein structure generation with sparse denoising models. https://doi.org/10.1101/2025.01.31.635780

(17) Multi-state Design of De Novo Copper Amino Oxidases

Tim Seidl[1], Sigrid Kaltenbrunner[1], Adrian Tripp[1], Gustav Oberdorfer[1]

[1] Graz University of Technology, Institute of Biochemistry, Graz, Austria

tim.seidl@student.tugraz.at

Protein design is a dynamic and rapidly evolving field. De novo designed proteins offer a wide range of potential applications in biomedicine and biotechnology. However, beyond generating functional designs, it is equally important to explore and expand the opportunities and limitations of current design methods. One way to do this is by attempting to mimic enzymes with complex reaction mechanisms.

This was pursued by targeting the reaction mechanism of copper amine oxidases (CuAOs). These enzymes require **autocatalytic maturation** of a topaquinone (TPQ) cofactor to perform their catalytic function. Additionally, the TPQ residue must be flexible, as the conformations required for maturation and catalysis differ. De novo CuAOs were designed using **RiffDiff**, a novel protein design workflow for one-shot enzyme design [1]. The initial designs demonstrated successful copper binding and were further evaluated for cofactor maturation, with limited but promising results.

The aim of this project was to enhance the TPQ formation. Therefore, the most promising de novo variants were redesigned using a **two-state approach**, incorporating both conformations—maturation and catalysis—as inputs. New sequences were generated using **ProteinMPNN**, utilizing its symmetry option, originally intended for oligomer design, to encode both states. These sequences were then assessed using **RosettaRelax**, **ESMFold**, and **AlphaFold3** to predict structures and evaluate TPQ flexibility.

The top-scoring variants were selected for expression and will be tested for cofactor maturation using intact mass spectrometry and photometric assays. Furthermore, promising candidates will also be crystallized to determine whether the designed two-state conformations are realized in the final protein structure.

References:

[1] M. Braun *et al*. “Computational enzyme design by catalytic motif scaffolding” in: *bioRxiv*, June 2025, doi: https://doi.org/10.1101/2024.08.02.606416.

(18) Multi-state Protein Design with DynamicMPNN

Alex Abrudan[1], Sebastian Pujalte Ojeda[1], Chaitanya K. Joshi[2], Matthew Greenig[1], Felipe Engelberger[3], Alena Khmelinskaia[4], Jens Meiler[3], Michele Vendruscolo[1], Tuomas P. J. Knowles[1]

[1]Yusuf Hamied Department of Chemistry, University of Cambridge, Cambridge, UK
[2]Department of Computer Science and Technology, University of Cambridge, Cambridge, UK
[3]Institute for Drug Discovery, Leipzig University, Leipzig, Germany
[4]Department of Chemistry, Ludwig Maximilians University, Munich, Germany

aca41@cam.ac.uk, sp2120@cam.ac.uk

Structural biology has long been dominated by **the one sequence, one structure, one function** paradigm, yet many critical biological processes depend on proteins that adopt multiple conformational states. Existing multi-state design approaches often rely on post-hoc aggregation of single-state predictions, with poor experimental success rates compared to single-state design.

We introduce **DynamicMPNN** [1], an end-to-end deep learning model for multi-state protein design that learns a single unified sequence distribution conditioned on a set of conformers. To that end, we construct a new ML-ready dataset of 46,033 conformational pairs using the sequence redundancy in the PDB. We also introduce a novel data processing pipeline to handle heterogeneous sequences across aligned structures. To move beyond simple sequence recovery metrics, we propose a rigorous multi-state self-consistency benchmark based on AlphaFold 3 [2], using target structures as templates and decoy structures for normalization. Evaluated on a challenging benchmark of metamorphic, hinge, and transporter proteins, **DynamicMPNN outperforms ProteinMPNN MSD** [3] by up to 25% on decoy-normalized RMSD and by 12% on sequence recovery, proposing an effective paradigm for designing functional, dynamic proteins.

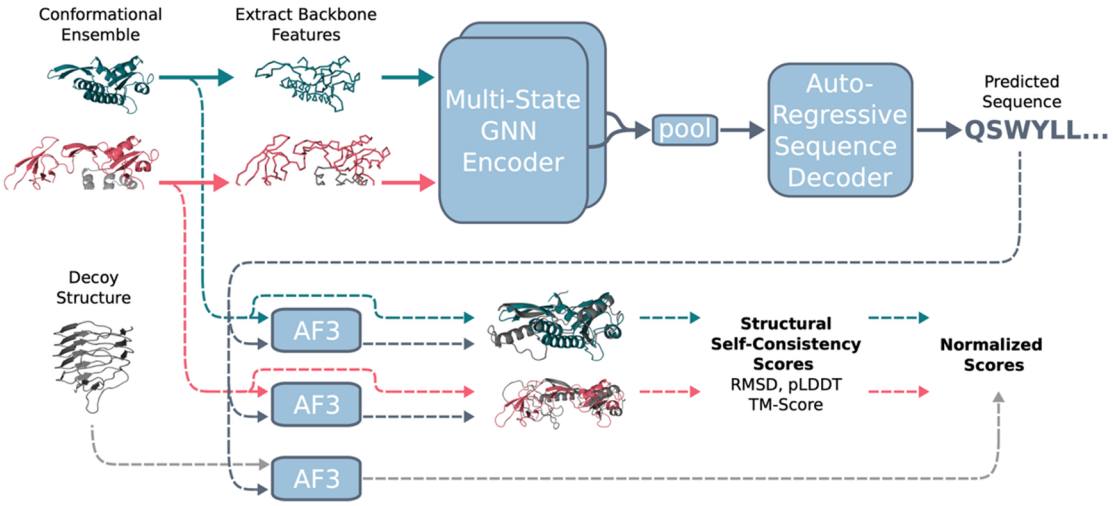


Figure 1: DynamicMPNN for multi-state protein design. Schematic of the DynamicMPNN pipeline. A multi-state GNN encoder processes a conformational ensemble to generate a pooled latent representation. An autoregressive decoder then predicts a sequence compatible with all input states. Designs are evaluated using a template-based AlphaFold 3 pipeline against both target and decoy structures to assess structural self-consistency.

References: [1] A. Abrudan, S. P. Ojeda, et al. Multi-state protein design with DynamicMPNN. *arXiv:2507.21938*, 2025. **[2]** J. Abramson, et al. Accurate structure prediction of biomolecular interactions with Alphafold 3. *Nature*, 2024. **[3]** Dauparas, J. et al. Robust deep learning-based protein sequence design using ProteinMPNN. *Science*, 378(6615), 49-56 (2022).

(19) Designing DNA-binding Proteins for Nucleic Acid-templated Protein Assemblies

Martin Fischlechner[1], Florian Praetorius[1]

[1] Institute of Science and Technology Austria (ISTA), 3400 Klosterneuburg, Austria.

martin.fischlechner@ista.ac.at
florian.praetorius@ista.ac.at

Aim of this work is to create building blocks for molecular scaffolding applications, which use DNA nanotechnology to control size, shape and composition of protein assemblies that present functional protein domains with molecular precision.

DNA nanotechnology [1] enables the assembly of structures up to the micrometer length scale, but the attachment of proteins usually relies on flexible linkers that significantly reduce the precision of spatial arrangements. Employing functionalized DNA-binding proteins in such assemblies should allow for highly defined display, while maintaining the design advantages that nucleic acid scaffolds afford.

The designed DNA-binders are composed of modified TALE-repeats (TALE = Transcription-Activator-Like-Effectors [2,3]) and bridging connectors. TALE-repeats feature programmable sequence-specific DNA binding, and functionality can be introduced via embedding functional motifs into connectors. Relative to the connectors, the second TALE-repeat on both sides is fixed in sequence, which facilitates combinatorial cloning and expression of constructs, while at the same time allowing for the design of custom connector-repeat connections. TALE-repeats are scored for functionality, rigidity, and low self-interaction. DNA-binding and sequence specificity is tested with HPLC and fluorescence polarization spectroscopy.

References:

[1] F. Hong, F. Zhang, Y. Liu, H. Yan. DNA Origami: Scaffolds for Creating Higher Order Structures. *Chem. Rev.*, 117:12584–12640, 2017.

[2] A. Nga-Sze Mak, P. Bradley, R. A. Cernadas, A. J. Bogdanove, B. L. Stoddard. The Crystal Structure of TAL Effector PthXo1 Bound to Its DNA Target. *Science*, 335:716-719, 2012.

[3] F. Praetorius, H. Dietz. Self-assembly of genetically encoded DNA-protein hybrid nanoscale shapes. *Science*, 355:eaam5488, 2017.

(20) Establishment and validation of a computational vaccine development pipeline for emerging viral diseases – a case study on Nipah Virus

Anna Nobis[1], Antonia Sophia Peter[1], Sascha Wagner[1], Jakob R. Riccabona[1,2], Moritz Ertelt[1,2], Max Beining[1,3], Dominic Rieger[1], Franz Dietzmeyer[1], Roman Trommler[1], Corinna Schüß[1], Hannes Junker[1], Birke Brumme[1], Christina Lange[1], Jens Meiler[1,2,4], Clara T. Schoeder[1,2]

[1] Institute for Drug Discovery, Leipzig University, Faculty of Medicine, Leipzig, Germany
[2] Center for Scalable Data Analytics and Artificial Intelligence ScaDS.AI, Dresden/Leipzig, Germany
[3] School of Embedded Composite Artificial Intelligence SECAI, Dresden/Leipzig, Germany
[4] Department of Chemistry and Center of Structural Biology, Vanderbilt University, Nashville, TN, USA

Anna.nobis@medizin.uni-leipzig.de

The COVID-19 pandemic highlighted the critical need for robust pandemic preparedness in global health systems. Rapid and cost-effective vaccine production is crucial during crises, achievable through the integration of computational methods into vaccine research. Therefore we created a computational vaccine design pipeline, applied it to different viruses, and stabilized their fusion proteins using insertion of disulfide bridges and single-point mutations. This study focuses on Henipaviruses, members of the Paramyxoviridae family with mortality rates exceeding 70%, surpassing the pathogenicity of related viruses like mumps and measles. Leveraging AlphaFold, ProteinMPNN and Rosetta, we identified vaccine candidates in a quick and cost-effectively way. Our computational strategy not only enables rapid development of novel vaccine candidates against Henipaviruses based on predicted protein structures within a few weeks but also establishes a generalizable framework for vaccine development.

(21) De novo Motif Scaffolding to Dissect FtsA–FtsW and FtsA–FtsN Interaction.

Luísa Rodrigues[1] , Florian Praetorius[1] & Martin Loose[1]

[1]Institute of Science and Technology Austria, 3400 Klosterneuburg, Austria.

luisa.rodrigues@ista.ac.at

Bacterial cytokinesis depends on the divisome, a large membrane-embedded complex whose core peptidoglycan synthases FtsW–FtsI and regulatory FtsQ–FtsB–FtsL subcomplex span the cell envelope, yet high-resolution insights into individual protein–protein contacts remain elusive [1] . FtsA serves as the critical hub that tethers the Z-ring to the membrane while also interacting with the cytoplasmic tail of FtsN, and recruiting downstream partners, including FtsW [2]. However, there is no soluble platform to dissect these contacts in isolation [3]. To overcome this barrier, we employed a de novo motif-scaffolding pipeline in which we isolated the minimal intracellular epitopes of FtsW and FtsN responsible for FtsA binding, enabling computational engineering and downstream biophysical validation.

Using RFdiffusion, we build structural backbones around the cytoplasmic loops of FtsW implicated in FtsA recruitment, then optimize amino-acid sequences with ProteinMPNN. In addition, we graft the minimal FtsN binding motif onto the same scaffold to generate bifunctional binders capable of engaging both partners. All designs are filtered by AlphaFold2 for local pLDDT > 90, predicted alignment error (pAE) < 8 Å, and RMSD < 2 Å to their native motifs. We express the top candidates in *E. coli*, purify them by His-Tag and SEC and performing fluorescence-based binding assays with Cy3- and Cy5-labeled full-length FtsA to determine affinities for each motif.

These binding scaffolds offer a versatile platform for dissection of divisome assembly, enabling the study of particular interactions without the complexity of the whole divisome.

References:

[1] Käshammer L, van den Ent F, Jeffery M, Jean NL, Hale VL, Löwe J. Cryo-EM structure of the bacterial divisome core complex and antibiotic target FtsWIQBL. Nat Microbiol. 2023 June;8(6):1149–59.

[2] Radler P & Loose M. A dynamic duo: Understanding the roles of FtsZ and FtsA for Escherichia coli cell division. *Mol. Microbiol.* (2024)

[3] Radler P, Baranova N, Caldas P, Sommer C, López-Pelegrín M, Michalik D, et al. In vitro reconstitution of Escherichia coli divisome activation. Nat Commun. 2022 May 12;13(1):2635.

(22) Leveraging structure prediction for mechanistic hypothesis generation in molecular and cell biology

Viktor Sadilek[1], **Florian Praetorius**[1]

[1]Institute of Science and Technology Austria

viktor.sadilek@ista.ac.at

Much of the mechanistic complexity in molecular and cell biology stems from the vast network of specific biomolecular interactions, many of which remain hidden ("dark matter") due to genetic redundancy and the limited sensitivity of classical forward genetic screens and biochemical assays [1]. Furthermore weak effects frequently lie beyond the practical reach of medium and high throughput methods unless guided by prior mechanistic insight or serendipitous observations.

AlphaFold-2 currently enables the prediction of multimeric protein assemblies which has given rise to in-silico screening pipelines such as AlphaPulldown [3]. However, these AlphaFold-2 based approaches remain restricted to protein–protein interactions.

Here, we explore Boltz-based protein structure prediction (Boltz-1 and Boltz-2) [4] as an in silico hypothesis generator to map potential interaction landscapes across diverse molecular classes.

We present preliminary benchmarks across multiple use cases: modeling DNA sequence specificity, predicting protein–protein interactomes and complex assemblies, conducting "in silico pulldown" experiments, and screening plant hormone–receptor pairs against proteomes of interest. Early results demonstrate that Boltz captures some key features of DNA binding but remains blind to the nuances underpinning sequence specificity, suggests novel interactors for known complexes, and identifies candidate hormone-binding proteins. Together, the in-silico observed interactions begin mapping putative mechanisms which warrant experimental follow-up.

References:

[1] Navarro-Quiles C, Lup SD, Muñoz-Nortes T, Candela H, Micol JL. The genetic and molecular basis of haploinsufficiency in flowering plants. Trends Plant Sci. 2024;29(1):72-85.

[2] Yu D, Chojnowski G, Rosenthal M, Kosinski J. AlphaPulldown—a python package for protein–protein interaction screens using AlphaFold-Multimer. Bioinformatics. 2023;39(1):btac749.

[3] Wohlwend J, Corso G, Passaro S, Reveiz M, Leidal K, Swiderski W, Portnoi T, Chinn I, Silterra J, Jaakkola T, Barzilay R. Boltz-1: Democratizing biomolecular interaction modeling. *bioRxiv*. 2024.

(23) Design and Characterization of a Binder Targeting the Tom22 MoRE to Understand Mitochondrial Protein Import Mechanisms

Maria Clara Miserendino[1], Florian Praetorius[1], Paul Schanda[1]

[1] Institute of Science and Technology, Klosterneuburg, Austria

mmiseren@ist.ac.at

Tom22 is an essential subunit of the Translocase of the Outer Membrane (TOM) complex, which serves as the main entry point for nuclear-encoded proteins into mitochondria. While its membrane-embedded regions are well characterized and known to support complex structure stability, the role of its large cytosolic domain remains unresolved.

To probe the role of a conserved segment within the cytosolic region of Tom22, we designed a specific binder targeting a short, transient α-helix, also referred to as a Molecular Recognition Element (MoRE). Using BindCraft v1.5, we generated a diverse set of de novo binders against an idealized helical model of the MoRE, with and without hotspot constraints. We applied Rosetta-based structural filters to prioritize candidates with favorable secondary structure and presence of charge–charge interactions at the interface.

Structural predictions were used to evaluate binder-target interactions, potential off-target binding, and homodimerization propensity. Based on these evaluations, a subset of binders was selected for experimental validation. Selected candidates were successfully expressed, purified, and characterized. Preliminary results indicate that several binders interact with Tom22, while showing no detectable interaction with predicted off-targets.

(24) Assessing the Quality of AlphaFold Structures in Single-Point Mutation Rosetta ddG Prediction

Harper Raiken[1], Beatriz Antunes, Aleksandra Panfilova, Amelie Stein

[1] University of Copenhagen, Department of Biology, Copenhagen, Denmark

zqt390@ku.dk

This study seeks to benchmark the performance of structures generated by AlphaFold [1] for the purpose of interaction ΔΔG prediction using Rosetta. As a control, we used ΔΔG calculations on structures available in the SKEMPI v2.0 database [2]. Predicted ΔΔGs were compared to SKEMPI scores (converted from binding affinities). AlphaFold structures were generated using FASTA sequences corresponding to structures in the SKEMPI v2.0 database. These structures were then subject to mutagenesis and passed to the cartesian Rosetta ΔΔG pipeline [3]. We find that AlphaFold-Multimer [4] and AlphaFold3 [5] structures produce similar quality predictions to those generated using crystallographic structures. Furthermore, the implementation of InterfaceAnalyzer provides more accurate binding ΔΔG predictions when compared to the pipeline alone.

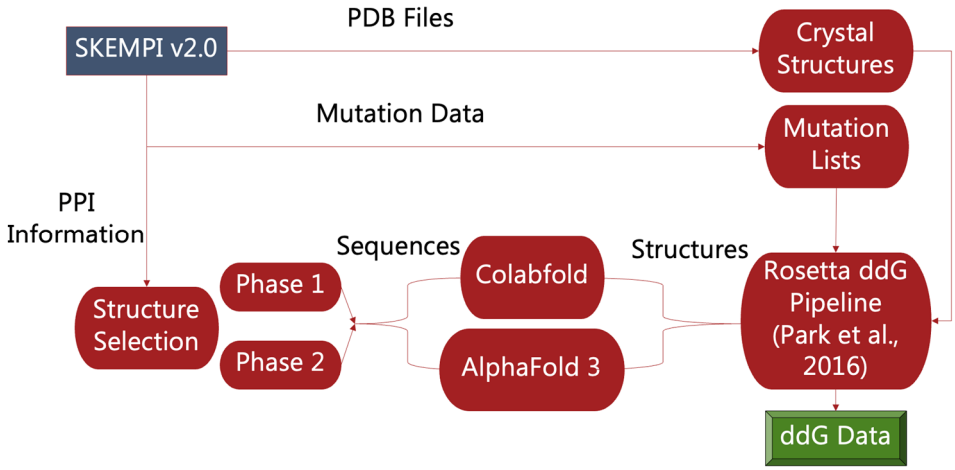


Figure 1: A depiction of the workflow used in the course of the study listed. Data from SKEMPI was used to collate crystal structures from the PDB and generate structures using AlphaFold. These structures were then subjected to mutagenesis to compare to SKEMPI scores and obtain accuracy calculations.

References:

[1] Jankauskaitė, J., Jiménez-García, B., Dapkūnas, J., et al., 2019. SKEMPI 2.0: an updated benchmark of changes in protein–protein binding energy, kinetics and thermodynamics upon mutation. Bioinformatics 35, 462–469.

[2] Jumper, J., Evans, R., Pritzel, A., et al., 2021. Highly accurate protein structure prediction with AlphaFold. Nature 596, 583–589.

[3] Park, H., Bradley, P., Greisen, P., et al., 2016. Simultaneous optimization of biomolecular energy function on features from small molecules and macromolecules. J Chem Theory Comput 12, 6201–6212.

[4] Evans, R., O'Neill, M., Pritzel, A., et al., 2021. Protein complex prediction with AlphaFold-Multimer.

[5] Abramson, J., Adler, J., Dunger, J., et al., 2024. Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature 630, 493–500.

(25) PEPE: Scalable generation of protein language model representations

Jahn Zhong[1,*], Niccolò Cardente[1], Habib Bashour[1], Maria Francesca Abbate[1], Geir Kjetil Sandve[2,3] and Victor Greiff[1,4,*]

[1] Department of Immunology, University of Oslo, Oslo, Norway
[2] Scientific Computing and Machine Learning section, Department of Informatics, University of Oslo, Oslo, Norway
[3] UiORealArt Convergence Environment, University of Oslo, Norway
[4] Imprint Labs, LLC. New York, NY, USA

[*] Corresponding authors

jahn.zhong@medisin.uio.no, victor.greiff@medisin.uio.no

Protein language models (PLMs) have demonstrated significant potential in capturing the complex interaction patterns between amino acids in protein sequences. These models, trained on large datasets of protein sequences, can generate embeddings, high-dimensional numerical representations that encode valuable information about the structure, function, and evolution of proteins[1]–[3]. However, conventional usage has largely been based on arbitrarily determined variable sets (embedding modes), including the choice of embedding layer, pooling method, and padding, which can potentially lead to a suboptimal representation with low information content for a given downstream task[4]–[6]. Currently state-of-the-art (SOTA) embedding extraction tools, such as PLMFit [7], are based on a sequential process that is limited by inefficiencies in both space (memory) and time (computation). (i) Accumulating all outputs in memory and writing them to disk in a single operation leads to a memory bottleneck. (ii) Additionally, repeated embedding of the same sequence to extract different embedding modes introduces unnecessary computational overhead and reduces throughput significantly.

Here, we present PEPE (Parallel Extraction for Protein Embeddings), a command-line tool designed for high-throughput multi-modal protein sequence embedding extraction. We show that PEPE's parallel process outperforms the previous SOTA method, in terms of total run time by several fold. We also demonstrate that SOTA's peak memory usage increases with scale and fails once available memory is exceeded, whereas PEPE's peak memory usage is independent of output size, allowing multi-modal embedding at previously unfeasible scales. PEPE supports a wide range of publicly available and custom protein language models and provides a simple command-line interface for researchers. PEPE thereby enables the generation of large-scale protein embedding datasets, facilitating the identification of optimal protein embedding settings for downstream analyses without requiring additional resources for fine-tuning.

References:

[1] Z. Lin *et al.*, "Evolutionary-scale prediction of atomic-level protein structure with a language model," *Science*, vol. 379, no. 6637, pp. 1123–1130, Mar. 2023.
[2] J. Barton, J. D. Galson, and J. Leem, "Enhancing antibody language models with structural information," *bioRxiv*, 04-Jan-2024.
[3] A. Villegas-Morcillo, A. M. Gomez, and V. Sanchez, "An analysis of protein language model embeddings for fold prediction," *Brief. Bioinform.*, vol. 23, no. 3, p. bbac142, May 2022.
[4] N. Deutschmann, A. Pelissier, A. Weber, S. Gao, J. Bogojeska, and M. R. Martínez, "Do domain-specific protein language models outperform general models on immunology-related tasks?," *Immunoinformatics (Amst.)*, vol. 14, no. 100036, p. 100036, Jun. 2024.
[5] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?," *arXiv [cs.LG]*, 06-Nov-2014.
[6] M. E. M. Gonzales, J. C. Ureta, and A. M. S. Shrestha, "Protein embeddings improve phage-host interaction prediction," *PLoS One*, vol. 18, no. 7, p. e0289030, Jul. 2023.
[7] T. Bikias, E. Stamkopoulos, and S. T. Reddy, "PLMFit: benchmarking transfer learning with protein language models for protein engineering," *Brief. Bioinform.*, vol. 26, no. 4, Jul. 2025.

(26) Designing Helical Bundles to Probe the CISS Effect

Paul Kluge[1], **Georg Künze**[1]

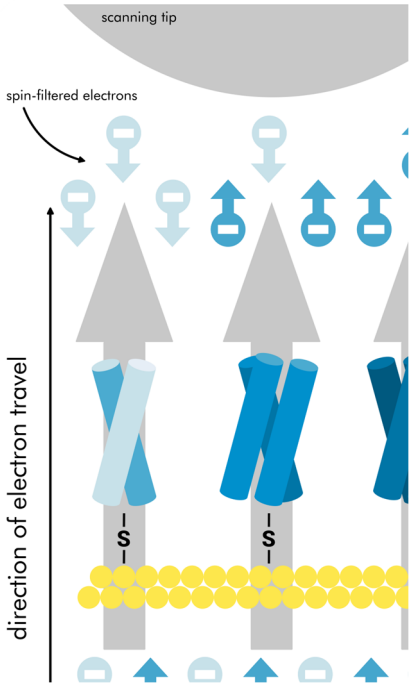
[1] Institute for Drug Discovery, Medical Faculty, University of Leipzig

paul.kluge@uni-leipzig.de

Electrons can travel through peptide backbones. Their movement is influenced by the chiral nature of the amino acids and the global structure of the peptide. The magnetic environment leads to a preferential transport of one electron spin state over the other. This phenomenon is called **chirality induced spin selectivity** (CISS). In this project we design novel helical bundles to explore how changes in the quaternary structure, such as superhelix twist, and parallelity influence CISS. We aim to expand the structural diversity of helical bundles by using rational and machine-learning tools.

We sourced amino acid patterns from external databases to improve the predictive power of ProteinMPNN, while harnessing the model to generate novel sequences. This approach enables exploration of previously unused regions of sequence–structure space for helical bundles.

The bundles will be characterized by STM, MOKE spectrosopy, and other techniques to probe their magnetic properties. Quantum mechanical simulations will allow us to gain deeper insight into the electron transport mechanisms.



(27) *De novo* protein design for efficient solar energy conversion

Louise Reimann[1], **Elisabet Romero**[1]

[1] Institute of Chemical Research of Catalonia, ICIQ, Av. Països Catalans, 16, 43007 Tarragona, Spain.

lkommers@iciq.es

Despite significant advances in protein engineering, the fundamental issue regarding artificial photosynthesis to develop the charge separation process, and to obtain precise control over chromophore-protein assemblies, remains a key obstacle. This research project is motivated by the goal of designing new photosynthetic proteins to enable sustainable energy solutions. The protein design needs to approach the right distance and angles between the chromophores to tune excitonic coupling and control absorption and excited state dynamics.[1] Furthermore, the photosynthetic proteins need to perform two fundamental processes: excitation energy transfer and charge separation.[1] *De novo* design is an attractive approach for constructing synthetic proteins with predetermined structures and functions, including the required chromophore-chromophore and chromophore-protein interactions.[2] Our research is focused on the design and modulation of different synthetic proteins, which could include different α-helices with specific position of the histidine binding site allowing the modulation of the chromophore's energy. Here we present six alpha-helices *de novo* protein with 4 binding sites, 2 at the top and 2 at the bottom (Fig. 1.b.), allowing the formation of chromophore-protein assemblies with 4 chromophores inside the protein, displaying an 11 nm red-shift in their absorption maximum compared to the maxima in organic solvent, together with a progressive broadening of the red tail at 700 nm with increasing the number of chromophores inserted (Fig. 1.c.). These observations suggest the stabilization of the chromophore's excited state by the protein, which could also be associated with excitonic interactions due to the dimer formation inside the protein and it is evidenced by circular dichroism spectroscopy. The absorbance difference signal centered at 669 nm has a partially conservative shape (Fig. 1.d.), which indicates the formation of the excitonically coupled dimers in, at least, a fraction of chromophore-protein assemblies. This result represents a first step towards the design of new solar-energy conversion systems.

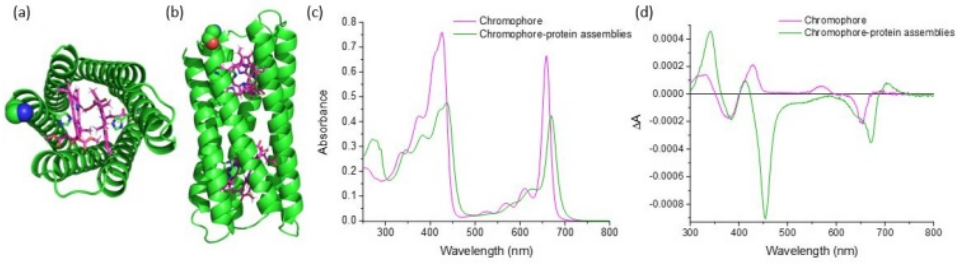


Figure 1. *(a) Top view of de novo chromophore-protein assembly. (b) Side view of de novo chromophore-protein assembly with 4 chromophores bound. (c) Absorption spectra of the chromophore in methanol (magenta) and the chromophore-protein assemblies (green) normalized by the chromophore concentration (20 μM), with the assemblies displaying an area difference of 0.90 times that of the chromophore in methanol. (d) Circular dichroism of chromophore-protein assemblies at 20 μM protein concentration.*

References:

[1] E. Romero, V. I. Novoderezhkin, R. van Grondelle, *Nature* **2017**, 543, 355.

[2] C.A. Floudas, H.K. Fung, S.R. McAllister, M. Mönnigmann, R. Rajgaria, *Chem. Eng. Sc.* **2006,** *61*, 966.

(28) Computational design of ligand-binding sites with increased pre-organization in *de novo* NTF2-like domains

Joana Albi-Puig [1,2], Marta Nadal[1], Paula M. Garcia-Franco[3], Sonia Vega[3], Adrián Velázquez-Campoy[3], Enrique Marcos[1]

[1] Protein Design and Modeling Lab, Department of Structural and Molecular Biology, Molecular Biology Institute of Barcelona (IBMB), CSIC, Baldiri Reixac 10, 08028 Barcelona, Spain
[2] Doctoral Program in Biotechnology, University of Barcelona, 08028 Barcelona
[3] Institute for Biocomputation and Physics of Complex Systems (BIFI), University of Zaragoza, Zaragoza, Spain

japcri@ibmb.csic.es

NTF2-like proteins offer a compact structure with internal cavities, making them attractive scaffolds for designing novel small-molecule binders and enzymes [1]. However, as creating ligand-binding pockets can often compromise folding stability. The de novo design of ligand-binding proteins remains challenging, with typically low success rates that require extensive experimental screening [2].

We present an automated **computational pipeline** that integrates deep learning and physics-based methods to design *de novo* proteins for **small-molecule binding** - involving molecular docking (DiffDock and PatchDock); sequence optimization (LigandMPNN); and *in silico* structural validation with (Rosetta and AlphaFold). The pipeline has been implemented on NTF2-like scaffolds developed in our lab, which were designed with enhanced structural stability through hydrophobic core expansion, while preserving ligand-binding pocket geometry. We have designed and experimentally tested binding sites for the small molecule drug Naproxen. **MD simulations** revealed that the designed homodimer interface rigidifies the pocket, compared to the monomers; increasing binding site preorganization and binding affinities (at least 20-fold, in the low μM range) [3].

These results highlight the potential of computational workflows to accelerate the design of both stable and functional ligand-binding proteins with folds featuring well-defined cavities.

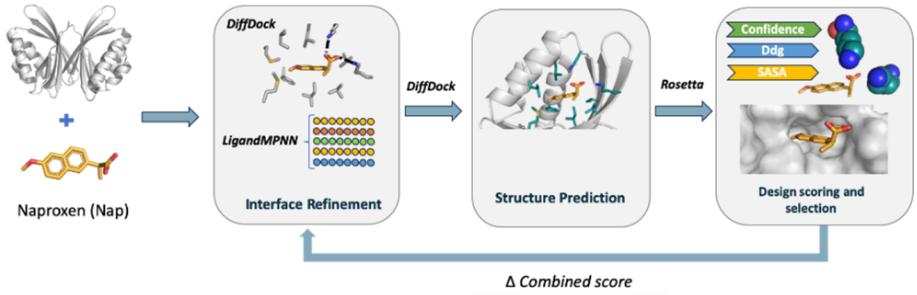


Figure 1: Computational pipeline for optimizing binding pockets. *Schematic overview of the iterative pipeline used to optimize a naproxen-binding site on a dimeric NTF2 scaffold.*

References:

[1] Marcos E, et a. Principles for designing proteins with cavities formed by curved β-sheets, Science, 355, 201-206, 2017.
[2] Lee, G. R et al. Small-Molecule Binding and Sensing with a Designed Protein Family. Biorxiv, November 2, 2023. DOI: 10.1101/2023.11.01.565201
[3] M. Nadal, et al. Buttressing ligand-binding pockets in de novo designed NTF2-like domains. Manuscript under revision, 2025.

(29) SKIPping down a DNA street

Carolina Jerez-Longres[1], Birte Höcker[1]

[1] Department of Biochemistry, University of Bayreuth, Bayreuth, Germany

carolina.jerez-longres@uni-bayreuth.de

Protein motors that transform mechanical energy into movement are essential for life and drive a plethora of processes, from cell division to muscle contraction. Building artificial protein motors allows us to design molecular machines with novel functions, and helps us understand how natural motors work. An existing toolkit of characterised molecular "parts", as well as advances in imaging techniques, protein engineering and design tools, enable the experimental realisation and characterisation of such motor concepts. To date, some examples of artificial protein motors have been built, among which are walkers that move along a track [1].

Here we present the first steps towards an experimental implementation of such a walker, the previously conceived **Synthetic Kinesin-Inspired Protein (SKIP)** [2]. SKIP moves along a DNA track, with "feet" made of repressor proteins that bind to their cognate DNA sequences in the presence of a ligand. Thus, SKIP is a clocked walker, whose motion is powered by changes in the ligands' concentrations. Classical engineering along with protein design and prediction tools are deployed to design linkers that connect the feet together to meet the optimal distance and flexibility requirements. SKIP is expected to move directionally along its track, over distances comparable to those of natural protein motors. A successful implementation of SKIP would advance the understanding of design principles for protein motors.

References:

[1] P. F. Nilsson *et al*. Walking by design: how to build artificial molecular motors made of proteins. *Nano Futures*, 8(4):042501, 2024

[2] M. J. Zuckermann *et al*. Motor properties from persistence: a linear molecular walker lacking spatial and temporal asymmetry. *New Journal of Physics*, 17(5):055017, 2015

(30) Improved Evaluation of loop-mediated protein-protein interactions with AlphaFold guided with ProteinMPNN

Carcelén Lourdes[1], Casadesús Alexandre[1], Marcos Enrique[1]

[1] Institute of Molecular Biology of Barcelona, Department of Structural and Molecular Biology, Barcelona, Spain

lcacri@ibmb.csic.es

The ability to predict binding interfaces directly from amino acid sequences is a critical test for evaluating computationally designed protein binders, as it tests both the folding of the protein binder and its docking orientation with the target. Despite recent advances in deep learning-based structure prediction such as AlphaFold2 [1], accurately modeling loop-mediated interfaces—common in antibodies, nanobodies, and other immunoglobulin-like scaffolds such as monobodies—remains a major challenge due to their hypervariable and flexible nature [2].

In this work, we present AF-ProteinMPNN, a novel method that combines synthetic multiple sequence alignments (MSAs), generated through ProteinMPNN-based interface design, with AlphaFold (AF2 or AF3) to guide structure prediction of binder–target complexes. These synthetic MSAs encode critical interface contacts, enabling AF to improve prediction accuracy for protein-protein interfaces of varying complexity. The requirement of a pre-designed model makes this method particularly well-suited for evaluating and refining binder candidates from prior design efforts, enabling effective discrimination between near-native (active) and incorrect (inactive) models at both the sequence and structural levels. AF-ProteinMPNN is computationally efficient and scalable, enabling high-throughput screening of synthetic binders. By facilitating accurate modeling of irregular and loop-driven interfaces, our approach broadens the scope of computational protein design and brings us closer to designing binders for previously intractable targets.

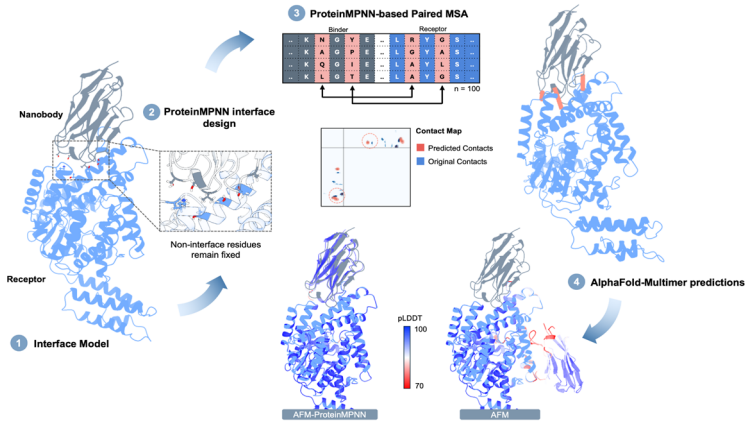


Figure 1: Prediction of loop-mediated protein-protein interactions with AF-ProteinMPNN. *Starting with a protein-protein complex model* ***(1)****, interface residues are mutated with ProteinMPNN* ***(2)*** *and the synthetic paired MSA is built* ***(3)****. Finally, structural predictions with AFM-ProteinMPNN show improved confidence and accuracy* ***(4).***

References:

[1] J. Jumper, et al. Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873):583–589, 2021

[2] J. A. Ruffolo, L.-S. Chu, S. P. Mahajan, and J. J. Gray. Fast, accurate antibody structure prediction from deep learning on massive set of natural antibodies. *Nature Communications*, 14(1):2389, 2023

(31) Target Identification and Design of Cancer Antigen Binding Proteins for CAR T Cell Therapy in Malign Melanoma

Christina K Wuest[1,2], Max A Lingner Chango[1,3], Manfred Kunz[3], Clara T Schoeder[1,2,4]

[1] Leipzig University Medical Faculty, Institute for Drug Discovery, Leipzig, Germany
[2] Fraunhofer Institute for Cell Therapy and Immunology, Leipzig, Germany
[3] Universitätsklinikum Leipzig, Dermatology, Leipzig, Germany
[4] Center for Scalable Data Analytics and Artificial Intelligence, Dresden/Leipzig, Germany

christina.wuest@medizin.uni-leipzig.de

Chimeric antigen receptor (CAR) T cell therapy has transformed cancer treatment by enabling potent tumor cell elimination. However, off-tumor toxicity due to antigen expression in healthy tissues, along with tumor antigen heterogeneity and mutation, continue to limit therapeutic efficacy and drive resistance. To address these challenges, we design *de novo* binders against clinically relevant tumor surface protein derived from cancer patient data.

Targeting overexpressed antigens in cancer with *de novo* proteins has shown promising results [1,2]. Our approach combines state-of-the-art protein design tools: RFdiffusion and hallucination for backbone generation, $MPNN_{prot}$ for sequence optimization, and AlphaFold2 for *in silico* structural validation. Initially generated ErbB3 binders partially show strong and specific binding *in vitro*. This preliminary data will be utilized for pipeline optimization. In our future work we want to establish a rational target identification pipeline that integrates systematic *de novo* binder design.

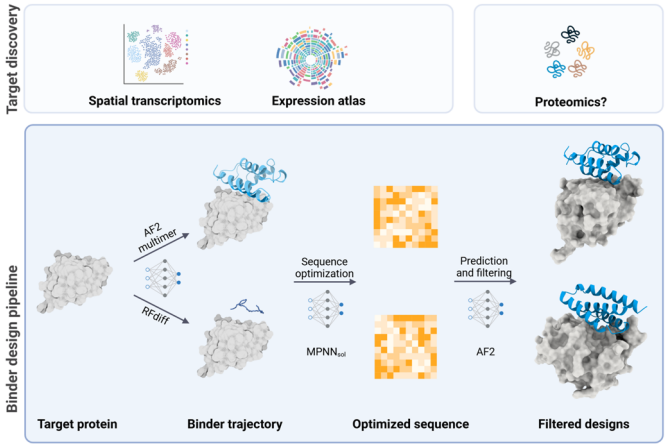


Figure 1: Schematic overview of the de novo protein design pipeline. *The projects aims to streamline target identification from cancer patient-derived data and generate viable antigen-binding domains for next-generation CAR T cell therapies.*

References:

[1] Cao, L., Coventry, B., Goreshnik, I. *et al.* Design of protein-binding proteins from the target structure alone. *Nature* **605**, 551–560 (2022). https://doi.org/10.1038/s41586-022-04654-9.

[2] Xia, Z., Jin, Q., Long, Z. *et al.* Targeting overexpressed antigens in glioblastoma via CAR T cells with computationally designed high-affinity protein binders. *Nat. Biomed. Eng* **8**, 1634–1650 (2024). https://doi.org/10.1038/s41551-024-01258-8.

(32) Application of Deep-Learning-Based Tools for Vector Engineering

Pablo Herrera-Nieto[1] , José Gallardo Hernanz[1], Mark J. van Raaij[1], Carmen San Martín[1]

[1] Department of Macromolecular Structures, Centro Nacional de Biotecnología (CNB-CSIC), 28049 Madrid, Spain

Recombinant human adenovirus vectors have been extensively studied for gene therapy, oncolytic virotherapy, and vaccine delivery. However, their clinical utility is often limited by strong host immune responses and suboptimal tissue specificity [1]. To overcome these challenges, the adenovirus capsid—composed of hexon, penton base, and fiber proteins—must be re-engineered to maintain structural integrity while reducing off-target interactions and immunogenic epitopes. Recent advances in deep-learning–based protein structure prediction and de novo design, including RFdiffusion [2], AlphaProteo [3], and BindCraft [4], are enabling precise modifications of viral capsid components. While these tools have shown success with single-chain proteins [5] and simpler viral vectors such as adeno-associated viruses [4], adapting them for large, multimeric assemblies like adenovirus remains an ongoing challenge. Adenovirus capsid engineering requires balancing structural stability with functional modifications across multiple interacting components. Here, we present a computational-to-experimental workflow to design, evaluate, and refine novel adenovirus capsid variants—termed “intelligent adenovirus” (iAds). Our strategy focuses on fiber knob engineering using BindCraft to optimize receptor interactions and modify tropism while preserving capsid stability (Figure 1). By integrating in silico design with iterative screening, we aim to generate vectors with enhanced specificity and reduced immunogenicity for transgene delivery. Ongoing work involves validating these designs through biochemical and functional assays, focusing on receptor binding, immune evasion, and stability to assess their potential for therapeutic applications.

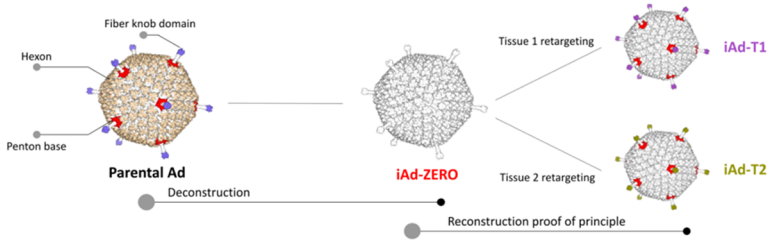


Figure 1: *Development pipeline for new adenoviral vectors.*

References

[1] Yamamoto, Y., Nagasato, M., Yoshida, T., & Aoki, K. (2017). Recent advances in genetic modification of
adenovirus vectors for cancer treatment. Cancer science, 108(5), 831-837.
[2] Watson, J. L., Juergens, D., Bennett, N. R., Trippe, B. L., Yim, J., Eisenach, H. E., ... & Baker, D. (2023).
De novo design of protein structure and function with RFdiffusion. Nature, 620(7976), 1089-1100.
[3] Zambaldi, V., La, D., Chu, A. E., Patani, H., Danson, A. E., Kwan, T. O., ... & Wang, J. (2024). De novo
design of high-affinity protein binders with AlphaProteo. arXiv preprint arXiv:2409.08022.
[4] Pacesa, M., Nickel, L., Schellhaas, C., Schmidt, J., Pyatova, E., Kissling, L., ... & Correia, B. E. (2024).
BindCraft: one-shot design of functional protein binders. bioRxiv, 2024-09.
[5] Vázquez Torres, S., Benard Valle, M., Mackessy, S. P., Menzies, S. K., Casewell, N. R., Ahmadi, S., ... & Baker, D. (2025). De novo designed proteins neutralize lethal snake venom toxins. Nature, 1-7.

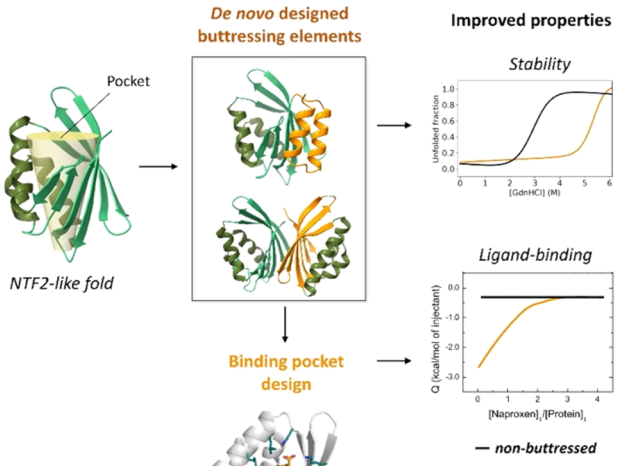
(33) Buttressing *de novo* designed NTF2-like domains for accommodating ligand-binding pockets

Marta Nadal[1], Joana Albi-Puig[1], Albert Castellví[2], Paula M. Garcia-Franco[4], Sonia Vega[4], Adrián Velázquez-Campoy[4], Enrique Marcos[1]

[1]Protein Design and Modeling Lab, Molecular Biology Institute of Barcelona (IBMB), CSIC, Baldiri Reixac 10, 08028 Barcelona, Spain
[2]Crystallographic Methods, Molecular Biology Institute of Barcelona (IBMB), CSIC, Baldiri Reixac 15, 08028 Barcelona, Spain
[4]Institute for Biocomputation and Physics of Complex Systems (BIFI), University of Zaragoza, Zaragoza, Spain

mnrcri@ibmb.csic.es

NTF2-like proteins, characterized by their compact α+β fold and distinctive cone-shaped architecture with an internal pocket, hold significant promise as scaffolds for designing small-molecule binders. A major hurdle in de novo protein design, however, is that engineering ligand-binding pockets often reduces folding stability. To address this, we developed strategies to stabilize NTF2-like domains while retaining their pocket geometry and accessibility. Our approach involved expanding the hydrophobic core through computationally designed α-helical subdomains or homodimer interfaces that support the β-sheet's convex face. Both approaches enhance structural stability without blocking pocket access on the concave side. Biochemical, biophysical, and crystallographic analyses confirm that our designs not only maintain the intended fold but also form well-defined hydrophobic ligand-binding pockets. Crucially, these engineered constructs demonstrated effective small-molecule binding, a capability absent in the original designs.



References:

[1] Nadal M, et al. "Buttressing ligand-binding pockets in de novo designed NTF2-like domains", (under review)
[2] Marcos E, et al. "Principles for designing proteins with cavities formed by curved β-sheets", *Science*, 355, 201-206 (2017)

(34) Design of *de novo* binder targeting paramyxovirus attachment proteins

Dominic Rieger[1], Anna Nobis[1], Lena Kiesewetter[1], Hannes Junker[1], Max Beining[1,3], Phillip Schlegel[1], Sascha Wagner[1], Jens Meiler[1,2,3,4], Clara T. Schoeder[1,2]

[1] Institute for Drug Discovery, Faculty of Medicine, Leipzig University, Leipzig, Germany
[2] Center for Scalable Data Analytics and Artificial Intelligence ScaDS.AI, Dresden/Leipzig, Germany
[3] School of Embedded Composite Artificial Intelligence (SECAI), Cooperation of University Leipzig and TU Dresden, Germany
[4] Department of Chemistry and Center of Structural Biology, Vanderbilt University, Nashville, TN, USA

dominic.rieger@medizin.uni-leipzig.de

Emerging viruses often present diagnostic and therapeutic challenges due to delays in developing effective antibodies or other protein-binding molecules. Traditional approaches depend on isolating antibodies from patients or animals or performing time-consuming high-throughput screenings. Here, we employ deep learning algorithms, including RFdiffusion [1], ProteinMPNN and AlphaFold [2], and BindCraft [3], together with biophysical calculations using Rosetta, to design binders targeting paramyxovirus attachment proteins—specifically those of Nipah, Langya, and Measles viruses. While some members of this family are well characterized, others lack known antibodies, making them ideal candidates for computational binder design.

The Nipah virus attachment protein served as the initial test case, with reference antibodies guiding assay development and validating the *de novo* binder design strategy. We then extended this approach to the Langya and Measles virus attachment proteins. Experimental screening identified positive binders among 10–16 candidates, demonstrating the potential of this "*in silico* first" strategy while highlighting opportunities to improve computational scoring.

Furthermore, we performed *in vitro* functional characterization and structural analysis of the positive protein binder, followed by affinity maturation in a low-throughput format to minimize extensive experimental screening. The resulting binders exhibited high specificity for viral antigens and favorable thermostability, supporting their potential as diagnostics and, with further refinement, as biotherapeutics.

References:

[1] Watson, J.L., Juergens, D., Bennett, N.R. et al. De novo design of protein structure and function with RFdiffusion. *Nature* 620, 1089–1100, 2023
[2] Bennett, N.R., Coventry, B., Goreshnik, I. et al. Improving de novo protein binder design with deep learning. *Nat Commun* 14, 2625, 2023
[3] Pacesa, M., Nickel, L., Schellhaas, C. et al. BindCraft: one-shot design of functional protein binders. bioRxiv 2024.09.30.615802, 2025

(35) Design of a FMN-free NADPH-dependent ene-reductase

Horst Lechner[1], Sebastian Greiter[1], Anna Reisenbichler[2], Tea Pavlow-Keller[2], Karl Gruber[2], Bernd Nidetzky[3], Johann H. Sattler[4], Gustav Oberdorfer[1]

[1] Graz University of Technology, Institute of Biochemistry, Graz, Austria

[2] University of Graz, Institute of Molecular Biosciences, Graz, Austria

[3] Graz University of Technology, Institute of Biotechnology and Biochemical Engineering, Graz, Austria

[4] University of Graz, Institute of Chemistry, Graz, Austria

horst.lechner@tugraz.at

Ene-reductases (ER) are well known to catalyze the asymmetric reduction of activated alkenes. Most ene-reductases are found in the so-called "old-yellow-enzyme" family, with the yellow flavin mononucleotide (FMN)-cofactor as name-giving feature. The enzymes' FMN-cofactor is reduced by a NAD(P)H cofactor. In the oxidative half reaction of FMNH2 a substrate, activated by suited amino acids in the binding pocket is reduced and thus, depending on substituents, creating up to two stereocenters. The enzymes are structurally and mechanistically very well characterized and the reaction is of high value in biocatalysis.

There are only very few examples of ene-reductases known not using FMN but employing directly NAD(P)H as reducing agent [1].

To gain new ene-reductases with possible different selectivities we considered to computationally design one. We redesign a member of aldo-keto reductase superfamily, the NAPH-dependent xylose reductase [2], to catalyse the reduction of an activated alkene. This enzyme was selected since it had some activating amino acids necessary for the ene-reduction following the mechanism of an OYE already at place.

We succeeded in gaining active designs possessing activity for the reduction of cinnamaldehyde to 3-phenylpropionaldehyde, solved the crystal structure of one of them and created several variants to alter selectivity as well as activity.

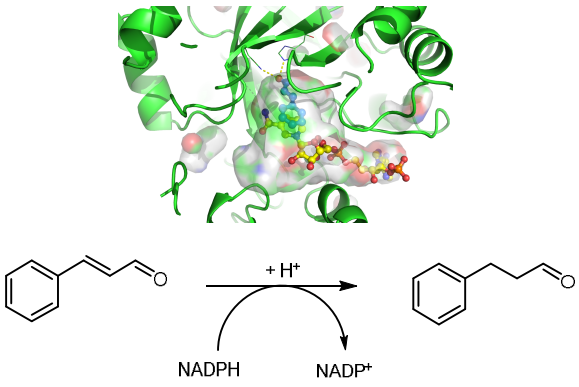


Figure 1. *Reduction of cinnamaldehyde using a designed NADPH dependent redesigned xylose reductase.*

References:

[1] A. Lygidakis, V. Karuppiah, R. Hoeven, A. NíCheallaigh, D. Leys, J. M. Gardiner, H. S. Toogood, N. S. Scrutton, *Angew.Chem.* **2016**, *128*, 9748 –9752; D. J. Mansell, H. S. Toogood, J. Waller, J. M. X. Hughes, C. W. Levy, J. M. Gardiner, N. S. Scrutton, *ACS Catal.* **2013**, *3*, 370−379

[2] R. Kratzer, S. Leitgeb, D. K Wilson, B. Nidetzky, *Biochem J*, **2006**, *393*, 51–58.

(36) In silico pipeline to inform ligandability and molecular format prioritization for emerging target ideas

Che Yang[1]*, David Rodríguez,[1]* Christos Nicolaou,[1] Rodrigo Alonso Ochoa,[1] Jovan Damjanovic,[1]Gerd Schluckebier,[1,2] Yulian Gavrilov,[3] Simone Fulle,[1] and Peter Clark[1]

[1] Digital Chemistry and Design, Novo Nordisk
[2] Structure and Interactions 2, Novo Nordisk
[3] Research Data Stewardship, Novo Nordisk

hchy@novornordisk.com

At Digital Chemistry and Design in Novo Nordisk, we sought to assist the selection of molecular formats for new drug discovery initiatives through effective programmatic searches and structure-based predictions. For that purpose, we developed the Target Tractability Platform, to determine:

- the ligandability of a novel target idea.
- the prioritization of the most viable molecular formats to effectively achieve the desired pharmacological modulation of that target. We have a diverse array of options to choose from, including small molecules, peptides, miniproteins, VHHs, antibodies, and oligonucleotides

In order to achieve this aspiration, we build on existing knowledge and toolkits, both internal and external sources.

- Molecular format precedence, i.e., bioactive compound and clinical data.
- Molecular format predictions, such as druggable sites amenable for specific molecular formats.
- Beyond just ligandability. Assess target selectivity and considerations with regards to cross-species translatability.
- Partner with idea submission teams. Provide unbiased and standardized feedback to an expert group for evaluating molecular format choices.

(37) PANTHER - **P**rotein-**A**ffinity for **N**ucleic **T**argeting, **H**ybridization, and **E**nergy **R**egression

Parisa Aletayeb [1], **Akash Deep Biswas,**[2*] **Stefano Rocca,**[1] **Carmine Talarico,**[2] **Guilio Vistoli,**[1] **and Alessandro Pedretti**[1*]

[1] Dipartimento di Scienze Farmaceutiche, Università degli Studi di Milano, Via Mangiagalli, 25, I-20133 Milano, Italy
[2] EXSCALATE, Dompé Farmaceutici S.p.A., Via Tommaso de Amicis 95, 80131, Napoli, Italy

Email First Author: parisa.aletayeb@unimi.it

Protein-RNA interactions are crucial to many biological functions, but predicting their binding free energies (ΔG) is challenging due to limited data and complex dynamics. To address this, we developed **PANTHER** (Protein Affinity for Nucleic Targeting, Hybridization, and Energy Regression), a machine learning model that estimates energy-based scores for protein-RNA complexes.

PANTHER uses a local-to-global strategy: it derives interaction energies from molecular dynamics simulations, extracts pairwise features, and applies noise-reduction techniques to build a reliable training dataset. The model was trained on 53 curated complexes and tested on an external set of 112. Random Forest Regression achieved the best performance (r = 0.80, MAE = 1.79 kcal/mol) and generalized well to the stress set (r = 0.63). Compared to tools like PredPRBA and PRA-Pred, PANTHER showed higher accuracy and robustness. It demonstrates how machine learning can bypass data scarcity and the computational limits of traditional quantum methods, making it a scalable tool for biomolecular research.

To enhance accessibility, **PANTHER is freely available as a webserver** at: **https://nova.disfarm.unimi.it/panther/**, where users can input PDB IDs or upload custom files to compute binding affinities.

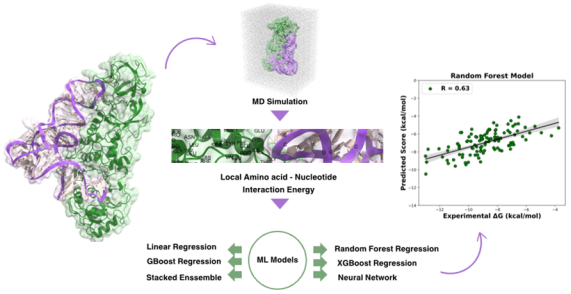


Figure 1. ***PANTHER workflow for predicting protein-RNA binding affinities score****. The pipeline includes MD simulations to extract local interaction energies, which are used to train ML models. The scatter plot shows the correlation (R = 0.63) between experimental ΔG and predicted scores using Random Forest model.*

References:

[1] K. Harini, M. Sekijima, M.M. Gromiha, PRA-Pred: Structure-based prediction of protein-RNA binding affinity, *Int J Biol Macromol* 259, 2024
[2] L. Deng, W. Yang, H. Liu, PredPRBA: Prediction of Protein-RNA Binding Affinity Using Gradient Boosted Regression Trees, *Front Genet* 10, 2019.

(38) Context-aware geometric deep learning for RNA sequence design

Parth Bibekar[1], Lucien F. Krapp[1], Matteo Dal Peraro[1]

[1] Institute of Bioengineering, School of Life Sciences, École Polytechnique Fédérale de Lausanne (EPFL) and Swiss Institute of Bioinformatics (SIB), Lausanne 1015, Switzerland

matteo.dalperaro@epfl.ch

RNA design has emerged to play a crucial role in synthetic biology and therapeutics. Although tertiary structure-based RNA design methods have been developed recently, they still overlook the broader molecular context, such as interactions with proteins, ligands,DNA, or ions, limiting the accuracy and functionality of designed sequences. To address this challenge, we present RISoTTo (RIbonucleic acid Sequence design from TerTiary structure), a parameter-free geometric deep learning approach that generates RNA sequences conditioned on both their backbone scaffolds and the surrounding molecular context. We evaluate the designed sequences based on their native sequence recovery rate and further validate them by predicting their secondary structures *in silico* and comparing them to the corresponding native structures. RISoTTo performs well on both metrics, demonstrating its ability to generate accurate and structurally consistent RNA sequences. Additionally, we present an *in silico* design study of domain 1 of the NAD+ riboswitch, where RISoTTo-generated sequences are predicted to exhibit enhanced binding affinity for both the U1A protein and the NAD+ligand.

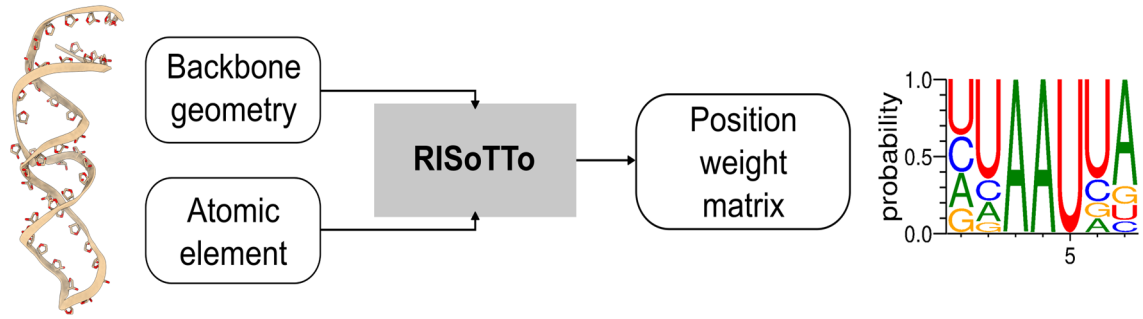


Figure 1: Overview of RISoTTo workflow for context-aware RNA sequence design

References:

[1] Bibekar, P., Krapp, L. F. & Dal Peraro, M. Context-aware geometric deep learning for RNA sequence design. (2025) doi:10.1101/2025.06.21.660801.

(39) Language guided evolution: Designing novel RuBisCO variants with protein language models

Alexander Braun[1], Birte Höcker[1]

[1] University of Bayreuth, Biochemistry III – Protein Design, Bayreuth, Germany

Alexander1.Braun@uni-bayreuth.de

Ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) is the most abundant enzyme on earth and the cornerstone of the Calvin-Benson-Bassham cycle. Despite its central role in sustaining life, RuBisCO is remarkably inefficient compared to other enzymes participating in central carbon metabolic pathways [1]. Its slow catalytic rate and propensity for oxygenation led to reduced photosynthetic efficiency. It is suggested the inefficiency of RuBisCO is based on its ancient origin in a high-CO2, low-O2 atmosphere, where its dual carboxylase/oxygenase activity was not detrimental. As oxygen levels rose, the enzymes inability to efficiently discriminate between the substrates led to the costly photorespiration [1,2]. Further RuBisCO remained remarkably conserved, evolving at one of the slowest rates among known genes, constrained more by phylogenetic inertia than catalytic trade-offs [2,3]. This inefficiency has motivated efforts to engineer RuBisCO variants with improved activity, yet without major successes [1].

We present a novel strategy for RuBisCO engineering that leverages the generative power of protein language models (pLMs) to explore sequence space far beyond natural diversity. By fine-tuning various pLMs with curated datasets, including form II RuBisCO (bacterial), ancestral, and synthetic sequences, we aim to create models capable of generating highly diverse, low identity RuBisCO variants, that are readily expressible in *E. coli*. These sequences could serve as starting points for laboratory evolution, putatively enabling the discovery of functional enzymes unconstrained by natural linage.

In first training attempts, we increased the fraction of generated sequences passing computational success metrics based on their natural counterparts (based on structure prediction scores for monomers and multimers, physics-based aggregation scores, and ProteinMPNN scoring) by four orders of magnitude compared to conditional generation using the base model. Our results highlight the potential of combining evolutionary insights with synthetic biology and machine learning to traverse the protein fitness landscape and unlock now possibilities for enzyme engineering.

References:

[1] N. Prywes, N.R. Phillips, O.T. Tuck, L.E. Valentin-Alvarado, D.F. Savage, Rubisco Function, Evolution, and Engineering, Annual review of biochemistry 92 (2023) 385–410.

[2] J.W. Bouvier, D.M. Emms, T. Rhodes, J.S. Bolton, A. Brasnett, A. Eddershaw, J.R. Nielsen, A. Unitt, S.M. Whitney, S. Kelly, Rubisco Adaptation Is More Limited by Phylogenetic Constraint Than by Catalytic Trade-off, Molecular biology and evolution 38 (2021) 2880–2896.

[3] J.W. Bouvier, D.M. Emms, S. Kelly, Rubisco is evolving for improved catalytic efficiency and CO2 assimilation in plants, Proceedings of the National Academy of Sciences of the United States of America 121 (2024) e2321050121.

(40) Rosetta-Based Simulation of Methylglyoxal Modifications Predicts Glycation-Driven Destabilization of Protein–RNA Interfaces

Abi Dozic[1], Xiao Yang[2], Anna Knoerlin[2], Yael David[2], Caleb Lareau[1]

1 Physiology, Biophysics, and Systems Biology Program, Weill Cornell Graduate School of Medical Sciences, New York, NY, USA

2 Chemical Biology Program, Sloan Kettering Institute, New York, NY , USA

abd4006@med.cornell.edu

Serine/arginine-rich splicing factor 2 (SRSF2) is a central RNA-recognition protein whose altered binding preferences drive widespread splicing defects in myelodysplastic syndromes and related leukemia. While mutations that weaken SRSF2-RNA affinity are well documented, the impact of non-enzymatic post-translational modifications remain unknown. Methylglyoxal (MGO), a reactive glycolytic by-product, readily modifies arginine, forming methylglyoxal-derived hydroimidazolone 1 (MGH-1). Here we combine structure-guided modeling and biophysical assays to explore how arginine glycation modulates SRSF2–RNA interactions.

Inspection of the SRSF2-RNA (PDB: 2LEC) NMR structure identified four arginines — R5, R61, R91, and R94 —positioned for direct contacts with RNA. To test the effect of glycation on SRSF2's RNA binding capability in silico, we parameterized MGH-1 as a non-canonical residue in Rosetta and applied previously developed Rosetta-Vienna RNP-ddG. Glycation of R61, R91, or R94 was predicted to destabilize binding, whereas modification of R5 was predicted to improve the binding to RNA. To validate these predictions, we purified recombinant SRSF2, treated it with MGO in vitro, and measured its RNA binding affinity by biolayer interferometry (BLI). Glycated SRSF2 showed a >20x fold decrease in binding affinity. In parallel, full-length SRSF2 arginine-to-tryptophan mutants (R→W mimicking steric bulk of MGH-1) were overexpressed in HEK293T cells for RNA immunoprecipitation with a Cy5-labelled SRSF2-binding RNA oligo. The R61W and R94W variants exhibited > 80 % and 60 % reductions in recovery, respectively, while R91W displayed a moderate 20 % decrease; R5W on the other hand had a 20% increase in recovery comparing to wild-type. Close examination of most energetically minimized SRSF2-RNA structures for R61MGH-1 and R94MGH-1 revealed that the loss of hydrogen bonds between a base and arginine upon glycation could be the main cause of observed binding loss. To our knowledge this is the first report of modeling the impact of protein glycation on RNA binding. These findings suggest glycation as a potential contributor to the splicing alterations observed in hyper-glycolytic malignancies and motivate further exploration of metabolite-driven epitranscriptomic control.

(41) FrustraMPNN: An ultra-fast deep learning tool for proteome-scale analysis of protein single residue frustration

Max Beining[1,3], Felipe Engelberger[1], Jens Meiler[1,2,3,4,5], Clara T. Schoeder[1,2]

[1] Institute for Drug Discovery, Leipzig University, Faculty of Medicine, Leipzig, Germany
[2] Center for Scalable Data Analytics and Artificial Intelligence ScaDS.AI, Dresden/Leipzig, Germany
[3] School of Embedded Composite Artificial Intelligence (SECAI), Cooperation of University Leipzig and TU 9 Dresden
[4] Department of Chemistry, Vanderbilt University, Nashville, Tennessee, United States of America
[5] Center for Structural Biology, Vanderbilt University, Nashville, Tennessee, United States of America

beining@izbi.uni-leipzig.de

Energetic frustration, where conflicting local interactions exist within a protein, is a key determinant of protein dynamics, allostery, and function. However, the immense computational cost of traditional single-residue frustration analysis has prevented its application at the proteome scale, creating a major bottleneck in structural biology. Here, we introduce FrustraMPNN, a message-passing neural network that predicts per-residue frustration indices orders of magnitude faster than existing methods while maintaining high accuracy. We validate FrustraMPNN on an E. coli proteome scale, a diverse external set of over 3,000 human protein structures, achieving 0.81 Spearman correlation over all frustration categories, demonstrating its robust generalizability. By conducting a thorough dataset ablation study, our results indicate that the model's performance improves from training on a wider range of protein sizes, highlighting an inherent limitation of datasets like Megascale. We provide FrustraMPNN as an open-source tool with an accessible Docker and Google Colab interface. We expect this model will enable exploring new hypotheses in fields like personalized structural biology, human and plant proteomics at unprecedented scale.

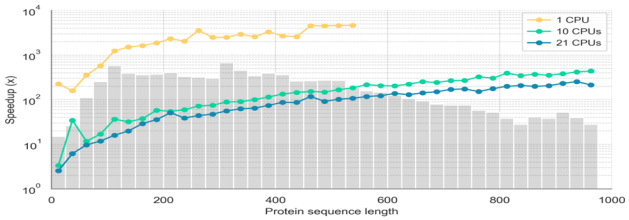


Figure 1: Computational acceleration achieved by FrustraMPNN for single-residue frustration analysis. *Mean speedup factor compared to frustrapy's single-residue mode plotted against protein sequence length in 25-residue bins. Single-residue analysis requires 20 frustration calculations per position (one for each amino acid), making it ~20× more computationally expensive than configurational or single-mutation analysis. Separate curves show frustrapy performance with 1 (blue), 10 (orange), and 21 (green) CPU cores (log scale speedup). Gray bars indicate the number of proteins in each size bin from the validation set. FrustraMPNN achieves >100× speedup for proteins of ~500 residues (10,000 individual calculations), with acceleration increasing super-linearly with protein size.*

(42) De Novo Design of Chromatin Binders

B. de Potter[1], L. Baan[1], A. Veltman[1], N. Rinzema[1], D. D. Sahtoe[1]

[1] Hubrecht Institute-KNAW, Oncode Institute, Uppsalalaan 8, Utrecht 3584 CT, Netherlands

d.sahtoe@hubrecht.eu

Chromatin folding and chemical modifications play central roles in regulating gene expression by influencing chromatin structure and accessibility [1-3]. Together, these mechanisms are essential for maintaining normal cellular function. When disrupted, they contribute to the development of diseases such as cancer [4].

Studying chromatin dynamics and gene expression in both healthy and diseased states requires tools that can bind chromatin and its chemical modifications with high specificity. Many of these modifications occur on histone tails, which are the intrinsically disordered regions of histone proteins. However, conventional protein-based tools struggle to target these disordered regions effectively [5].

To address this challenge, we present a design approach aimed at overcoming these limitations. Our method involves the design of scaffolds with the ability to effectively bind to histone tails by forcing them in β-strand conformations. Initial results indicate successful binding to the H4 histone tail, demonstrating that the approach works, and we are currently expanding our approach to also bind histone modifications. These binders will shed new light on the molecular basis of epigenetic gene regulation and may ultimately enable us to target and/or modify undesired chromatin states through epigenome editing approaches.

References:

[1] Luger K, Dechassa ML, Tremethick DJ. New insights into nucleosome and chromatin structure: an ordered state or a disordered affair? Nat Rev Mol Cell Biol. 2012 Jul;13(7):436–47.
[2] Arya G, Schlick T. Role of histone tails in chromatin folding revealed by a mesoscopic oligonucleosome model. Proc Natl Acad Sci. 2006 Oct 31;103(44):16236–41.
[3] Greenberg MVC, Bourc'his D. The diverse roles of DNA methylation in mammalian development and disease. Nat Rev Mol Cell Biol. 2019 Oct;20(10):590–607.
[4] Mitchener MM, Muir TW. Oncohistones: Exposing the nuances and vulnerabilities of epigenetic regulation. Mol Cell. 2022 Aug 18;82(16):2925–38.
[5] Veggiani, G., Villaseñor, R., Martyn, G.D. et al. High-affinity chromodomains engineered for improved detection of histone methylation and enhanced CRISPR-based gene repression. Nat Commun 13, 6975 (2022).

(43) Computationally Engineered Ligands for High Affinity Chromatography Resins

Victoria Most[1,2], Jon Kapla[1], Åsa Hagner McWirther[1], Per Liden[1], Ismet Dorange[3]

[1] Cytiva, Björkgatan 30, 753 23 Uppsala, Sweden
[2] AFRY, Frösundaleden 2A, 169 75 Solna, Sweden
[3] Danaher, 2200 Pennsylvania Avenue, NW, Washington, DC, United States

victoria.most@cytiva.com

Cytiva chromatography resins enable scalable protein purification, supporting workflows from early-stage research to full-scale manufacturing. Affinity Ligand Chromatography is a powerful and widely adopted technique that purifies target proteins through reversible, selective interactions with ligands immobilized on a resin matrix. Cytiva's HiTrap™ Protein A columns exemplify this approach, delivering high-performance immunoaffinity purification of most immunoglobulin Gs via selective binding to the Fc region [1]. While traditional methods have achieved significant success, recent advances in structural modeling and protein engineering have unlocked new possibilities for designing highly specific affinity ligands. Cytiva aims to harness these innovations to develop and offer tailored chromatography resins optimized for distinct and demanding applications.

References:

[1] https://www.cytivalifesciences.com/en/us/solutions/bioprocessing/products-and-solutions/downstream-bioprocessing/protein-a-chromatography-solutions, 2025

(44) Expanding the *Arabidopsis thaliana* interactome using coevolutionary signals and deep learning

Laura Pilgram, Georg Künze

Leipzig University, Institute for Drug Discovery, Leipzig, Germany

laura-marie.pilgram@medizin.uni-leipzig.de

Protein-protein interactions (PPIs) are central to most cellular processes. While experimental approaches such as yeast two-hybrid (Y2H), affinity purification mass spectrometry (AP-MS), and cross-linking mass spectrometry (XL-MS) have been widely used to detect PPIs, they often suffer from limited reproducibility and high rates of both false positives and false negatives. Computational methods, particularly coevolution-based approaches as used in recent large-scale interactome studies [1–3], can complement experimental techniques by predicting PPIs directly from sequence data using statistical models and machine learning.

In this study, we aim to construct and structurally characterize a more comprehensive interactome of *Arabidopsis thaliana* by integrating experimental data from resources such as STRING and the Protein Data Bank (PDB) with computational predictions. Our pipeline, based on the approach of Humphreys et al. [1–2], generates paired multiple sequence alignments (pMSAs) of orthologous protein pairs to detect coevolutionary signals using Direct Coupling Analysis (DCA) [4] and deep learning models such as RoseTTAFold2-Lite [1] and AlphaFold2 [5]. Our proteome-wide screening identifies both novel and known PPIs, and the predicted complex structures provide insights into interaction mechanisms and their potential modulation by mutations.

References:

[1] I. R. Humphreys, J. Zhang, M. Baek et al. Protein interactions in human pathogens revealed through deep learning. Nat Microbiol 9, 2642–2652, 2024

[2] I. R. Humphreys, J. Pei, M. Baek et al. Computed structures of core eukaryotic protein complexes. Science 374, eabm4805, 2021

[3] J. Zhang, J. Pei, J. Durham, T. Bos, Q. Cong. Computed cancer interactome explains the effects of somatic mutations in cancers. Protein Science, 31(12):e4479, 2022

[4] M. Weigt, R. A. White, H. Szurmant, J. A. Hoch, T. Hwa. Identification of direct residue contacts in protein–protein interaction by message passing. Proc. Natl. Acad. Sci. U.S.A. 106 (1) 67-72, 2009

[5] J. Jumper, R. Evans, A. Pritzel et al. Highly accurate protein structure prediction with AlphaFold. Nature 596, 583–589, 2021

(45) Lipid A modifications in *Klebsiella pneumoniae*: impairing polymyxin penetration via distinct mechanisms

Wendong Ma[1], Sue Nang[1], Kaijie Mu[1], Xukai Jiang[2], Meiling Han[1], Jian Li[1]

[1] Biomedicine Discovery Institute, Monash University, Australia
[2] National Glycoengineering Research Center, Shandong University, China

wendong.ma@monash.edu

Polymyxin-resistant Gram-negative bacteria, including *Klebsiella pneumoniae*, represent a significant public health threat globally. Lipid A modification is the major mechanism of polymyxin resistance. In *K. pneumoniae*, the simultaneous presence of multiple lipid A modifications, such as phosphoethanolamine (pEtN) and 4-amino-4-deoxy-L-arabinose (Ara4N), induces complex outer membrane (OM) remodeling. How OM remodeling associated with polymyxin resistance impacts polymyxin activity remains unclear, hindering the discovery of more efficacious lipopeptide antibiotics. To address this, we used all-atom molecular dynamics (MD) simulations and OM models based on quantitative membrane lipidomics, we discovered that extensive lipid A modifications increased OM heterogeneity and decreased stability. A higher proportion of modified lipid A impaired polymyxin B (PMB) binding and hindered its penetration by preventing the formation of folded conformations, thereby reducing activity. These findings were supported by minimum inhibitory concentration (MIC) and time-kill assays. Furthermore, we identified distinct interaction modes between PMB and two key lipid A modifications. The flexible linear structure of pEtN impeded polymyxin penetration by anchoring the molecule at the OM surface through electrostatic attraction. In contrast, Ara4N acted as an electrostatic barrier, strongly hindering the initial insertion of polymyxins into the glycerol region of the OM. These findings provide atomic-level insights into how the extent of OM remodeling governs polymyxin–OM interactions, and how different lipid A modifications hinder polymyxin penetration through distinct mechanisms. Our results may inform the discovery of novel lipopeptide antibiotics to combat life-threatening Gram-negative pathogens.

(46) The Docking Problem: Revisiting a Challenging Case of Local Protein Flexibility

Rebecca Würz[1], Christoph Sotriffer[1]

[1]University of Würzburg, Institute of Pharmacy and Food Chemistry, Germany

rebecca.wuerz@uni-wuerzburg.de

Docking-based virtual screening of small molecules plays a key role in structure-based drug design and drug discovery pipelines. The success critically depends on the availability of a target structure and its quality. Even an excellently resolved structure, however, is of little value if it does not represent the binding conformation for a particular ligand. This is due to a major limitation of many docking programs applied in virtual screening: the rigid treatment of the protein.[1] Ideally, side chain and backbone flexibility should be considered, but this increases the search space and complicates the scoring of the generated docking poses.[2] Nevertheless, many developers within and outside the Rosetta Commons community have made efforts in recent years to include flexibility into small-molecule docking.[1-3]

Here, current docking programs and workflows considering full protein flexibility are analysed with respect to aldose reductase as a model system known for its local but challenging conformational changes in the backbone and side chains upon ligand binding.[4] Among the more than 80 different complexes of aldose reductase in the PDB, two binding-pocket conformations are dominating: one is considered as the closed state, while the other shows an open transient pocket. Besides that, however, two additional conformations were found with two particular ligands that are closely related analogues. The question addressed in this work is whether and how a docking workflow can successfully reproduce their binding modes if only the two dominating protein conformations are available. In addition, it has to be elucidated which further conformations may potentially be accessible to new inhibitors.

References:

[1] K. Wu et al. *Advancements in small molecule drug design: A structural perspective*. Drug Discov. Today. **2023**, 28, 103730

[2] J. Paggi et al. *The Art and Science of Molecular Docking.* Annu. Rev. Biochem. **2024**, 93, 389-410

[3] N. Pagadala et al. *Software for Molecular Docking: A Review.* Biophys. Rev. **2017**, 9, 91-102

[4] M. Zentgraf et al. *How Reliable Are Current Docking Approaches for Structure-Based Drug Design? Lessons from Aldose Reductase.* Angew. Chem. Int. Ed. **2007**, 46, 3575-3578

(47) Polar Interactions Design

Serj Koshian[1], Danny Sahtoe[1]

[1] Hubrecht Institute, Utrecht, The Netherlands

s.koshian@hubrecht.eu

Substantial advances have occurred in recent years in our ability to design proteins de novo driven by the use of deep learning techniques. This has enabled us to design proteins faster and with higher success rates than previously possible. However, despite these advances, our ability to design predominantly polar interfaces remains limited. Polar interactions, in order to form, require amino acids to be positioned at precise geometries to form a network of interactions, which typically involves water molecules as well – features that we are currently unable to design for well with our existing tools. Thus, to address this problem, we are setting up a high-throughput assay to investigate polar protein-protein interfaces by screening thousands of binding pairs with mostly polar interactions at their interfaces. Analyses of the top hits could help us better understand polar interfaces and aid in polar interface design.

(48) PolyCraft: A specificity-driven protein binder design

Valentas Brasas[1,2], Beatrice Scapolo[1,3], Charlotte R. Christensen[1,3], Kristoffer H. Johansen[3], Timothy P. Jenkins[1]

[1]Technical University of Denmark, Department of Biotechnology and Biomedicine, Lyngby, Denmark
[2]Technical University of Denmark, Novo Nordisk Foundation Center for Biosustainability, Lyngby, Denmark
[3]Department of Health Technology, Technical University of Denmark, Kongens Lyngby, Denmark

valbra@dtu.dk

Designing proteins with precisely controlled binding properties remains one of the central ambitions in computational biology. Current approaches are typically optimized for a single target[2] and become less effective when extended to situations where multiple molecular interactions must be simultaneously considered. In this work, we introduce a framework called Polycraft that adapts optimization-guided generative strategies for multi-target aware protein design, enabling the coordinated tuning of binding preferences across several targets.

Our approach integrates information from minibinder-target complexes predicted by AlphaFold2[1] across multiple targets, balancing binding towards one or more desired targets while actively reducing unwanted interactions with other targets. This makes it particularly relevant for applications such as therapeutic minibinders, where specificity or cross-reactivity needs to be jointly optimized[3].

In silico results suggest that the framework can explore sequence–structure landscapes efficiently. Polycraft generates minibinders with desired specificity in some of biology's most challenging settings, i.e., point-mutant specificity against pMHCs. More broadly, this work highlights how protein structure prediction can be integrated with generative optimization to address multi-objective design problems, providing a path toward frameworks that better capture the trade-offs and complexities of realistic binding scenarios.

References:

[1] J. Jumper et al. Highly accurate protein structure prediction with AlphaFold. *Nature*, 596:583–589, 2021.
[2] M. Pacesa, L. Nickel, C. Schellhaas, J. Schmidt, E. Pyatova, *et al.* BindCraft: one-shot design of functional protein binders. *bioRxiv*, 2024. doi:10.1101/2024.09.30.615802.
[3] K. H. Johansen, D. S. Wolff, B. Scapolo, M. L. Fernández-Quintero, C. R. Christensen, *et al.* De novo-designed pMHC binders facilitate T cell–mediated cytotoxicity toward cancer cells. *Science*, 389(6758):380–385, 2025. doi:10.1126/science.adv0422.

(49) *De novo* design of a polycarbonate hydrolase

Freja Rønne[1,2], Laura H. Holst[2], Niklas G Madsen[2], Freja T Toftgård[2], Ioana-Malina Moise[2], Evamaria I Petersen[2], Peter Fojan[2]

[1] Novonesis, Copenhagen, Denmark*
[2] Aalborg University, Department of Materials and Production, Aalborg, Denmark

frq@novonesis.com / filbr@mp.aau.dk

Esterases, lipases, and PETases have been used to engineer enzymes capable of degrading synthetic plastics, however, the family of serine hydrolases occupies a narrow sequence space. In our approach, with the help of Rosetta, an active site catalysing the hydrolysis of polycarbonate was introduced into a set of thermostable protein scaffolds to generate *de novo* enzymes, which overcome the limits of sequence space by which natural proteins are bound. Through computational evaluation, a potential PCase was selected and produced recombinantly in *Escherichia coli*. Thermal analysis suggests that the designed PCase has a melting temperature of > 95°C. The designed PCase shows activity towards a model substrate, and activity toward polycarbonate was confirmed using atomic force spectroscopy (AFM), proving the successful design of a PCase.

*Not affiliated with Novonesis at the time of data generation and publication [1].

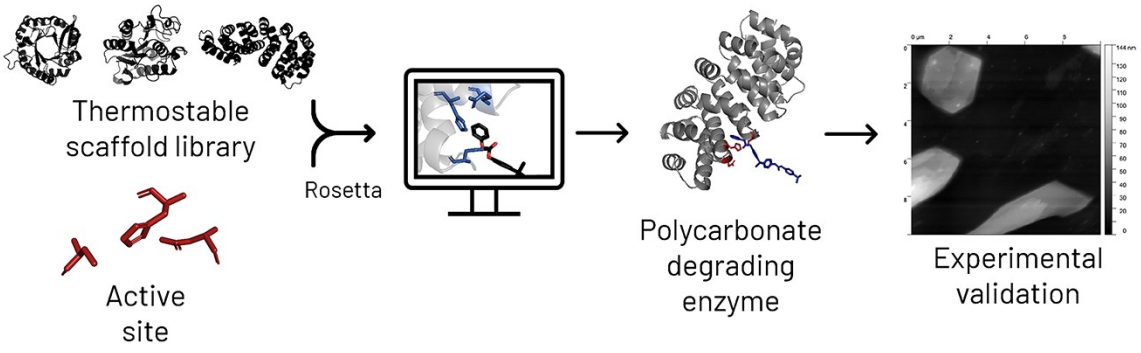


Figure 14: de novo design and validation of a PCase. *Active site matched to scaffolds and optimised using Rosetta, computationally evaluated, and experimentally validated.*

References:

[1] Laura H Holst, Niklas G Madsen, Freja T Toftgård, Freja Rønne, Ioana-Malina Moise, Evamaria I Petersen, Peter Fojan, *De novo* design of a polycarbonate hydrolase, *Protein Engineering, Design and Selection*, Volume 36, 2023, gzad022, https://doi.org/10.1093/protein/gzad022

(50) The design of heterochiral DNA-binding peptides

Tobias Wörtwein[1,2], Kateryna Maksymenko[2], Murray Coles[2], Mohammad El-Gamacy[1,2]
[1] Internal Medicine II, University Hospital Tübingen, Tübingen, Germany
[2] Max Planck Institute for Biology, Department of Protein Evolution, Tübingen, Germany

Tobias.woertwein@tuebingen.mpg.de

DNA binding proteins are relatively large-sized or unstable. Moreover, targeting these proteins to a specific point within a genome is a challenging design problem. To solve this, I am designing peptides that can bind specifically to pre-defined DNA sequence patterns by extending across the major and/or minor groove. These then can be linked to regulatory payloads which to home in on a target sequence. To further stabilize these peptides, I leverage non-canonical amino acids, which could readily achieve protease resistance, and can possess wider structural diversity. To this end, I developed a specific sampling protocol in Damietta capable of sculpting novel peptide backbones and sequences. Preliminary work to validate this design pipeline has yielded sub-micromolar DNA binders with extreme protease resistance.

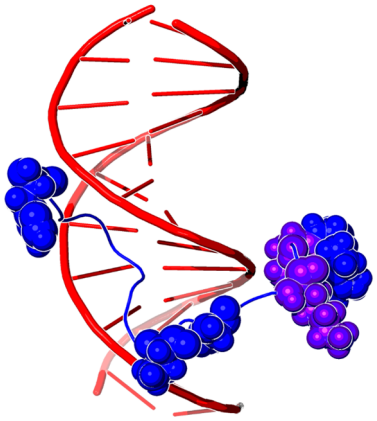


Figure 1: Peptide binding on the minor grove. *This shows the current design to validate the Damietta protein design suite to work with d-amino acids (shown in blue) as well as DNA. This peptide is binding specific on the minor groove.*

(51) FrustraMPNN : Predicting of local frustration in proteins using deep learning

Max Beining[1,3], Felipe Engelberger[1,2], Onur Elaldi[1], Jens Meiler[1,2,3,4,5], and Clara T. Schoeder[1,2]

[1]Institute for Drug Discovery, Leipzig University, Faculty of Medicine, Leipzig, Germany
[2]Center for Scalable Data Analytics and Artificial Intelligence ScaDS.AI, Dresden / Leipzig, Germany
[3]School of Embedded Composite Artificial Intelligence (SECAI), Cooperation of University Leipzig and TU Dresden
[4]Department of Chemistry, Vanderbilt University, Nashville, Tennessee, United States of America
[5]Center for Structural Biology, Vanderbilt University, Nashville, Tennessee, United States of America

elaldi.onur@uni-leipzig.de

Strong energetic conflicts are minimized in folded native states following the principle of minimal frustration, allowing a sequence to spontaneously fold. However, local deviations from this principle create the complex energy landscapes necessary for active biological functions. Notably, catalytic sites themselves are often highly frustrated, regardless of the protein oligomeric state. We propose a neural network for predicting local frustration patterns in proteins, building on ProteinMPNN with ThermoMPNN-inspired architectures. By capturing energetic conflicts, our model aims to reveal functional complexity, particularly in catalytic sites, enabling frustration-guided protein design. Our model demonstrated high predictive accuracy for protein frustration patterns, achieving approximately 0.8 Spearman correlation across the comprehensive Megascale dataset. At the cost of minimal loss in accuracy, FrustraMPNN provides a computationally efficient oracle for integrating frustration analysis into deep learning workflows for protein engineering.

(52) Lipid phase separation within outer membrane vesicles mediates bacterial resistance to antimicrobial peptides

Kaijie Mu[1,2], Wendong Ma[2], Zhenyu Ma[1], Jianzhong Ye[3], Jian Li[2], Xukai Jiang[1]

[1] National Glycoengineering Research Center, Shandong University, Qingdao 266237, China
[2] Biomedicine Discovery Institute, Monash University, Melbourne 3800, Australia
[3] Department of Clinical Laboratory, Key Laboratory of Clinical Laboratory Diagnosis and Translational Research of Zhejiang Province, The First Affiliated Hospital of Wenzhou Medical University, Wenzhou 325000, Zhejiang Province, China

xukai.jiang@sdu.edu.cn

Multidrug-resistant Gram-negative pathogens have represented a major global medical challenge due to lack of new antibiotics. Outer membrane vesicles (OMV) in Gram-negative bacteria significantly contribute to their resistance to antimicrobial peptides (AMPs), particularly lipopeptide polymyxins. However, the lack of knowledge about how OMV interacts with AMPs impedes the understanding of the unique resistance mechanism. Here, we employed large-scale coarse-grained molecular dynamics simulations and enhanced sampling techniques to explore the structural dynamics of OMVs and their interactions with polymyxins and other AMPs. Our results demonstrated that lipid phase separation occurred between the lipopolysaccharide (LPS) and phospholipid (PL) molecules within the outer leaflet of OMVs, forming LPS-rich regions, PL-rich regions, and LPS-PL interfaces. Interestingly, small geometric defects appeared at LPS-PL interfaces due to mismatched orientation between LPS and PL molecules. These defects enhanced polymyxin binding to OMVs with folded conformations, in which their hydrophobic regions inserted into the PL-rich phases while the positively charged residues bound to the exposed phosphate groups of lipid A. Free energy calculations confirmed that polymyxins penetrated OMVs more effectively at LPS-PL interfaces than at LPS-rich regions. Importantly, this biased location at LPS-PL interfaces was found across six other types of AMPs, including Melittin, LL-37, Magainin 2, Tachyplesin 1, Protegrin 1, and Capitellacin. Our findings elucidate that lipid phase separation enables OMVs to act as decoys, absorbing AMPs and thereby reducing their binding to bacterial cells and enhancing bacterial survival. These mechanistic insights will inspire the design of novel AMPs that can evade the protective effect of OMVs.

(53) Bacteriophage Protein Design Incorporating the Rosetta Energy Function and Elitist Evolutionary Algorithms

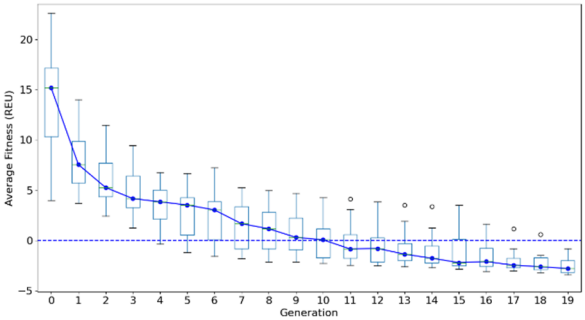
Rolando Armas[1], Cristofer Motoche-Monar[2,3], Francisco Hidrobo[1], Jose A. Castillo[2,3]

[1] School of Mathematical and Computational Sciences, Yachay Tech University, Imbabura, Ecuador
[2] School of Biological Sciences and Engineering, Yachay Tech University, Imbabura, Ecuador
[3] Phagen Tech, Imbabura, Ecuador

cristofer.motoche@yachaytech.edu.ec

This study presents a proof-of-concept for enhancing protein-protein interactions using a single-objective elitist evolutionary algorithm. The research focuses on a *de novo* structural complex formed by the bacteriophage Rs551's Receptor Binding Protein (RBP) [1] and the *R. solanacearum* bacterial pilus [2]. The primary objective was to minimize the interaction energy within the defined 38-position interface zone to design a more robust molecular interaction. To achieve this, a genetic algorithm (GA) was integrated with the Rosetta Macromolecular Modeling suite. The algorithm employs a specialized mutation operator that identifies stabilizing and destabilizing residues and probabilistically selects one for replacement, guided by a Multiple Sequence Alignment (MSA) probability matrix that reflects naturally occurring substitutions.

The experimental results demonstrate the effectiveness of this computational approach. Over 16 independent runs and 20 generations, the algorithm consistently found improved designs. A clear convergence was observed, with the average fitness value (interaction energy) decreasing from an initial average of 15.78 to -2.77 in the final generation, indicating a significant enhancement in binding affinity. Analysis of the final 400 solutions revealed specific mutation patterns, identifying key amino acid positions and substitutions that contribute to lower energy states. Interface quality metrics further confirmed that the evolved protein designs had better energy scores without significantly altering the interface's structural packing. This work successfully validates the evolutionary algorithm as a tool for automating the search of optimal solutions in protein design, effectively highlighting energetically favorable mutations. As future work, extensive validation and benchmarks must be incorporated, along with transforming the genetic algorithm from single objective to multi-objective.



1: Algorithm convergence. *We calculate the mean fitness value (measured in Rosetta Energy Units) of all decoys in each generation. In only 20 generations, we observe an essential convergence of the fitness values.*

References:

[1] A. Ahmad, M. J. Stulberg, J. P. Mershon, D. S. Mollov, and Q. Huang, "Molecular and biological characterization of ϕRs551, a filamentous bacteriophage isolated from a race 3 biovar 2 strain of Ralstonia solanacearum," PLoS ONE, vol. 12, no. 9, Sep. 2017, doi: 10.1371/journal.pone.0185034.

[2] Y. Kang, H. Liu, S. Genin, M. A. Schell, and T. P. Denny, "Ralstonia solanacearum requires type 4 pili to adhere to multiple surfaces and for natural transformation and virulence," Molecular Microbiology, vol. 46, no. 2, pp. 427–437, 2002, doi: 10.1046/j.1365-2958.2002.03187.x.

(54) *De novo* design of transmembrane fluorescence-activating proteins

Jingyi Zhu[1,2,3,4*], Mingfu Liang[1,2,3,4*], Ke Sun[2,3,4*], Yu Wei[2,3,4], Ruiying Guo[2,3,4], Lijing Zhang[2,3,4], Junhui Shi[2,3,4], Dan Ma[2,3,4], Qi Hu[2,3,4], Gaoxingyu Huang[2,3,4], Peilong Lu[2,3,4#]

[1] College of Life Sciences, Zhejiang University, Hangzhou, Zhejiang, China
[2] Westlake Laboratory of Life Sciences and Biomedicine, Hangzhou, Zhejiang, China
[3] Key Laboratory of Structural Biology of Zhejiang Province, Research Center for Industries of the Future, School of Life Sciences, Westlake University, Hangzhou, Zhejiang, China
[4] Institute of Biology, Westlake Institute for Advanced Study, Hangzhou, Zhejiang, China
*These authors contributed equally to this work
liangmingfu@westlake.edu.cn

The recognition of ligands by transmembrane proteins is essential for the exchange of materials, energy, and information across biological membranes. Progress has been made in designing transmembrane proteins from scratch, as well as designing water-soluble proteins to bind small molecules, but de novo design of transmembrane proteins that tightly and specifically bind to small molecules remains an outstanding challenge. In this study, we present the accurate design of ligand-binding transmembrane proteins by integrating deep learning and energy-based methods. We designed pre-organized ligand-binding pockets in high-quality 4-helix backbones for a fluorogenic ligand, and generated transmembrane span using gradient-guided hallucination. The designer transmembrane proteins specifically activated fluorescence of the target fluorophore with mid-nanomolar affinity, exhibiting higher brightness and quantum yield compared to EGFP. These proteins were highly active in the membrane fraction of live bacterial and eukaryotic cells upon expression. Crystal and cryo-electron microscopy structures of the designer protein-ligand complexes were very close to the design model. We showed that the interactions between ligands and transmembrane proteins within the membrane can now be accurately designed. Our work paves the way for the creation of novel functional transmembrane proteins, with a wide range of applications including imaging, ligand sensing, and membrane transport.

(55) Application of deep learning-based protein design tools to generate soluble Ghrelin receptor constructs and Ghrelin receptor-targeting de novo mini binders

Hannes Junker, Florian J. Olander, Dominic Rieger, Clara T. Schoeder

Institue for Drug Discovery, Medical Faculty, Leipzig, Germany

hannes.junker@medizin.uni-leipzig.de

G protein-coupled receptors represent one of the pharmacologically most relevant protein families. Despite being targeted extensively, fine-tuning their signaling pathways remains a major unresolved challenge.

For this objective generating GPCR-targeting binders using generative modelling methods such as RFDiffusion is a promising approach, however, inducing signaling with these *de novo* binders showed limited success. GPCRs pose a particularly challenging class of targets due to their tight binding pockets, as well as their limited extracellular surface exposure. This complicates the design of binders capable of both binding the receptor and modulating its conformational ensemble.

Here, we present our computational pipeline to design *de novo* mini binders and peptides targeting both the active and inactive state of the Ghrelin receptor. This involves backbone generation with RFDiffusion, ProteinMPNN sequence design and subsequent *in silico* and experimental validation.

In a separate project we investigate the application of SolubleMPNN on the receptor with the aim to generate soluble constructs of the Ghrelin receptor in both active and inactive states. While this likely eliminates the signal transduction cascade, preseving the extracellular binding pocket during design holds the potential to facilitate faster screening of GPCR-targeting drug candidates.

(56) De novo design of a de novo nanopore steroid sensor

Thomas Guiziou[1], **Anastassia Vorobieva**[2]
[1] VIB-VUB Center for Structural Biology, Brussels, Belgium

thomas.guiziou@vub.be

Measuring biomarkers is critical for medical diagnostics and biosafety. Currently, the primary method for detecting biomarkers is LC-MS, which requires heavy machinery, expert workers, and high costs. A cheaper, user-friendly biosensing technology is needed, and nanopore fingerprinting is a promising alternative. Thus, we will build upon recent successes in nanopore design[1] and develop a nanopore capable of detecting and differentiating within a small molecule family. For this proof-of-concept, we will focus on steroids, due to their importance in human health and the impact of steroid endocrine disruptors on ecosystems.

We hypothesize the commonality between steroids will enable the design of a general binding site, while their differences will produce distinct electrophysiology signals, thus allowing for steroid fingerprinting. We will focus on a subset of steroids that share a common motif: the *A-B* rings with a C3 carbonyl. First, we will produce a library of TMB with a *de novo* helical domain. The lumen of the TMB is very constrained, making most backbone moves unfavorable. Thus, we start by docking helical fragments in the TMB, then resample their structure with RFdiffusion[2] and design the sequence with ProteinMPNN[3]. Increasing the constriction of the TMB will allow the nanopore to produce a higher electric signal. Experimental characterization for those designs is underway. Secondly, we will design the *de novo* steroid binding site using COMBS[4], focusing the design on the C3 carbonyl and the A-B rings of the steroids. Thirdly, we will record an ensemble of electrophysiology signals and build a classifier to trace the electric signal to the steroids.

This project will represent the first report of the *de novo* design of constriction in nanopores and proof of concept of a nanopore steroid biosensor. Such a biosensor could then be transformed into a commercial steroid sensor.

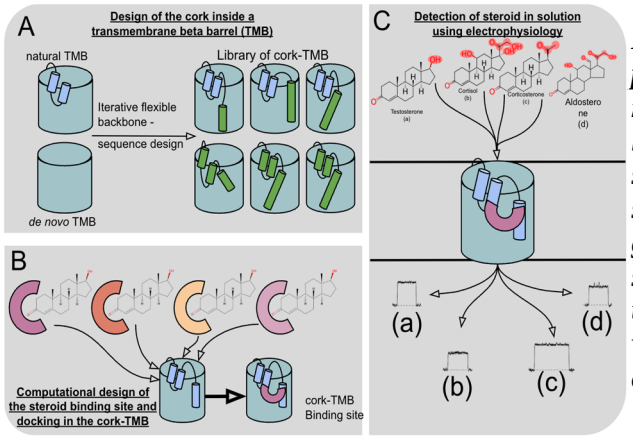


Figure 1: Scientific objectives of the proposal. *Design of the cork inside a transmembrane β- barrel (TMB) nanopore based on natural and de novo designed scaffolds (A). Design of a steroid binding site by targeting the common C3 carbonyl group (B). Detection and fingerprinting of steroid molecules by electro- physiology; the steroid core contributes to binding while variable substituents contribute to distinct signal features ("blockades")(C).*

References:

[1] A. A. Vorobieva *et al.*, "De novo design of transmembrane β barrels," *Science*, vol. 371, no. 6531, p. eabc8182, Feb. 2021.
[2] J. L. Watson *et al.*, "Broadly applicable and accurate protein design by integrating structure prediction networks and diffusion generative models," *bioRxiv*, Dec. 10, 2022. doi: 10.1101/2022.12.09.519842.
[3] J. Dauparas *et al.*, "Robust deep learning-based protein sequence design using ProteinMPNN," *Science*, vol. 378, no. 6615, pp. 49–56, Oct. 2022.
[4] N. F. Polizzi and W. F. DeGrado, "A defined structural unit enables de novo design of small-molecule-binding proteins," *Science*, vol. 369, no. 6508, pp. 1227–1233, Sep. 2020.

(57) *De novo* or *de naturae?* – Exploration of *de novo* protein properties and identification of natural proteins with similar stability

Johannes Klier[1,2], Dominic Rieger[1], Moritz Ertelt[1,3], Jens Meiler[1,2,3], Clara T. Schoeder[1,3]

[1] Institute for Drug Discovery, Leipzig University, Faculty of Medicine, Leipzig, Germany.
[2] School of Embedded Composite Artificial Intelligence, SECAI, Dresden/Leipzig, Germany.
[3] Center for Scalable Data Analytics and Artificial Intelligence ScaDS.AI, Dresden/Leipzig, Germany

klier@izbi.uni-leipzig.de

De novo proteins designed with computational methods, especially, with artificial intelligence, often exhibit high yields and high thermostability while displaying rigid and idealized structures, limiting their flexibility and subsequently their function. In this study, a classifier was trained on a set of *de novo* proteins using structure-derived features, indicating that *de novo* and natural proteins can still be differentiated. Turning this argument around, it was assessed whether proteins exist in nature through a search of the AlphaFold database that display similar properties as *de novo* proteins but are evolved. Organisms, such as archaea, that can be considered evolutionary old, possessed a higher proportion of *de novo*-like proteins. Expression of *de novo* classified natural proteins from different species also revealed that these proteins possess similarly high expression rates and thermostability compared to those of computationally designed *de novo proteins*. This finding supports the idea that *de novo* proteins represent only a small proportion of protein space and yet have to uncover more complex topologies and functions.

(58) Designing an enzymatic site into a dynamic protein

Boje Ingwersen[1], Estefani Gaytan Nuñez[1], Mark Kriegel[1], Birte Höcker[1]

[1] University of Bayreuth, Department of Biochemistry, Bayreuth, Germany

Boje.Ingwersen@uni-bayreuth.de

Many naturally occurring enzymes exhibit dynamic motions, ranging from large scale domain rearrangements to subtle switching of individual rotamers, to facilitate transition state stabilization [1] [2]. In conventional enzyme design, these dynamic contributions are often overlooked. Instead, the inside out methodology focuses on stabilizing a single conformation that supports one or more transition states, thereby enabling catalysis [3].

Here, we implement the Kemp elimination reaction into the dynamic, two-lobed periplasmic binding protein PotF [4]. Utilizing the conventional single state approach, we show that the Kemp elimination reaction can be introduced into dynamic proteins. However, our kinetic and structural data suggests that our design approach leaves room for improvement in achieving optimal catalytic efficiency and conformational control. We give an outlook on future refinements of the methodology, utilizing multi state approaches. This work explores how state-of-the-art design and prediction tools can extend the traditional single-state design workflow to better emulate the dynamic behaviour observed in natural enzymes.

References:

[1] Eisenmesser, E.Z. *et al.* Intrinsic dynamics of an enzyme underlies catalysis. *Nature* 438, 117-121 (2005).

[2] Agarwal, P. K., Bernard, D. N., Bafna, K. & Doucet, N. Enzyme Dynamics: Looking Beyond a Single Structure. *ChemCatChem* 12, 4704–4720 (2020).

[3] Kiss, G., Çelebi-Ölçüm, N., Moretti, R., Baker, D. & Houk, K. N. Computational Enzyme Design. *Angew. Chem. Int. Ed.* 52, 5700–5725 (2013).

[4] Kröger, P., Shanmugaratnam, S., Ferruz, N., Schweimer, K. & Höcker, B. A comprehensive binding study illustrates ligand recognition in the periplasmic binding protein PotF. *Structure* 29, 433-443.e4 (2021).

(59) Redesign of hard-to-express proteins from fungal or plant origin to yield folded protein in bacteria

Christoph Sojer[1], Ahmed Alagic[1], Gustav Oberdorfer[1], Horst Lechner[1]

[1]Graz University of Technology, Institute of Biochemistry, Graz, Austria

csojer@student.tugraz.at

Many proteins of interest are marginally stable or challenging to express in heterologous hosts, limiting their use in biocatalysis and making them poor candidates for directed evolution. This project aims to successfully produce proteins of fungal and plant origin in heterologous expression systems.

Instability often arises from intrinsic thermodynamic properties and evolutionary pressures specific to the organism's environment. Existing computational tools like PROSS [1] and FireProt2 [2] can enhance protein stability and expression but rely on multiple sequence alignments, making them less effective for proteins with limited homologous sequences or from organisms that lack selective pressure to evolve stable proteins.

This project seeks to develop a robust, generalizable method using machine-learning-based sequence design algorithms to improve protein production yields and stability. Advanced tools such as ProteinMPNN [3] and Frame2seq [4] will be applied. We will optimize three challenging heme-containing unspecific peroxygenases (UPOs; E.C. 1.11.2.1), including two that have resisted functional heterologous expression in *Escherichia coli*. Our chosen targets serve as ideal candidates to validate our approach and also their relevance in biocatalysis is generally known in the community.

This approach is anticipated to produce stable, highly expressible proteins that are independent of evolutionary constraints, offering valuable starting points for directed evolution campaigns and facilitating the study of hard-to-produce enzymes.

References:

[1] A. Goldenzweig *et al.*, "Automated Structure- and Sequence-Based Design of Proteins for High Bacterial Expression and Stability," *Mol. Cell*, vol. 63, no. 2, pp. 337–346, July 2016, doi: 10.1016/j.molcel.2016.06.012.

[2] M. Musil *et al.*, "FireProt 2.0: web-based platform for the fully automated design of thermostable proteins," *Brief. Bioinform.*, vol. 25, no. 1, p. bbad425, Jan. 2024, doi: 10.1093/bib/bbad425.

[3] J. Dauparas *et al.*, "Robust deep learning–based protein sequence design using ProteinMPNN," *Science*, vol. 378, no. 6615, pp. 49–56, Oct. 2022, doi: 10.1126/science.add2187.

[4] D. Akpinaroglu, K. Seki, A. Guo, E. Zhu, M. J. S. Kelly, and T. Kortemme, "Structure-conditioned masked language models for protein sequence design generalize beyond the native sequence space," Dec. 15, 2023, *bioRxiv*. doi: 10.1101/2023.12.15.571823.

(60) Guiding structure-based protein design with UniProt annotations

Fernando A. Meireles[1], Parth Bibekar[1], Lucien F. Krapp[1], Matteo Dal Peraro[1]

Laboratory for Biomolecular Modeling, EPFL, Lausanne, Switzerland

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.

(61) Creating and comprehending *de novo* flavocytochromes

Ethan Bungay[1], Meiyue Liu[2], Julea Butt[3], Ross Anderson[1]

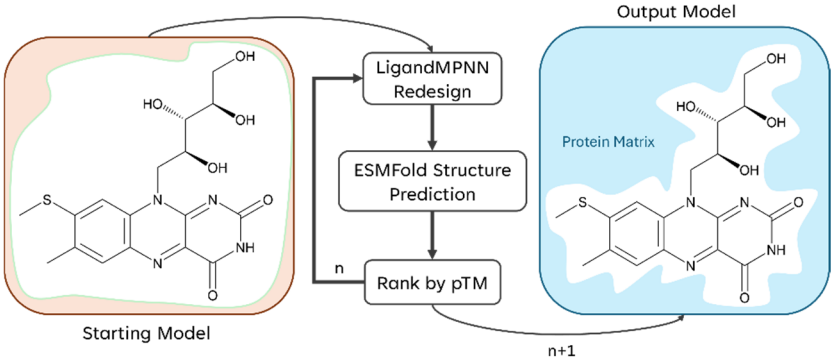
[1] University of Bristol, School of Biochemistry, Bristol, UK
[2] University of Bristol, School of Chemistry, Bristol, UK
[3] University of East Anglia, School of Chemistry and School of Biological Sciences, Norwich, UK

ethan.bungay@bristol.ac.uk

De novo redox protein design provides an alternate route towards understanding natural bioenergetic pathways and engineering novel molecular machinery. Design within the Anderson group focuses on 4-helix bundles (4HBs), these 'maquettes', inspired by a core subunit of transmembrane oxidoreductases, are free of evolutionary complexity, creating a blank slate for bottom-up design. The group's blueprint is the diheme 4D2, which has since been elongated and mutated to configure alternative ligands, with previous work investigating its amenability for covalent flavin attachment [1,2]. Flavins boast desirable bioenergetic properties, including photosensitivity and electron bifurcation, but their amphipathic structure hindered previous attempts to bind them uniformly into these minimalist scaffolds.

This work presents a suite of novel flavocytochromes, created through an *in silico* directed-evolution pipeline (Figure 1) leveraging advances in both protein design and structure prediction. By integrating a covalent flavin site in an optimised cytochrome scaffold, we have developed proteins that bolster impressive thermostability, heme-affinity, and structural homogeneity. By combining data from EPR (electron paramagnetic resonance), PFE (protein film electrochemistry), and ultrafast transient absorption spectroscopy, we can probe the flavin's redox states and dynamics, down to the femtosecond timescale. Correlating this to structural data, future work aims to use the electrochemical behaviour to guide the next generation of design, looking to achieve longer-lived semiquinones, altered redox potentials, and potentially unlock novel catalytic functions.

Figure 1: *Example of an iterative pipeline employed to introduce stabilising interactions to a covalent flavin binder*



References:

[1] Hutchins, G. H.; Noble, C. E. M.; Bunzel, H. A.; Williams, C.; Dubiel, P.; Yadav, S. K. N.; Molinaro, P. M.; Barringer, R.; Blackburn, H.; Hardy, B. J. An expandable, modular de novo protein platform for precision redox engineering. *Proc. Natl. Acad. Sci. U. S. A.* 2023, *120* (31), e2306046120 DOI: 10.1073/pnas.2306046120

[2] Dubiel, P. *De novo* design of photoactivatable electron transfer proteins. Ph.D. Thesis, University of Bristol, Bristol, UK.

(62) *De novo* designed protein binders for improved sensitivity and modularity in competition-based assays

Tallie Godschalk[1,2*], Yosta de Stigter[1,2*] and Maarten Merkx[1,2]

[1] Laboratory of Chemical Biology, Department of Biomedical Engineering, Eindhoven University of Technology, The Netherlands
[2] Institute for Complex Molecular Systems, Eindhoven University of Technology, The Netherlands

t.godschalk@tue.nl

Competition-based assays are a common strategy for detecting small molecules. However, these assays generally rely on the availability of a structural mimic or labeled version of the target analyte that binds to the same epitope of an (immobilized) binder to displace the target and generate a readout signal. The similar binding affinity of the target and competitor analyte inherently limits assay sensitivity, as competitor displacement requires relatively high target concentrations. Recent developments in deep learning-based protein design approaches, including RFdiffusion [1] and BindCraft [2], have enabled the rapid generation of custom **protein binders** against a wide range of molecular targets focused on specific binding epitopes. These protein-based binders offer a promising alternative to traditional competitor domains, as their binding properties can be readily tuned and optimized without the need for chemical modification or labeling. We tested the implementation of *de novo* designed binders as **synthetic competitors** in a solution-based competition assay with a bioluminescent readout (Figure 1) [3]. In this assay, displacement of the competitor by the target analyte enables the reconstitution of a split luciferase, resulting in the emission of bright blue light. Using digoxin as a model analyte, we demonstrate **enhanced assay sensitivity** compared to conventional formats. This approach circumvents the need for chemically modified competitors and offers a more **modular framework** adaptable to a broad range of small molecule (and protein) targets.

Figure 1: Schematic overview bioluminescent competition assay for small molecule detection. *The sensor domain consisting of the capture protein fused via flexible peptide linkers to the luciferase fragments is kept in a dark off-state by the de novo competitor domain. Target detection displaces the competitor, resulting in reconstitution of the luciferase, and the emission of blue light.*

References:

[1] J. L. Watson et al, "De novo design of protein structure and function with RFdiffusion", *Nature*, 620, 1089–1100 (2023)
[2] M. Pacesa et al, "BindCraft: one-shot design of functional protein binders", *bioRxiv*, 2024.09.30.615802
[3] E. A. van Aalen et al, "Turning Antibodies into Ratiometric Bioluminescent Sensors for Competition-Based Homogeneous Immunoassays", *ACS Sens.*, vol. 9, 1401–1409, (2024)

(63) Variational Autoencoder-Based Generation of Loop Conformations in Antibodies and Other Proteins

Johannes Peterlechner[1], Gustav Oberdorfer[1]

[1] Technical University of Graz, Institute of Biochemistry, Graz Austria

johannes.peterlechner@tugraz.at

Predicting flexible and disordered protein regions, especially loop structures remain challenging due to their variability in sequence and conformation. Yet loops often mediate ligand binding, catalysis, and molecular recognition, making accurate modeling crucial for functional protein design.

ln this study we introduce a Variational Autoencoder for protein structure generation, trained on a large dataset of PDB-derived loops. The learned latent space captures structural relationships, enabling sampling of diverse geometries and identification of similar conformations.

The VAE can generate, diverse loop conformations, which was applied on the sampling of Ty1 Nanobody CDR loops. Generated loops preserved backbone conformation while varying in detail.

Results show VAEs are effective for stochastic sampling, supporting both targeted and diversity-maximized generation and have strong potential for integration into rapid protein design workflows.

(64) Application of Deep-Learning-Based Tools for Vector Engineering

Pablo Herrera-Nieto[1], José Gallardo[1], Mark J. van Raaij[1], Carmen San Martín[1]

[1] Department of Macromolecular Structures, Centro Nacional de Biotecnología (CNB-CSIC), 28049 Madrid, Spain

pablo.herrera@cnb.csic.es

Recombinant human adenovirus vectors have been extensively studied for gene therapy, oncolytic virotherapy, and vaccine delivery. However, their clinical utility is often limited by strong host immune responses and suboptimal tissue specificity[1]. To overcome these challenges, the adenovirus capsid—composed of hexon, penton base, and fiber proteins—must be re-engineered to maintain structural integrity while reducing off-target interactions and immunogenic epitopes.

Recent advances in deep-learning–based protein structure prediction and de novo design, including RFdiffusion [2], AlphaProteo [3], and BindCraft [4], are enabling precise modifications of viral capsid components. While these tools have shown success with single-chain proteins [5] and simpler viral vectors such as adeno-associated viruses [4], adapting them for large, multimeric assemblies like adenovirus remains an ongoing challenge. Adenovirus capsid engineering requires balancing structural stability with functional modifications across multiple interacting components.

Here, we present a computational-to-experimental workflow to design, evaluate, and refine novel adenovirus capsid variants—termed "intelligent adenovirus" (iAds). Our strategy focuses on fiber knob engineering using BindCraft to optimize receptor interactions and modify tropism while preserving capsid stability (Figure 1). By integrating in silico design with iterative screening, we aim to generate vectors with enhanced specificity and reduced immunogenicity for transgene delivery. Ongoing work involves validating these designs through biochemical and functional assays, focusing on receptor binding, immune evasion, and stability to assess their potential for therapeutic applications.

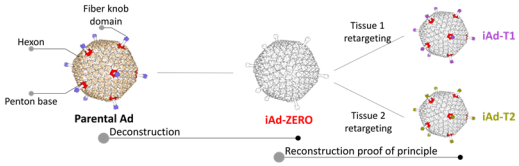


Figure 1. *Development pipeline for new adenoviral vectors.*

References:

[1] Yamamoto, Y., Nagasato, M., Yoshida, T., & Aoki, K. (2017). Recent advances in genetic modification of adenovirus vectors for cancer treatment. Cancer science, 108(5), 831-837.

Watson, J. L., Juergens, D., Bennett, N. R., Trippe, B. L., Yim, J., Eisenach, H. E., ... & Baker, [2] D. (2023). De novo design of protein structure and function with RFdiffusion. Nature, 620(7976), 1089-1100.

[3] Zambaldi, V., La, D., Chu, A. E., Patani, H., Danson, A. E., Kwan, T. O., ... & Wang, J. (2024). De novo design of high-affinity protein binders with AlphaProteo. arXiv preprint arXiv:2409.08022.

[4] Pacesa, M., Nickel, L., Schellhaas, C., Schmidt, J., Pyatova, E., Kissling, L., ... & Correia, B. E. (2024). BindCraft: one-shot design of functional protein binders. bioRxiv, 2024-09.

[5] Vázquez Torres, S., Benard Valle, M., Mackessy, S. P., Menzies, S. K., Casewell, N. R., Ahmadi, S., ... & Baker, D. (2025). De novo designed proteins neutralize lethal snake venom toxins. Nature, 1-7.

(65) *De novo* design of anchors that bind to termini of designed protein fibres

Alina Konstantinova[1,2], Neža Ribnikar[2], Ajasja Ljubetič[1,3]

[1] National Institute of Chemistry, Department of Synthetic Biology and Immunology, Ljubljana, Slovenia
[2] University of Ljubljana, faculty of medicine, Ljubljana, Slovenia
[3] EN-FIST Centre of Excellence, Ljubljana, Slovenia

alina.konstantinova@ki.si, ajasja.ljubetic@ki.si

Protein nanomachines perform a great variety of functions in living organisms, mediating vital cellular processes, such as cell division, maintaining structure or moving cargo inside the cell. De novo protein design has seen considerable development in recent years, and opens up many possibilities for designing functional proteins from scratch.

Our goal is to design protein nanomachines capable of performing specific functions, such as directed movement and moving cargo. For that we're using a designed protein system consisting of de novo designed fibre tracks and symmetric "walker" proteins, moving along them. In this work we aim to design symmetric oligomeric anchors that would specifically attach to fibre ends, halting its assembly, which would enable controllable and directed growth and attachment.

In our design pipeline we used a combination of RFDiffusion[1], ProteinMPNN[2] and AlphaFold2[3] to design cyclical symmetric oligomers with symmetry roughly corresponding to that of a fibre track, that would flexibly attach to fibre monomers, forming an "anchor". After multiple rounds of computational design, we selected best ranking oligomers for experimental evaluation. We tested the oligomeric state using SEC and mass photometry.

Designs capable of forming correct oligomeric state were further characterized with cryo-EM. Selected designs were subsequently used in oligomeric anchors comprised of a symmetric oligomer fused to fibre "caps" with a flexible linker. Fibre "caps" were created by redesigning one of the interfaces of interaction within fibre monomers, so that caps interacting with fibre ends interfere with their further assembly. Various anchor's interactions with de novo designed fibres have been tested in vitro using TIRF microscopy, in order to evaluate the assembly and behaviour of anchors, as well as their colocalization with fibre ends.

References:

[1] Watson, J. L. et al. De novo design of protein structure and function with RFdiffusion. Nature 620, 1089–1100 (2023).
[2] Dauparas, J. et al. Robust deep learning–based protein sequence design using ProteinMPNN. Science 378, 49–56 (2022).
[3] Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. Nature 596, 583–589 (2021).

(66) De novo design of a Michealase by simultaneous catalytic array scaffolding and substrate affinity optimization

David Stoll[1], Adrian Tripp[1], Markus Braun[1], Noelia Ferruz[2], Gustav Oberdorfer[1] Horst Lechner[1],

[1]Institute of Biochemistry, Graz University of Technology, Petersgasse 12/2, 8010-Graz, Austria
[2]Systems and Synthetic Biology, Centre for Genomics Regulation, C/ Dr. Aiguader, 88 08003 Barcelona, Spain

david.stoll@student.tugraz.at

The Michael addition reaction represents one of the most fundamental carbon-carbon bond forming reactions in organic chemistry, widely utilized in pharmaceutical synthesis and natural product formation. Despite its synthetic importance, naturally occurring enzymes that catalyze intermolecular Michael additions (Michaelases) are extremely rare, creating significant interest in developing artificial biocatalysts for this transformation. Such enzymes could enable sustainable, stereoselective synthesis under mild aqueous conditions. Here we describe the de novo design of a Michaelase. Our approach follows the previously reported RiffDiff[1] approach and expands it by integrating a GNINA-based molecular-docking metric into the pipeline, to enable structure-guided active-site residues optimization for enhanced atalytic efficiency.[1] To improve substrate binding specificity, we incorporated GNINA's scoring function into the selection workflow, evaluating each design for predicted binding affinity. This additional filtering enhanced the computational quality metrics of the generated designs. Active site preorganization was assessed by using HTMD molecular dynamics simulations for the apo and holo states of the designs. This also allowed to determine the distribution of Bürgi-Dunitz and Flippin-Lodge angles as well as the Bürgi-Dunitz distance, which describe the nucleophilic attack from the active site lysine to the substrate's carbonyl carbon. A subset of the designed sequences were experimentally validated for Michael addition activity using a colorimetric turn-on probe.[2] This integrated computational-experimental approach demonstrates the potential for designing novel enzymes to catalyze synthetically valuable reactions absent in nature.

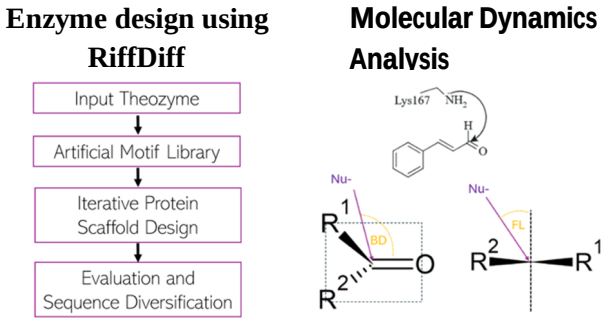


Figure 1: Workflow of Michaelase Design and Evaluation. *This figure shows the workflow. The input theozyme is based on the active site of DERA-MA*[3] *(PDB entry 7P76). GNINA was implemented in the evaluation and sequence diversification steps. The MD analysis was performed using python based HTMD.*

References:

[1] Braun, M. L., Tripp, K. W. et al. Computational enzyme design by catalytic motif scaffolding. bioRxiv preprint at https://doi.org/10.1101/2024.08.02.606416 (2024)
[2] Biewenga, L. et al. Selective colorimetric 'turn-on' probe for efficient engineering of iminium biocatalysis. *Nat. Chem.* **14**, 927–935 (2022).
[3] Kunzendorf, A. et al. Unlocking asymmetric Michael additions in an archetypical class I aldolase by directed evolution. *Nat. Commun.* **12**, 5177 (2021).

(67) Rapid screening of *de novo* designed binders using split luciferase complementation

Yosta de Stigter[1,2*], Tallie Godschalk[1,2*] and Maarten Merkx[1,2]

[1] Laboratory of Chemical Biology, Department of Biomedical Engineering, Eindhoven University of Technology, The Netherlands.
[2] Institute for Complex Molecular Systems, Eindhoven University of Technology, The Netherlands.

y.d.stigter@tue.nl

Recent advances in deep learning-based protein design have made considerable progress in designing **protein binders**. State-of-the-art methods such as RFdiffusion [1] or Bindcraft [2] show impressive success rates, but obtaining a binder with the desired affinity and/or kinetic profile still requires experimental screening of tens to hunderds of designs. Biochemical characterization of all candidates is time consuming and often unfeasbile for labs without a sophisticated high-throughput infrastructure. To facilitate *de novo* binder screening, we developed a simple **bioluminescent screening assay** based on binding-induced complementation of split luciferase fragments, genetically fused to binder and target proteins. Binding of the *de novo* binder to its target brings the luciferase fragments into close proximity, reconsistuting the luciferase protein to emit bright blue light (Figure 1) [3]. The increase in light can directly be used as a measure for binding and is easily recorded using a standard platereader or digital camera. We tested our assay to screen 20 *de novo* designed binders as competitors for an antibody binding site, which **successfully identified hits** and allowed for subsequent titrations experiments to determine binding affinities. Our assays performed similar when testing directly in crude cell lysate, showing that laborious protein purification is not neccessary for initial screening. With its **easy-to-implement** workflow and compatibility with **complex sample matrices**, our bioluminescent screenings assay offers a fast, attractive alternative for the initial screening and chacterization of *de novo* designed binders.

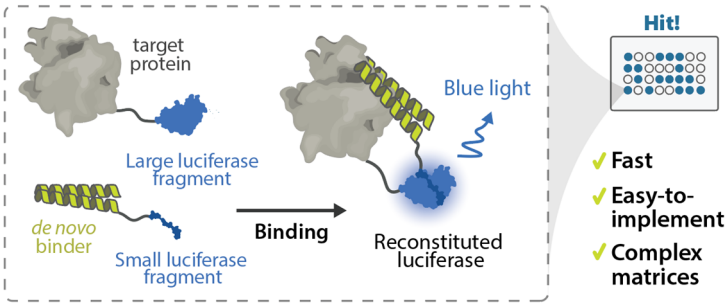


Figure 1: Schematic of bioluminescent binder screening assay. *Target protein and de novo binder are fused to split luciferase fragments that upon binding reconstitute and emit blue light.*

References:

[1] Watson, J. L. et al. De novo design of protein structure and function with RFdiffusion. *Nature,* 620, 1089–1100 (2023)
[2] Pacesa, M. et al. BindCraft: one-shot design of functional protein binders. *bioRxiv*, 2024.09.30.615802
[3] Dixon, A. S. et al. NanoLuc complementation reporter optimized for accurate measurement of protein interactions in cells. *ACS Chem. Biol.* 11, 400–408 (2016).

(68) Investigation of the Binding Pocket in a Polyester Hydrolase

Norbert Graefe[1], Georg Künze[1,2,3]

[1] Leipzig University, Institute for Drug Discovery, Medical Faculty
[2] Interdisciplinary Center for Bioinformatics, Leipzig University, 04107 Leipzig, Germany
[3] Center for Scalable Data Analytics and Artificial Intelligence, Leipzig University, 04105 Leipzig, , Germany

norbert.graefe@medicine.uni-leipzig.de

The hydrolysis of post-consumer plastics by PET hydrolases is a promising strategy for the closed-loop recycling of polyethylene terephthalate (PET). In this study, we investigated the role of critical residues in the binding pocket of the Polyester Hydrolase Leipzig 7 (PHL7) [1].

Within polyester hydrolases the β8-α6 loop region, plays a key role in catalytic function, as it harbors the catalytic histidine and regulates access to the active site. A major structural difference between type I and type II PET hydrolases lies in this loop. Type I enzymes contain two to three fewer amino acids, resulting in a more rigid loop structure and a more defined active site geometry. To account for the additional flexibility, type II PET hydrolases require a disulfide bridge connecting the β8-α6 loop with the β7-α5 loop, on which the catalytic aspartate resides, to maintain the active site integrity [2].

We performed an evolutionary conservation analysis and introduced naturally occurring amino acids into the loop region of PHL7. Thereby, we observed that mutations that broaden the entrance to the binding pocket positively influence the enzyme´s activity.

In a subsequent approach, we broadened the search space to investigate more residues of the binding pocket by employing htFuncLib, which first applies phylogenetic filtering to select candidate residues before evaluating them using the Rosetta energy function [3]. This study examines how the introduced mutations within the binding pocket affect the catalytic activity and thermostability of PHL7 through rational modification of its amino acid sequence.

References:

[1] Sonnendecker, Christian, et al. "Low carbon footprint recycling of post-consumer PET plastic with a metagenomic polyester hydrolase." ChemSusChem 15.9 (2022): e202101062.
[2] Fecker, Tobias, et al. "Active site flexibility as a hallmark for efficient PET degradation by I. sakaiensis PETase." Biophysical journal 114.6 (2018): 1302-1312.
[3] Weinstein, Jonathan Yaacov, et al. "Designed active-site library reveals thousands of functional GFP variants." Nature communications 14.1 (2023): 2890.

(69) Training-free approach to *de novo* design of protein binders

Kateryna Maksymenko[1], Julia Skokowa[2], Andrei N. Lupas[1], Mohammad ElGamacy[2]

[1] Max Planck Institute for Biology Tübingen, Department of Protein Evolution, Tübingen, Germany
[2] University Hospital Tübingen, Division of Translational Oncology, Tübingen, Germany

kmaksymenko@tuebingen.mpg.de

Computational design of on-demand protein binders can address a wide range of medical and biotechnological needs. However, *de novo* binder design remains an outstanding challenge. Recent progress in this area relies mainly on machine learning tools that have two major limitations: 1) they are constrained by the structural and binding patterns learned from natural complexes, and 2) their output is difficult to rationalize. Therefore, we sought to develop a training-free solution to the binder design problem.

Our design pipeline first identifies suitable scaffolds with high complementarity to the target epitope using a massive-scale docking software (HECTOR). Next, the selected scaffold is mutated to carve a binding site *de novo.* Sequence design is performed by the Damietta engine [1], which employs tensorized energy calculations. Finally, design candidates undergo rigorous in silico validation through accelerated molecular dynamics.

We used the described framework to design binders against two oncogenic targets, VEGF and IL-7Rα. Experimental testing of several candidates identified binders with nanomolar affinities to their targets and strong inhibiting activities of the downstream signaling [2]. By relying on first principles, our design approach can be readily extended to small-molecule binders, as well as synthetic binders that incorporate artificial amino acids.

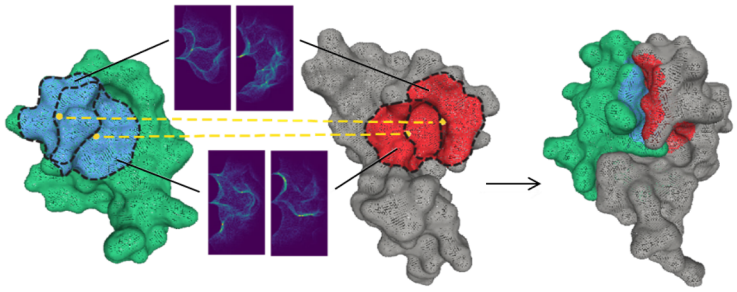


Figure 1: The design workflow decouples the docking step from interface design. *The docking step relies on the HECTOR fingerprinting method that analytically identifies surface patches complementary to a given target patch.*

References:

[1] I. Grin et al. The Damietta Server: a comprehensive protein design toolkit. *Nucleic Acids Research*, 52(W1), W200-W206, 2024.
[2] K. Maksymenko et al. A complementarity-based Approach to De Novo Binder Design. *Advanced Science*, e02015, 2025.

(70) Computational design of HIV-1 Envelope germline targeting nanoparticles.

Svearike Oeverdieck[1,2,3], Sebastian Ols[3], Yoann Aldon[1,2], Neil King[3], Rogier Sanders[1,2]

[1] Department of Medical Microbiology, Amsterdam UMC, University of Amsterdam
[2] Amsterdam Institute for Infection and Immunity, Infectious Diseases, University of Amsterdam
[3] Institute for Protein Design, University of Washington

s.g.a.oeverdieck@amsterdamumc.nl

HIV-1 remains a major global public health issue as there is no cure or vaccine. The Envelope glycoprotein (Env) presented on the surface of the virions or infected cells is the primary target for neutralizing antibodies and the focus of HIV-1 vaccine development [1]. Although broadly neutralizing antibodies (bNAbs) have been isolated and extensively characterized from infected individuals, their induction through vaccination remains challenging [1].

Env apex, especially the V1V2 supersite, is an attractive epitope since it is the target of several potent bNAbs like PG9. However, when using the entire Env as a vaccine candidate the immune response often directs itself towards immuno-dominant epitopes such as the exposed base of the Env trimer and/or strain specific glycan holes [2]. Our work here centers on the redesign and scaffolding of the V1V2 supersite into nanoparticle building blocks, utilizing computational methods like RFdiffusion, so that the whole nanoparticle surface is our epitope of interest. We hypothesize that this will focus the immune response after vaccination towards these epitopes and enhance the affinity of the elicited bNAbs. Using nanoparticles in this scheme will enable us to multivalently display the desired epitopes and increase vaccine efficacy [3].

References:

[1] T. G. Caniels et. al. Germline-Targeting HIV Vaccination Induces Neutralizing Antibodies to the CD4 Binding Site. *Science Immunology* 30;9, 2024
[2] E.I.M.M. Reiss et al. Fine-mapping the immunodominant antibody epitopes on consensus sequence-based HIV-1 envelope trimer vaccine candidates. *npj Vaccines* **7**, 152, 2022
[3] M. Brinkkemper et al. Mosaic and mixed HIV-1 glycoprotein nanoparticles elicit antibody responses to broadly neutralizing epitopes. *PLoS Pathog* 3;20, 2024

(71) Prediction of pre- and postfusion conformations of class I fusion proteins with AlphaFold

Sevilay Güleşen[1], Victoria Most[1], Jens Meiler[1,2,3,4*], Clara T. Schoeder[1,2*]

[1] Institute for Drug Discovery, Faculty of Medicine, Leipzig University, Leipzig, Germany
[2] Center for Scalable Data Analytics and Artificial Intelligence ScaDS.AI, Dresden/Leipzig, Germany
[3] Department of Chemistry, Department of Pharmacology, Center for Structural Biology, Institute of Chemical Biology, Center for Applied Artificial Intelligence in Protein Dynamics, Vanderbilt University, Nashville, Tennessee, United States of America
[4] Institute for Computer Science, Faculty of Mathematics and Informatics, University Leipzig, Leipzig, Germany; Wilhelm Ostwald Institute for Physical and Theoretical Chemistry, Faculty of Chemistry and Mineralogy, University Leipzig, Leipzig, Germany
[*] authors contributed equally

sevilay.guelesen@medizin.uni-leipzig.de

Arenaviruses such as Marburg virus (MARV) use their surface glycoproteins (GPs) to mediate attachment to host cells and fusion of the viral membrane[1]. Epitopes on the viral GP serve as targets for the humoral immune response, particularly antibodies. During the fusion process, the GP undergoes conformational changes triggered by pH differences. These structural states are typically classified as **prefusion**, intermediate (middle), and **postfusion conformations**. However, experimentally determined different conformations of GP structures are underrepresented in the RCSB especially for **class I fusion proteins** [2]. Different conformations such as postfusion structures are critical for computational prediction of possible epitope regions. Recent advances in computational protein structure prediction have greatly improved the ease and accuracy of modeling these conformations. In this project, we observed that **AlphaFold**2-multimer (AF2-M) can predict different GP conformations, but the overall frequency of states other than the prefusion conformation predictions is low [3]. Therefore, we hypothesized that adapting AF2-M sampling is necessary to enrich for these desired conformations. The primary objective of this study is to manipulate AF2-M to predict multiple conformations of a given GP. AF2-M requires sequence input and internally uses multiple sequence alignments, trained weights, or template information. We tested the use of template data to enrich for pre- or postfusion conformations and demonstrated that our approach enables highly accurate prediction of class I fusion protein structures in both states, with the template dataset playing a crucial role in guiding modeling toward the intended conformation. Additionally, we have shown that the lack of correlation between pLDDT and TM-scores suggests that low pLDDT values may obscure the presence of valid alternative conformations.

References:

[1] J. M. White, S. E. Delos, M. Brecher, and K. Schornberg, “Structures and Mechanisms of Viral Membrane Fusion Proteins: Multiple Variations on a Common Theme,” *Crit Rev Biochem Mol Biol*, vol. 43, no. 3, pp. 189–219, Jan. 2008, doi: 10.1080/10409230802058320.
[2] N. Blais *et al.*, “Characterization of Pre-F-GCN4t, a Modified Human Respiratory Syncytial Virus Fusion Protein Stabilized in a Noncleaved Prefusion Conformation,” *J Virol*, vol. 91, no. 13, Jul. 2017, doi: 10.1128/JVI.02437-16.
[3] R. Evans *et al.*, “Protein complex prediction with AlphaFold-Multimer,” 2021, doi: 10.1101/2021.10.04.463034.

(72) Designing protein binders to instriniscally disordered regions for modulating condensation of DDX4

Christian Peder Jacobsen [a, b*], Rasmus Krogh Norrild [a], Alexander Kai Buell [a], Timothy Patrick Jenkins [a], Morten Gronbech Rasch [b], Andreas Hald [b]

[a] Department of Biotechnology and Biomedicine, Technical Universityof Denmark, 2800 Kgs Lyngby, Denmark

[b] Department of Protein Research, Novo Nordisk, Novo Nordisk Park 1 2760 Måløv, Denmark

DDX4 is an RNA helicase containing an intrinsically disordered region (IDR) that promotes biomolecular condensate formation. Recently, collaborators Norrild et al. used mRNA display of overlapping peptides to map regions within DDX4 that drive condensation, based on partition free energy measurements. This condensation map presents a potential strategy for identifying therapeutically relevant sites in diseases associated with IDR-mediated condensation and for elucidating the mechanism behind IDR mediated molecular interactions. Building on this map, we aim to de novo design protein binders that inhibit IDR-driven condensation - both to validate the map's utility for therapeutic targeting and to assess whether designed binders can effectively reduce condensate formation. Several binder design strategies have been employed to target the IDR of DDX4. These include RFdiffusion-based one-sided and two-sided design approaches, as well as a method that embeds peptides into predefined scaffold binding clefts and redesigns interface residues using inverse folding and molecular dynamics simulations. In addition, we have explored hallucination-based design strategies guided by an inhouse protein–protein interaction predictor trained using contrastive learning on protein-language model embeddings from ~60.000 protein-interaction obtained from the PDB. Diverse binder libraries will be synthesized as gene pools and screened using a high-throughput yeast surface display system capable of testing thousands of designs. High-affinity candidates will be enriched through iterative selection and next-generation sequencing, after which selected binders will be validated based on their ability to modulate DDX4 aggregation in vitro and bind target peptides using biolayer interferometry. This project aims to demonstrate how a condensation map can guide therapeutic discovery by enabling rational targeting of IDR regions driven condensation, while also serving as a benchmar k for evaluating protein binder design strategies against the challenging target space of intrinsically disordered regions.

(73) Computational Design of Programmable Sequence-Specific RNA Binding Proteins

Iori Morita[1], Florian Praetorius[1]

[1] Institute of Science and Technology Austria, Klosterneuburg, Austria
iori.morita@ista.ac.at

Detection and regulation of ssRNA is of great interest in synthetic and cell biology. It is highly desirable to develop a programmable ssRNA-binding protein. Taking inspiration from natural RNA-binding proteins, we propose to build a programmable ssRNA-binding protein.

With this goal in mind, we take two different approaches: (A) refining naturally existing modular RNA binders, and (B) building a new-to-nature scaffold based on natural RNA-binding motifs.

In the approach (A), we utilized PUF family proteins [1], the naturally occurring modular RNA binders. As shown in Figure 1a, the proteins have a repetitive, arch-shaped structure, and each repeat unit is responsible for the recognition of one RNA base. We extracted the repeat units from X-ray structures, copied and rearranged them into an arch. The amino acid sequences were designed with LigandMPNN. After testing 48 such designs by fluorescence polarization assays, we identified one protein that binds to poly(U) with K_d < 1 nM. We are currently testing variants of this design to obtain proteins that bind to other RNA sequences.

In the approach (B), we extracted RNA-binding motifs from PDB, copied and arranged them repetitively in 3D space. We used RiffDiff [2] to identify helix fragments that accommodate the motifs, and built the rest of the scaffold with RFdiffusion. The amino acid sequences were again designed with LigandMPNN. As a result, we have gotten much more diverse structures *in silico*, and we are currently working on the experimental screening.

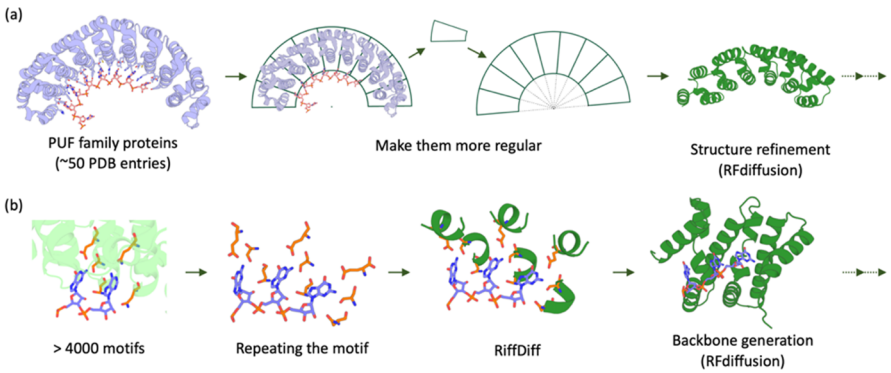


Figure 1: Computational approaches to design programmable RNA binders. *(a) Refining the naturally occurring modular RNA binders. (b) Building a new-to-nature scaffold around natural RNA-binding motifs.*

References:

[1] X. Wang, J. McLachlan, P. D. Zamore, T. M. Tanaka Hall, Modular Recognition of RNA by a Human Pumilio-Homology Domain, *Cell*, **2002**, *110*, 501-512.

[2] M. Braun, A. Tripp, M. Chakatok, S. Kaltenbrunner, C. Fischer, D. Stoll, A. Bijelic, W. Elaily, M. G. Totaro, M. Moser, S. Y. Hoch, H. Lechner, F. Rossi, M. Aleotti, M. Hall, G. Oberdorfer, Computational enzyme design by catalytic motif scaffolding. *bioRxiv*, **2024**.

(74) Uncovering and Leveraging Bias in Structure-Conditioned Protein Design Models

Laura Dillon[1], Oliver Crook[1]

[1]Department of Chemistry, University of Oxford, South Parks Road, Oxford, United Kingdom.

laura.dillon@chem.ox.ac

AI-driven protein design models, including tools like ProteinMPNN and various language models, have achieved remarkable success [1]. However, these models encode systematic biases that remain poorly understood [2]. Our research investigates how a model's underlying architecture—specifically, structure-conditioned versus sequence-only—creates distinct bias patterns that shape design outputs.

Using variance decomposition on 8,000 proteins across 144 species, we quantify the taxonomic information each model retains. Our analysis reveals that structure-conditioned models like ProteinMPNN retain minimal unexplained species-level variance (<3%), whereas sequence-only models like ESM2 show substantial taxonomic bias (19-25%).

However, structure-conditioned models exhibit different systematic biases beyond taxonomic preferences. They show systematic preference for compact, thermostable proteins with high β-sheet content, which are characteristics overrepresented in crystallographic databases. We demonstrate active shifts in designs toward these properties regardless of the template's origin, showing that inherent biases directly shape functional outputs.

Interestingly, taxonomic preferences invert between model types, such as ProteinMPNN favouring archaeal proteins while ESM2 favours eukaryotic sequences. Our preliminary fine-tuning experiments suggest these architectural biases can be adjusted for specific design tasks without compromising performance.

Understanding these biases is critical for strategic model selection and for developing more effective and transparent protein engineering pipelines. Our findings highlight the importance of clarifying biases to enhance the reliability and versatility of AI-driven protein design tools.

References:

[1] Dauparas et al. Robust deep learning-based protein sequence design using ProteinMPNN. Science 378, 49-56 (2022)
[2] Ding & Steinhardt. Protein language models are biased by unequal sequence sampling across the tree of life arXiv: 2024.03.07.584001(2024)

(75) Drug-silicone oil interactions predictive model for biologics

Dimitra Škrekovski [1,2], Jure Cerar[2], Blaž Likozar[1]

[1] National Institute of Chemistry, Department of Catalysis and Chemical Reaction Engineering, Ljubljana, Slovenia
[2] Novartis Pharmaceutical Manufacturing LLC, Drug product development, Mengeš, Slovenia

dimitra.skrekovski@ki.si

Pre-filled syringes (PFSs), particularly in autoinjector formats, are widely used for biologics due to their convenience for self-administration. However, silicone oil (SO), employed as a lubricant in PFSs, can interact with therapeutic proteins, leading to **SO depletion** into the drug solution [1,2,3]. These **interactions at the SO-liquid interface** promote protein aggregation and detachment of SO particles, increasing subvisible particle counts and potentially compromising product specifications, efficacy, and immunogenic safety [1,3]. Such issues often remain undetected during early development due to limited testing, making late-stage mitigation complex and resource-intensive.

In this study, we utilized molecular dynamics (MD) simulations to explore the interactions of various monoclonal antibodies (mAbs) with SO at the SO – liquid interface. Advanced simulation techniques such as umbrella sampling were utilized to simulate mAbs' adsorption to the SO. These simulation methods allow us to calculate protein-SO interaction energies and correlate them with experimental data (e.g., microflow imaging, break-loose and gliding force measurements, and RapID) to validate our models.

Our goal is to develop a **predictive model** (Figure 1) that will warn us about possible SO-protein interactions in new drug candidates already in **early development**, allowing to adapt and shift development strategy, saving both time and resources.

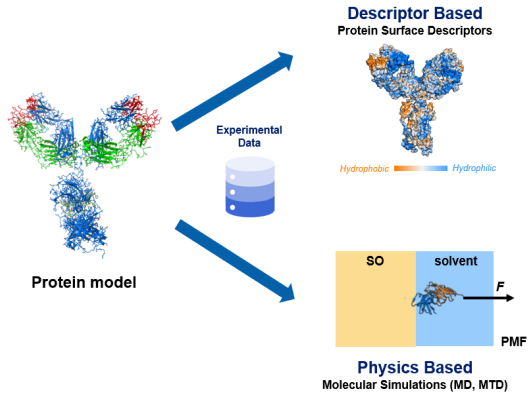


Figure 15: ***Schematic representation of predictive model creation workflow.***

Two different approaches will be tested); first, physics-based approach by studying the systems of interest with MD simulations. Secondly, once we gain a better understanding of the underlying mechanisms of protein-SO interactions, our model will be simplified by correlating protein surface properties responsible for the adsorption (e.g., hydrophobicity, charge distribution).

References:

[1] Moll, F.A.,et al. The silicone depletion in combination products induced by biologics. *European Journal of Pharmaceutics and Biopharmaceutics*, 203: 114418, 2024.
[2] Saurabh, S., et al. Adsorption of monoclonal antibody fragments at the water-oil. A coarse-grained molecular dynamics study. *APL Bioengineering*, 8: 026128, 2024.
[3] Dandekar, R., et al. Monoclonal Antibody Aggregation near Silicone Oil−Water Interfaces, *Langmuir*, 37: 1386-1398, 2021.

(76) Computational Design of a Bivalent Protein Inhibitor: Targeting Different Cytokine Receptors

Matej Milijaš-Jotić[1], Olga Klimenkova[2], Andrei Lupas[1], Julia Skokowa[2] and Mohammad ElGamacy[1,2]

[1] Department of Protein Evolution, Max Planck Institute for Biology, Tübingen, Germany
[2] Translational oncology, Internal Medicine II, University Hospital Tübingen, Tübingen, Germany

matej.jotic@tuebingen.mpg.de, mohammad.elgamacy@tuebingen.mpg.de

Targeting multiple clinically relevant receptors with a single therapeutic agent can improve treatment efficacy and address the limitations of monotherapy. In this work, we aimed to develop single-domain, bivalent blockers for IL-1R-1 and G-CSFR – two key cytokine receptors implicated in a variety of inflammatory and autoimmune disorders, as well as several cancers. We merged a peptide known for its inhibition of IL-1R-1 with a potent four-helical bundle inhibitor that represses G-CSF. This expanded the binding interface to IL-1R-1, which we further optimized using the Damietta protein design software to improve binding affinity in the bound form, as well as solubility and stability in the unbound state. Finally, we used molecular dynamics-based rankings to select a handful of candidates for experimental characterization. The best candidates are high-yielding, thermostable and bind each of the targets individually with nanomolar affinity. Moreover, they can inhibit both the G-CSF and IL-1 signaling pathways in cells-based assays and are capable of blocking either of the receptors in primary human hematopoietic stem and progenitor cells. This bivalent approach may help mitigate drug resistance, while providing enhanced avidity, potency and specificity compared to current therapies, which largely consist of chemotherapy or antibody monotherapy.

(77) VHH-MPNN: optimizing nanobody developability with inverse folding

Adrien Bitton[1], Anastasia Pentina[1], Sven H. Giese[1], Santiago Villalba[1]

[1] Bayer AG, Pharmaceutical Division, Müllerstr. 178. 13353 Berlin, Germany

adrien.bitton@bayer.com

Nanobodies (VHHs) are an attractive therapeutic modality, due to their compact single-chain format which enhances stability and eases delivery and manufacturing. VHH binders are usually discovered through camelid immunization, yet they must undergo optimization of many properties besides the targeted antigen to meet the requirements of therapeutic products.

Inverse folding models, tasked to generate diverse sequences constrained to fold according to a target backbone, have become essential protein design tools thanks to their ability to suggest sequences with improved developability traits and the structure-based functions [1]. Previous studies demonstrated their applicability at engineering antibody [2] and VHH [3] binding, here we introduce VHH-MPNN finetuned on millions of predicted structures spanning both sequencing and patent datasets, showing its effectiveness at VHH developability optimization.

We evaluate inverse folding models and VHH-MPNN on a hold-out dataset of 480 VHHs with measured developability profiles (e.g. stability, solubility, …), along with computational predictors to assess the in-silico success rate of the designed sequences. Finally, we discuss methods to bias and steer VHH-MPNN generation towards sequences with co-optimized developability properties and reduced immunogenicity.

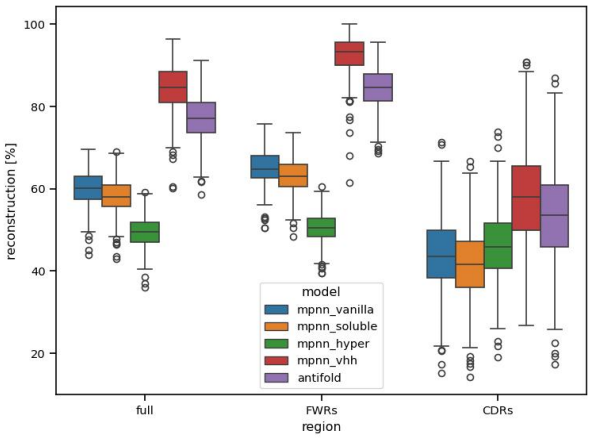


Figure 1: Native sequence recovery from inverse folding. *Models are first evaluated at reconstructing sequences from the hold-out 480 VHHs. MPNN "vanilla" and "soluble" are the original general protein checkpoints. MPNN "hyper" was finetuned on thermophilic sequences. Antifold was finetuned on antibodies and we use its heavy chain decoder. VHH-MPNN (ours) was specifically trained on nanobodies, consequently it achieves the highest sequence recovery.*

References:

[1] J. Dauparas et al. Robust deep learning-based protein sequence design using ProteinMPNN. *Science*, 2022

[2] M. H. Høie et al. AntiFold: Improved antibody structure-based design using inverse folding. *Bioinformatics Advances*, 2025.

[3] D. Chomicz et al. nanoFOLD: sequence design of nanobodies via inverse folding. *BioRxiv preprint*, 2025.

(78) Target Identification and Design of Cancer Antigen Binding Proteins for CAR T Cell Therapy in Malign Melanoma

Christina K Wuest[1,2], Max A Lingner Chango[1,3], Manfred Kunz[3], Clara T Schoeder[1,2,4]

[1] Leipzig University Medical Faculty, Institute for Drug Discovery, Leipzig, Germany
[2] Fraunhofer Institute for Cell Therapy and Immunology, Leipzig, Germany
[3] Universitätsklinikum Leipzig, Dermatology, Leipzig, Germany
[4] Center for Scalable Data Analytics and Artificial Intelligence, Dresden/Leipzig, Germany

christina.wuest@medizin.uni-leipzig.de

Chimeric antigen receptor (CAR) T cell therapy has transformed cancer treatment by enabling potent tumor cell elimination. However, off-tumor toxicity due to antigen expression in healthy tissues, along with tumor antigen heterogeneity and mutation, continue to limit therapeutic efficacy and drive resistance. To address these challenges, we design *de novo* binders against clinically relevant tumor surface protein derived from cancer patient data.

Targeting overexpressed antigens in cancer with *de novo* proteins has shown promising results [1,2]. Our approach combines state-of-the-art protein design tools: RFdiffusion and hallucination for backbone generation, $MPNN_{prot}$ for sequence optimization, and AlphaFold2 for *in silico* structural validation. Initially generated ErbB3 binders partially show strong and specific binding *in vitro*. This preliminary data will be utilized for pipeline optimization. In our future work we want to establish a rational target identification pipeline that integrates systematic *de novo* binder design.

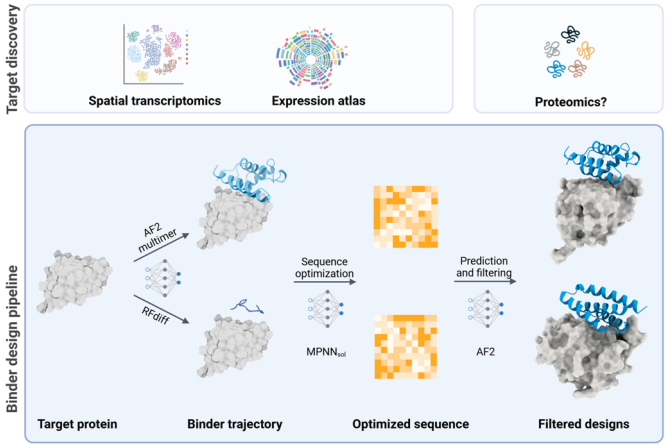


Figure 1: Schematic overview of the de novo protein design pipeline. *The projects aims to streamline target identification from cancer patient-derived data and generate viable antigen-binding domains for next-generation CAR T cell therapies.*

References:

[1] Cao, L., Coventry, B., Goreshnik, I. *et al.* Design of protein-binding proteins from the target structure alone. *Nature* **605**, 551–560 (2022). https://doi.org/10.1038/s41586-022-04654-9.

[2] Xia, Z., Jin, Q., Long, Z. *et al.* Targeting overexpressed antigens in glioblastoma via CAR T cells with computationally designed high-affinity protein binders. *Nat. Biomed. Eng* **8**, 1634–1650 (2024). https://doi.org/10.1038/s41551-024-01258-8.

(79) Deep learning-based design of nitrile hydratase enzymes for industrial applications and bioremediation

Joe Scaife[1], Rachel Noel[1], Binuraj Menon[2]

[1] Defense Science and Technology Laboratory, Human and Biological Advantage, Salisbury, United Kingdom
[2] University of Portsmouth, Centre for Enzyme Innovation, Portsmouth, United Kingdom

jscaife@dstl.gov.uk

Nitrile hydratases (NHases) are metalloenzymes which catalyse the conversion of nitrile groups into amides. NHases are important industrial enzymes which produce 600,000 tonnes of acrylamide a year and are also used for the bioremediation of nitrile pollutants from industrial waste (1). Despite their utility, NHases exhibit poor thermostability and low activity against specific nitrile substrates, limiting the scope of their use.

This project aims to engineer new NHases to overcome these limitations using deep learning-based protein design tools. Novel NHases will be developed through sequence redesign approaches, such as ProteinMPNN (2), and *de novo* design strategies to create entirely new enzyme scaffolds. High-throughput assays will be developed and validated to evaluate the performance of engineered enzymes. Computational analyses will then be conducted on top-performing variants to understand their functionality and guide further optimisation of the designs.

This work addresses the urgent need for sustainable chemical manufacturing processes and effective strategies for environmental cleanup, while also demonstrating the use of deep learning in complex enzyme engineering tasks.

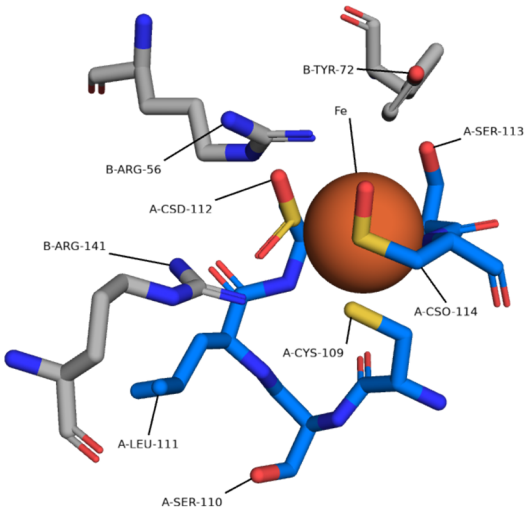


Figure 1. The active site structure of the Fe-type Nhase enzyme from *Rhodococcus erythropolis* (PDB:2ZPB). This diagram highlights **CSLCSC** metal binding motif with post translationally modified cysteines, which are required for enzyme activity.

References:

[1] Asano Y, Babich L, Ponzano BC, Díaz-Rodríguez A, Bertau M. Science of Synthesis: Biocatalysis in Organic Synthesis: Thieme; 2015.
[2] Dauparas J, Anishchenko I, Bennett N, Bai H, Ragotte RJ, Milles LF, et al. Robust deep learning–based protein sequence design using ProteinMPNN. Science. 2022;378(6615):49-56.

(80) Designing Filament-Specific MDA5 Binders

Till Gundlach[1], Florian Praetorius[1] & Carrie Bernecky[1]

[1] Institute of Science and Technology Austria (ISTA), 3400 Klosterneuburg, Austria

till.gundlach@ista.ac.at
florian.praetorius@ista.ac.at

MDA5 is a cellular sensor of viral double-stranded RNA and a critical component of the innate immune system in vertebrates [1]. The protein forms filaments around the RNA, triggering an interleukin response to virus infection [2]. While MDA5-dsRNA complexes have been resolved using cryo-EM [3], details into its structure and function remain lacking due to the dynamic nature of MDA5 filaments.

Here, we present work on a *de novo* protein to specifically recognize MDA5 filaments, but not its monomeric form. Our state-specific MDA5 binder could be used to detect or to stabilize MDA5 filaments. This could enable deeper insights into cellular filament dynamics and allow for further structural investigations using cryo-EM. MDA5 has been shown to play a role in autoimmune disorders [5] and is of interest as a target for cancer immunotherapy [4]. With better understanding of MDA5 functionality our state-specific binders could also be used to control its cellular behavior in the context of therapeutics.

Our state-specific MDA5 binder consists of two binders against different domains of MDA5, which are rigidly linked to fit the spatial arrangement of monomers in filaments. Single-target binder designs are created using RFDiffusion [6] or BindCraft [7] pipelines. We experimentally verify binding to single domains of MDA5 using size-exclusion chromatography. We then link compatible designs using RFDiffusion and mutationally tune binder affinity as needed.

References:

[1] Reikine, Stephanie, Jennifer B. Nguyen, and Yorgo Modis. 2014. "Pattern Recognition and Signaling Mechanisms of RIG-I and MDA5." *Frontiers in Immunology* 5 (July). https://doi.org/10.3389/fimmu.2014.00342.

[2] Berke, Ian C, and Yorgo Modis. 2012. "MDA5 Cooperatively Forms Dimers and ATP-sensitive Filaments upon Binding Double-stranded RNA." *The EMBO Journal* 31 (7): 1714–26. https://doi.org/10.1038/emboj.2012.19.

[3] Yu, Qin, Alba Herrero del Valle, Rahul Singh, and Yorgo Modis. 2021. "MDA5 Disease Variant M854K Prevents ATP-Dependent Structural Discrimination of Viral and Cellular RNA." *Nature Communications* 12 (1): 6668. https://doi.org/10.1038/s41467-021-27062-5.

[4] Rice, Gillian I., Yoandris del Toro Duany, Emma M. Jenkinson, et al. 2014. "Gain-of-Function Mutations in IFIH1 Cause a Spectrum of Human Disease Phenotypes Associated with Upregulated Type I Interferon Signaling." *Nature Genetics* 46 (5): 503–9. https://doi.org/10.1038/ng.2933.

[5] Duewell, Peter, Ebba Beller, Sabrina V Kirchleitner, et al. 2015. "Targeted Activation of Melanoma Differentiation-Associated Protein 5 (MDA5) for Immunotherapy of Pancreatic Carcinoma." *OncoImmunology* 4 (10): e1029698. https://doi.org/10.1080/2162402X.2015.1029698.

[6] Watson, Joseph L., David Juergens, Nathaniel R. Bennett, et al. 2023. "De Novo Design of Protein Structure and Function with RFdiffusion." *Nature* 620 (7976): 1089–100. https://doi.org/10.1038/s41586-023-06415-8.

[7] Pacesa, Martin, Lennart Nickel, Joseph Schmidt, et al. 2024. "BindCraft: One-Shot Design of Functional Protein Binders." Preprint, bioRxiv, October 1. https://doi.org/10.1101/2024.09.30.615802.

(81) Directed Evolution Of Novel Dna Binding Domains

William Cheney[1], Carl de Boer[1]

[1] School of Biomedical Engineering, Department of Applied Science and Medicine, Vancouver, British Columbia, Canada

wcheney@student.ubc.ca

DNA binding domains (DBD) are found in diverse types of proteins, including transcription factors, restriction enzymes, and transposases. These proteins are central to many biological processes and are frequently adapted for scientific use, ranging from standard cloning to next-generation sequencing. Despite widespread use, the engineering of custom DNA-binding proteins remains a challenge due to the complex relationship between protein structure and binding specificity. Understanding the space of possible DNA-binding proteins would provide additional insight into both their engineering and evolutionary origins. To explore this space, we aim to evolve DNA-binding proteins starting from random protein sequences. Despite the vast sequence space of possible proteins, DBDs may be surprisingly common. DNABIND, a SVM model that predicts DNA binding proteins, predicts that >1% of random sequences contain DNA binding properties. We will accomplish this through ribosome display, a high-throughput in vitro-directed evolution technique. Beginning with a random protein library, we will selectively pull-down proteins that interact with DNA, mutagenize the binders, and repeat this process multiple times to enrich proteins with DNA-binding properties. These results will shed light on the space of all DBDs and can provide powerful training data for generative protein models. Together, this information will enable the engineering of novel DBDs for specific sequences.

(82) BINDMED: machine-learning-based *de novo* design and lead optimization of high affinity BINDers for biomarkers of bioMEDical relevance

Marco Orlando[1], Ario de Marco[1]

[1] Lab of Environmental and Life Sciences, University of Nova Gorica, Vipavska cesta 13, Rožna Dolina, 5000 Nova Gorica, Slovenia

marco.orlando@ung.si

I am going to present the project I will develop during my **PostDoc** at the **University of Nova Gorica**, funded under the **SMASH MSCA COFUND action** [1]

Machine Learning (ML) methods represent the most promising frontier for *de novo* protein design of binders selective for specific antigen epitopes, such as antibodies [2]. Indeed, such methods may represent one of the few available options when working with poorly immunogenic target molecules, unfolded epitopes, highly conserved antigens, and when the binders belong to advantageous but poorly represented proteins in public databases, such as nanobodies and adhirons [3]. Despite the great accelaration that *de novo* desing tools is playing in this context in the last couple of years [4-6], >> **10 designs have to be characterized to obtain binders with low nM affinity** , while there is **still not all-in-one** solution **for combining** the best in terms of **binding affinity**, **foldability**, **structural stability** and **immugenicity**.

The **first part** of this project will investigate customized **approaches to generate and score ML-based *de novo* (re)designed Nbs** and **adhiron binders, starting from alternative target protein scaffolds** (**folded** and **unfolded**); the **second part** is set to develop a method to **condition the generation process** towards **biophysical features** (binding affinity, flexibility/stability, solubility, immunogenicity) and select < 10 designs **for experimental characterization** in the host laboratory, targeting both a folded and an unfolded epitope; **validated designs** will be used in a **third step** to develop a **few-shot approach** to **steer** a p**re-traied ML-model** for **generating binders** for the **specific target** and **optimized** to reach the **pareto frontier across multiple target properties,** while **having pM affinity** and **minimizing further experiments**.

Altogether, **this project** is expected to provide an ***in-silico* pipeline** for **the generation** and **data-driven lead optimization** of **Nb** and **other custom protein binders** to **target non-canonical biomarkers** for which **high-throughput binding assays** are **not possible** and **classic screening methods are prone to fail.**

References:

[1] https://smash.ung.si/

[2] Akbar, R., Bashour, H., Rawat, P., Robert, P. A., Smorodina, E., Cotet, T. S., ... & Greiff, V.. Progress and challenges for the machine learning-based design of fit-for-purpose monoclonal antibodies. *MAbs* 14(1):2008790. 2022

[3] D'Ercole, C., De March, M., Veggiani, G., Oloketuyi, S., Svigelj, R., & De Marco, A. Biological applications of synthetic binders isolated from a conceptually new adhiron library. *Biomolecules*, 13(10), 1533. 2023

[4] Wang, R., Wu, F., Gao, X., Wu, J., Zhao, P., & Yao, J. Iggm: A generative model for functional antibody and nanobody design. *bioRxiv*. 2025

[5] Bennett, N. R., Watson, J. L., Ragotte, R. J., Borst, A. J., See, D. L., Weidle, C., ... & Baker, D. Atomically accurate de novo design of single-domain antibodies. *bioRxiv*. 2024

[6] Liu, C., Wu, K., Choi, H., Han, H. L., Zhang, X., Watson, J. L., ... & Baker, D. Diffusing

(83) Predicting Experimental Success in De Novo Binder Design: A Meta-Analysis of 3,766 Experimentally Characterised Binders

Andreas Rygaard[1,2,*], Max D. Overath[1,*,†], Christian P. Jacobsen[1]
Valentas Brasas[1], Oliver Morell[1], Pietro Sormanni[2], Timothy P. Jenkins[1,*]

[1] Department of Biotechnology and Biomedicine, Technical University of Denmark, Kongens Lyngby, Denmark
[2] Yusuf Hamied Department of Chemistry, University of Cambridge, Cambridge CB2 1EW, UK
*Correspondence: maxove@dtu.dk, anryg@dtu.dk, tpaje@dtu.dk † Equal contribution

Designing **high-affinity de novo protein binders** has become increasingly tractable, yet in vitro prioritisation continues to depend on heuristics in the absence of systematic analysis. Here, we present a large-scale meta-analysis of **3,766 experimentally tested binders** across **15 structurally diverse targets**. Using a unified, high-throughput pipeline that re-predicts each binder–target complex with AF2 (initial guess and ColabFold), AF3, and Boltz-1, we extract over **200 structural, energetic, and confidence features** per design. We show that interface-focused metrics, most notably the AF3-derived interaction prediction Score from Aligned Errors (**ipSAE**), outperform commonly used scores such as ipAE and ipTM, with a 1.4-fold increase in average precision compared to ipAE.

We further show that combining these metrics with orthogonal physicochemical interface descriptors, including Rosetta ΔG/ΔSASA and interface shape complementarity, **improves predictive performanc**e, especially when used in **pairwise interaction terms.** While overall performance varies by target, simple linear models trained on a small number of AF3-derived features generalize well across datasets.

We propose interpretable, **target-agnostic filtering strategies**, such as combining AF3 ipSAE_min rankings with structural filters, to improve precision in selecting binders for testing. Finally, we **release the complete dataset,** establishing a **community resource to benchmark** and accelerate de novo binder discovery.

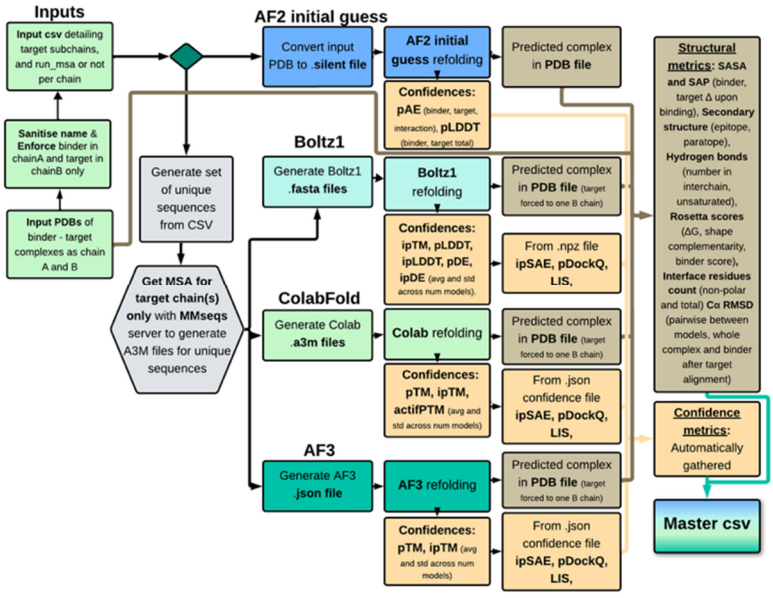


Figure 1: A unified high-throughput pipeline for systematic binder evaluation. *To improve prioritisation in* de novo *binder design, we developed a scalable pipeline that refolds binder–target complexes using AF2, Boltz-1, ColabFold, and AF3. Standardised MSA generation and model-specific formatting enable consistent input handling. From the predictions, over 200 structural and confidence-based features are extracted, including ipSAE, pDockQ2, and Rosetta-derived metrics. The pipeline supports large-scale, interpretable assessment of binder quality across*

(84) DAPPER: A Context-Aware Deep Learning Framework for continuous Peptide–Protein Binding Affinity Prediction

Nitesh Mani Tripathi, and Anupam Bandyopadhyay*

Biomimetic Peptide Engineering Laboratory, Department of Chemistry, Indian Institute of Technology, Ropar, Punjab-140001, India.
*Correspondence: AB, anupamba@iitrpr.ac.in

The PD-1/PD-L1 immune checkpoint pathway plays a central role in tumor immune evasion by suppressing cytotoxic T-cell activity within the tumor microenvironment. Therapeutic blockade of PD-L1 protein with emerging peptide inhibitors has revolutionized cancer immunotherapy by reactivating antitumor T-cell responses. Peptide-based inhibitors targeting PD-L1 protein offer promising potential for enhanced specificity and improved pharmacokinetic properties, advancing precision immune checkpoint modulation in oncology. Here, we present DAPPER, a lightweight, modular deep neural network regressor model designed for continuous peptide-protein binding affinity prediction. The model integrates peptide sequences, protein binding region embeddings, and peptide-level physicochemical features through a hybrid sequence-context fusion paradigm. It employs a bidirectional LSTM (BiLSTM) architecture for peptide encoding, combined with pre-trained ESM-2 embeddings from functional regions of the target protein and domain-specific biophysical features, to produce continuous affinity scores (on a 0-1 scale) for peptide candidate ranking. The model is trained on a dataset comprising experimental IC50 values and augmented via label-aware strategies to enhance generalization, with hyperparameters optimized via Bayesian search [1]. This approach offers an efficient in silico screening of millions of de novo peptides alternative to labor-intensive experimental screening, demonstrating strong performance in prioritizing high-affinity beta-hairpin peptides against PD-L1 protein.

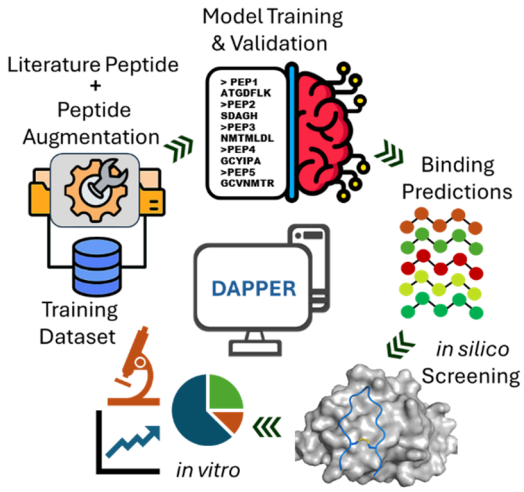


Figure 1: *Overview of the DAPPER workflow for peptide–protein binding affinity prediction, integrating peptide sequences, physicochemical features, and protein binding region embeddings to rank candidate peptides for structural modeling and validation.*

References:

[1] DAPPER: A Context-Aware Deep Learning Framework for Continuous Peptide–Protein Binding Affinity Prediction. **Nitesh Mani Tripathi**, and Anupam Bandyopadhyay *(Manuscript under preparation).*

(85) Design of Metal-Binding Peptides from Protein Fragments

Ján Michael Kormaník[1], Erik Andris[1], Tadeáš Kalvoda[1], Robin Kryštůfek[1], Milan Kožíšek[1], Lubomír Rulíšek[1]

[1] Institute of Organic Chemistry and Biochemistry of the Czech Academy of Sciences, Prague, Czech Republic

kormanik@uochb.cas.cz

Small metallopeptides offer a big potential for various applications, including medicine, biosensors or catalysis, but this territory is still under explored. Unfortunately, ML-based approaches are not developed to reliably design short metal binding peptides. In our recent study, we have designed a small peptide by taking fragments from proteins with known structure and joining them together. This peptide, which harbors the tetrahedral Cys_2His_2 motif commonly appearing in zinc fingers, binds Zn^{2+} ions with K_d = 220 nM.[1] In the present study, we aim to expand the scope to improve this result by including Asp and Glu as possible binding amino acids and also employing a more thorough computational protocol to select the best designs. Using a fluorescene-assay based screening, we obtained several good candidates that show strong metal-binding properties.

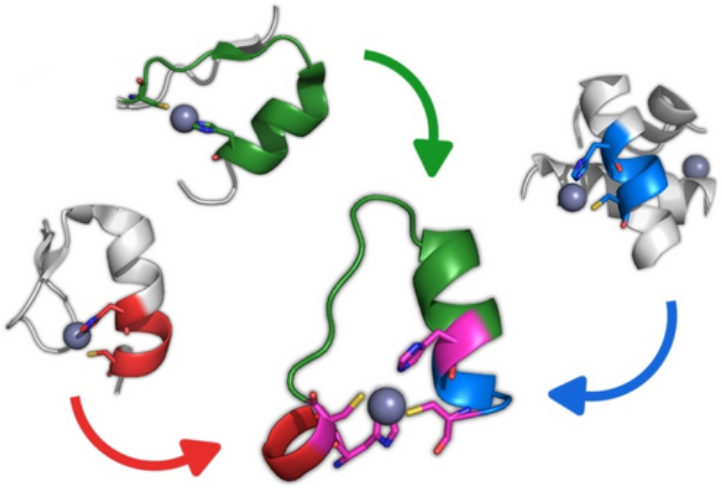


Figure 1: Design process. *Peptides are designed by taking pieces from proteins with known structure and joining them together, to obtain a tetrahedral metal-binding site. Best candidates are experimentally verified.*

References:

[1] Kormaník, J. M., Herman, D., Andris, E., Culka, M., Gutten, O., Kožíšek, M., Bednárová, L., Srb, P., Veverka, V., Rulíšek, L. (2025). Design of Zn-Binding Peptide (s) from Protein Fragments. *ChemBioChem*, *26*(7), e202401014.

(86) The novel rotamer library NuConf enables the design of protein-nucleic acid interactions in the protein design software MUMBO

Marharyta O. Makarova[1], Martin T. Stiebritz[1], Birthe Lemke[1], Beatrix Suess[2,3] and Yves A. Muller[1]

[1] Division of Biotechnology, Department of Biology, Faculty of Sciences, Friedrich-Alexander Universität Erlangen-Nürnberg, Germany
[2] Department of Biology, Technische Universität Darmstadt, Germany
[3] Centre for Synthetic Biology, Technische Universität Darmstadt, Germany

marharyta.makarova@fau.de

Current AI-tools for designing macromolecules have been facing limitations with structure prediction and design of nucleic acids, as the amount of experimental data available for training is significantly lower than that of proteins [1]. Consequently, developing alternative approaches remains of interest. MUMBO is a program that allows designing protein-protein interactions and protein-ligand binding pockets by means of side-chain packing algorithms, where alternative amino acids and their rotamers are generated from libraries on a fixed backbone and the most favourable ones are selected based on the lowest overall energy [2]. In order to extend the program's capabilities to designing nucleic acids, we developed NuConf – a discrete pseudorotational angle-dependent nucleotide-specific rotamer library. NuConf was derived by means of statistically analyzing pseudorotational and dihedral angles of more than 175,000 nucleotides from experimental structures and validated by rebuilding more than 20,000 nucleotides in a custom dataset. Notably, it became evident that nucleotides are predicted at least as accurate as amino acids. We show that the implementation of the NuConf library in MUMBO enables modelling and designing DNA and RNA sequences on a fixed backbone together with protein-nucleic acid interaction interfaces. Here, we also provide examples of its application for specific research questions. Because its approach and format are program agnostic, NuConf can be used by other molecular design frameworks as well.

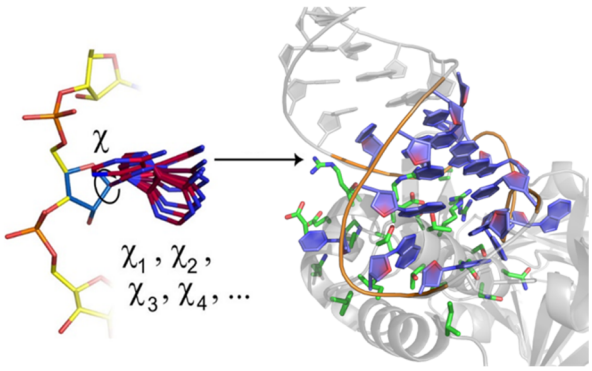


Figure 16: Utilizing nucleotide rotamer library allows for designing protein-RNA interface. *A schematic representation of nucleotide rotamer modelling in MUMBO is shown to the left. An interaction interface modelled with MUMBO between DNA-binding domain of NagR and the RNA aptamer is shown to the right, with protein side chains in green and nucleotides in blue.*

References:

[1] B. Schneider, B.A. Sweeney, A. Bateman, J. Cerny, T. Zok and M. Szachniuk. When will RNA get its AlphaFold moment? *Nucleic Acid Res*, 51:9522-9532, 2023

[2] M.T. Stiebritz and Y.A. Muller. MUMBO: a protein-design approach to crystallographic model building and refinement. *Acta Crystallogr D Biol Crystallogr*, 62:648-658, 2006.

(87) *De novo* design and functionalization of miniproteins for aldol reaction catalysis

Anna Szczepańska[1], Manish Kumar[1], Łukasz Berlicki[1]

[1] Wroclaw University of Science and Technology, Department of Bioorganic Chemistry, Wrocław, Poland

anna.szczepanska@pwr.edu.pl

The high conformational stability and compact size of miniproteins make them attractive scaffolds for *de novo* design of functional peptides, which can be readily synthesized using solid-phase peptide synthesis (SPPS). These characteristics make them particularly well-suited for rational enzyme design, allowing for development of novel, highly stable catalysts and deepening our understanding of structure- activity relationships.

In this study, we present the *de novo* design and characterization of miniproteins catalysing the aldol reaction between phenylacetaldehyde and 4-nitrobenzaldehyde under mild aqueous conditions. Variants with different charge distributions have been characterized and evaluated.

The miniproteins were designed using Rosetta, utilizing the BluePrint and FastDesign protocol, ProteinMPNN, and further fine-tuned using Alphafold3.[1][2][3] To promote folding and enhance rigidity, we incorporated a cyclic β-amino acid *trans*-ACPC into the helical secondary structure elements.[4] The resulting miniproteins were synthesized via SPPS and characterized by circular dichroism and nanoDSF, confirming high thermal stability and well-folded structures.

Catalytic activity was assessed by HPLC following 1 hour of reaction in phosphate buffer (pH 7.4). Additional variants were created by dividing the structure into two parts along the assumed reaction axis. Selected amino acid residues on both sides were replaced to adjust the dipole moment of the molecule.

Here we report the results for a series of miniproteins with HEEEH (helix-β-sheet-helix) topology. Ongoing studies aim to further optimize the activity and specificity. Moreover, the obtained miniproteins demonstrate high thermal stability, making them re-usable in the future as scaffolds for different applications.

References:

[1] B. Kuhlman, G. Dantas, G. C. Ireton, G. Varani, B. L. Stoddard, and D. Baker. Design of a novel globular protein fold with atomic-level accuracy. *Science*, 302(5649):1364–1368, 2003.
[2] J. Dauparas, I. Anishchenko, N. Bennett, et al. Robust deep learning-based protein sequence design using ProteinMPNN. *Science*, 378(6615):49–56, 2022.
[3] J. Jumper, R. Evans, A. Pritzel, et al. Highly accurate protein structure prediction with AlphaFold. *Nature*, 596:583–589, 2021.
[4] L. M. Johnson and S. H. Gellman. Methods for studying protein folding and stability. *Methods in Enzymology*, 523:407–429, 2013.

(88) Inhibitor design for bacterial nanomachines

K.S. Stroh[1], A. Hayda[2], F. Parmeggiani[3], E. Durand[2], R. Pellarin[1]

[1] École Normale Supérieure de Lyon, Laboratory of Biology and Modeling of the Cell, Lyon, France
[2] Laboratoire de Chimie Bactérienne, Marseille, France
[3] School of Pharmacy and Pharmaceutical Sciences, Cardiff University, Cardiff, UK

kai.stroh@ens-lyon.fr

The Type VI Secretion System (T6SS) is a contractile nanomachine employed by Gram-negative bacteria to inject toxic effectors into competing microbes and host cells [1,2]. Disrupting T6SS assembly offers a promising anti-virulence strategy to disarm antibiotic-resistant pathogens without driving strong selective pressure for resistance. Here, we report a computational pipeline for de novo design of inhibitors to specifically block the anchoring interaction between the membrane complex and the baseplate in Bacteroides fragilis.

We leveraged RFdiffusion [3] to generate initial backbone scaffolds guided by structural templates of the protein TssN, a key assembly hub in the B. fragilis T6SS. Candidate scaffolds were sequence-designed using ProteinMPNN [4]. Designs were filtered by AlphaFold2 initial guess confidence metrics [5]. Top candidates were further evaluated via Rosetta-derived physics-based observables—including interface ΔΔG, buried surface area, and hydrogen-bonding networks [7, 8]. Finally, we applied Chai-1 [6], as an orthogonal structure prediction network, to score designs without prior information about the binding pose.

References

[1] E.R. Green and J. Mecsas, Microbiol. Spectr. 4 (2016).
[2] M. Yekani et al., Microb. Pathog. 149, 104506 (2020).
[3] J. L. Watson et al., Nature. 620, 1089–1100 (2023).
[4] J. Dauparas et al., Science. 378, 49–56 (2022).
[5] N. R. Bennett et al., Nat. Commun. 14, (2023).
[6] J. Boitreaud et al., bioRxiv (2024).
[7] S. Chaudhury et al., Bioinformatics, 26, 689–691 (2010).
[8] M. D. Tyka et al., J. Mol. Biol., 405, 607–618 (2011).

(89) Conformationally selective *de novo* protein binders for phosphate monitoring

Wouter M.A.J. Langers[1,3], Peter Zijlstra[1,3], Maarten Merkx[2,3]

[1] Department of Applied Physics and Science Education, Molecular Biosensing, Eindhoven University of Technology, The Netherlands.
[2] Laboratory of Chemical Biology, Department of Biomedical Engineering, Protein Engineering, Eindhoven University of Technology, The Netherlands.
[3] Institute for Complex Molecular Systems, Eindhoven University of Technology, The Netherlands.

W.M.A.J.Langers@tue.nl

High phosphate levels contribute to eutrophication, posing a significant threat to aquatic ecosystems. Closely monitoring the phosphate concentration in bodies of water is essential to mitigate these environmental impacts. However, current detection methods often lack the sensitivity, specificity, and continuity for robust environmental monitoring. This research leverages computational protein design to develop a phosphate sensing platform based on phosphate binding protein (PBP) and **conformationally selective**, **reversible** *de novo* binders.

We exploit the natural affinity (KD ~0.1 µM) and selectivity of PBPs, which undergo a distinct conformational change upon phosphate binding. This conformational change provides the structural basis for designing *de novo* binders that selectively recognize the phosphate-bound form of the PBP, enabling indirect detection of phosphate through PBP-binder interactions (Figure 1). AlphaFold2 structure predictions of PBP consistently yield the ligand-bound conformation. Therefore, by using the AlphaFold2-based design pipeline **BindCraft**, we could generate candidate binders predicted to interact specificially with the phosphate-bound conformation of PBP [2].

Preliminary binding assays (SPR) confirm the interaction between the designed binders and PBP in the presence of phosphate. Ongoing work aims to validate the conformation-specific nature of these interactions. Moreover, by characterizing the binding kinetics we are able to select for reversibility, enabling continous phopshate monitoring. Ultimately, this biosensing platform could be integrated into existing biosensor setups using single-molecule plasmon-enhanced fluorescence, enabling more effective phosphate monitoring [1].

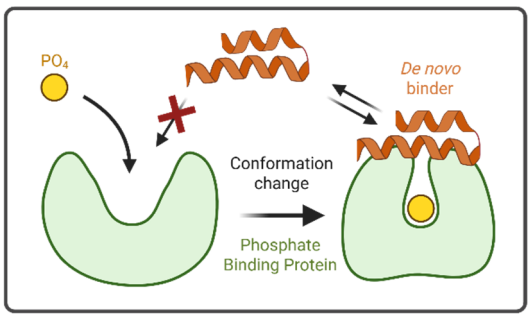


Figure 1: Schematic of conformational specific binder. *Phosphate Binding Proteins (green) undergo a distinct conformational change upon phosphate binding (yellow ball). A de novo designed binder (orange) selectively interacts with the phosphate-bound conformation of PBP.*

References:

[1] Lamberti, V., Dolci, M., & Zijlstra, P. (2024). Continuous monitoring biosensing mediated by Single-Molecule Plasmon-Enhanced fluorescence in complex matrices. ACS Nano. https://doi.org/10.1021/acsnano.3c12428

[2] Pacesa, M. et al. BindCraft: one-shot design of functional protein binders. bioRxiv, 2024.09.30.615802

(90) OVO, an Open-Source Ecosystem for De-novo Design

David Prihoda[1], Marco Ancona[1], Tereza Calounova[1], Adam Kral[1], Danny A. Bitton[1]

[1] Applied Research and Innovation, MSD Czech Republic, Prague, Czech Republic.

david.prihoda@msd.com

Recent advances in computational biology have transformed protein design, enabling accurate structure prediction, sequence generation, and de-novo design of novel proteins with broad biomedical applications. These breakthroughs have spawned a diverse array of specialized models and tools that are often chained into complex pipelines. However, the current landscape is fragmented: tools are distributed as standalone scripts or notebooks, lack standardized installation and require significant computational expertise to integrate into end-to-end pipelines and operate at scale. This fragmentation impedes reproducibility, scalability, and rapid adoption of new models. Moreover, existing pipelines rarely provide intuitive user interfaces, file management, or interactive visualization capabilities. Selecting promising candidates from thousands of designs demands multi-criteria filtering beyond confidence scores, incorporating sequence, structure, and physics-based features, which is rarely supported in a unified platform.

Here, we introduce OVO, the first end-to-end de-novo protein design ecosystem tailored to address these challenges. OVO features a modular, scalable backend with pre-defined workflows for design and evaluation that can be extended by the community via third party plugins. Alongside standard designability metrics, OVO features ProteinQC, a novel method for protein sequence and structure evaluation, and Promb, a novel method for protein humanness evaluation. Coupled with an interactive visualization interface, this empowers users to explore and select designs efficiently. By consolidating models and workflows into a cohesive open-source platform, OVO demystifies the design process, facilitates reproducibility, and fosters community-driven benchmarking and standards development. This ecosystem not only accelerates protein design at industrial scale but also provides a foundation for continuous methodological innovation in this rapidly evolving field.

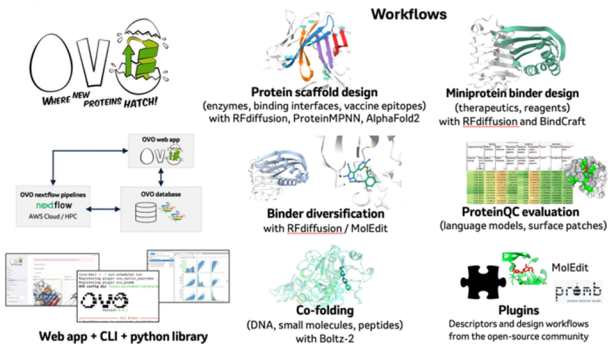


Figure 1: OVO ecosystem for de novo design. OVO enables executing end-to-end de novo design and evaluation pipelines on cloud or HPC infrastructure. Pre-defined workflows include protein scaffold design with RFdiffusion [1], miniprotein binder design with RFdiffusion or Bindcraft [2], small molecule binder diversification with MolEdit [3], and more.

References:

[1] Watson, J.L., Juergens, D., Bennett, N.R. *et al.* De novo design of protein structure and function with RFdiffusion. *Nature* **620**, 1089–1100 (2023).

[2] Pacesa, M., Nickel, L., Schellhaas, C., BindCraft: one-shot design of functional protein binders. *BioRxiv* (2025).

[3] Lin, X., Xia, Y., Li, Y. et al. In-silico 3D molecular editing through physics-informed and preference-aligned generative foundation models. Nat Commun 16, 6043 (2025)

(91) The Protein Engineering Tournament: A Community Benchmark For Predictive And Generative Models In Protein Engineering

Bijoy Desai[1]

[1]The Align Foundation

bijoy@alignbio.org

The Align Foundation is a nonprofit organization that works to propel biology into a new era of reproducible, scalable, and shareable research. Our vision is to accelerate the field by development of powerful predictive models. To get there, we have developed a suite of programs designed to enable collaborative, large, data intensive projects supported by automated methods. As a part of our community benchmarking initiative we host the Protein Engineering Tournament that is designed to benchmark predictive and generative models in protein engineering, offering a platform for critical assessment and collaborative learning-similar to what Critical Assessment of Structural Prediction (CASP) has achieved for protein structure prediction.

This year's Tournament will focus on the enzyme target polyethylene terephthalate hydrolase (PETase), with real world impact in eliminating microplastics, and development of a circular plastic economy. We are in the late stage of planning the Tournament, which will take place in two consecutive phases: 1) Predictive, where competitors will be challenged to predict withheld protein function data, and 2) Generative, where competitors will be challenged to design novel enzymes with improved properties. As a part of Align Foundation's mission to make data for AI open, reproducible, and shareable, we will generate and release a training dataset around PETase sequences that the participants can use for supervised learning in the predictive and generative phases. In the generative phase we will obtain novel PETase sequences designed by the different computational modeling methods employed by the participants and generate experimental data for them to assess the ability of each model to improve multiple properties. The tournament is scheduled to begin in winterof 2025 and culminate in summer of 2026.

(92) Distilling Protein Interfaces with Design to Scale Experimental Mapping of Protein-Protein Interactions

Lisa S. Brandenburg[1], **Basile I. M. Wicky**[1]

[1] ETH Zürich, Department of Biosystems Science and Engineering, Basel, Switzerland
lisa.brandenburg@bsse.ethz.ch

Understanding protein-protein interactions (PPIs) is central to unraveling the complexity of cellular networks and advancing therapeutic targeting of key regulatory nodes. While the Protein Databank has enabled the rise of deep learning-based structure prediction models that revolutionized our ability to infer protein structures from their sequences, accurately predicting PPI affinities and specificities remains constrained by the absence of large, quantitative experimental datasets [1,2]. To address this gap, we are developing a framework that integrates protein design and high-throughput experimental screening to create a distilled dataset of protein interfaces and their interactions. The approach involves (i) deriving a non-redundant set of structural interface motifs from natural PPIs [3], (ii) embedding these motifs into compact de novo protein scaffolds using generative design [4], and (iii) screening this distilled dataset with a multiplexed bacterial two-hybrid assay using next-generation sequencing for quantifiying interactions [5]. This framework aims to enable scalable mapping of millions of protein-protein interactions. We have largely established the motif scaffolding pipeline and initiated preliminary interaction screens, providing the first indications of feasibility for this approach. I will present the rationale, design strategy, and ongoing progress toward building this distilled "interfaceome," which has the potential to improve predictive models of PPI specificity and support applications in therapeutic protein engineering and synthetic biology.

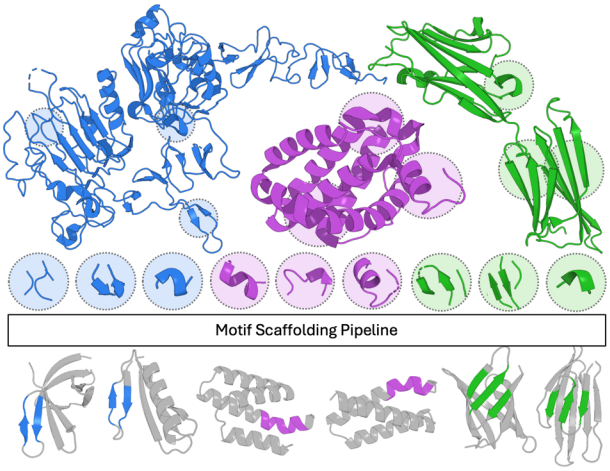


Figure 1: Computational motif scaffolding framework.

Motif extraction from three structurally distinct model proteins (HER2 in blue, MCL1 in purple, PD-L1 in green) and the subsequent scaffolding of their surfaces into small de novo proteins (de novo structure shown in gray) for experimental screening.

References:

[1] Lee, C. Y. et al. Systematic discovery of protein interaction interfaces using AlphaFold and experimental validation. Mol. Syst. Biol. 20, 75–97 (2024).
[2] Barrio-Hernandez, I. et al. Clustering predicted structures at the scale of the known protein-universe. Nature 622, 637–645 (2023).
[3] Mackenzie, C. O., Zhou, J. & Grigoryan, G. Tertiary alphabet for the observable protein structural universe. Proc. Natl. Acad. Sci. 113, E7438–E7447 (2016).
[4] Watson, J. L. et al. De novo design of protein structure and function with RFdiffusion. Nature 620, 1089–1100 (2023).
[5] Boldridge, W. C. et al. A multiplexed bacterial two-hybrid for rapid characterization of protein–protein interactions and iterative protein design. Nat. Commun. 14, 4636 (2023).

(93) Benchmarking state-of-the-art *In silico* peptide design, docking, and optimization methods in a practical setting

Mark Fonteyne[1,2], Nada Badr[1,2], Peter J.K. Kuppen[1], Olivier J.M. Béquignon[2], Lysanne D. A. N. de Muynck[1], Alexander L. Vahrmeijer[1], Gerard J. P. van Westen[2*], Willem Jespers[3*]

[1] Department of Surgery, Leiden University Medical Center, 2333 ZA Leiden, The Netherlands.
[2] Medicinal Chemistry, Leiden University (LACDR), 2333 CC Leiden, The Netherlands.
[3] Department of Pharmacy, University of Groningen, 9713 AV Groningen, The Netherlands.
* These authors contributed equally as last authors.

Recently, an increasing amount of peptides have emerged as promising therapeutics owing to their target specificity, strong binding affinity, and cost-effective production.[1-3] While deep learning-based structure prediction tools has greatly enhanced structural predictions for protein–peptide complexes, it has simultaneously marginalized physics-based interaction models, favoring structural coherence over biophysically grounded accuracy. Hence, it is essential to assess the strength and weaknesses of structure-based and physics-based approaches to facilitate the design of novel peptides in a practical setting. In this benchmark study, the molecular docking software AutoDock-CrankPep and HADDOCK were evaluated, together with structure-based protocols such as AF2-multimer, AF3, and Boltz-2 based on three main categories: target binding site identification, peptide virtual screening, and peptide optimization. To simulate a practical setting, the non-redundant PepPro database was used to evaluate the methods on their ability to account for protein flexibility.[4] The PepPro database contains 57 experimentally obtained apo- and holo-protein structures with peptide sequence lengths ranging from 5-30 amino acids. In this study, co-folding techniques performed best at target binding site identification and peptide virtual screening. Boltz-2 outperformed every method with a high-quality, one-hit pose prediction success rate of 68.4% (figure 1). However, Boltz-2 may exceed beyond its true predictive capability in more difficult cases, as it reuses structural information from the native PDB in the MSA to greater extend then AF3, rather than relying on independent evolutionary data. Peptide pLDDT consistently exhibited a stronger correlation with the DockQ score than other metrics across all co-folding methods, indicating that peptide pLDDT is a more reliable predictor of binding quality than the conventional use of iPTM rankings. This benchmark provides more insights on the strengths and weaknesses of the state-of-the-art peptide docking and optimization methods and highlights how to use them in a practical setting.

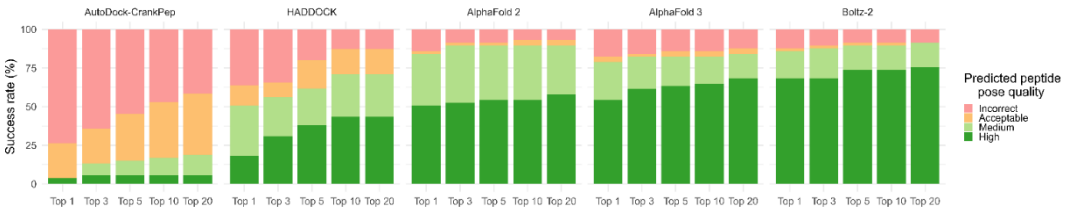


Figure 1: Success rate of pose prediction quality for Protein-peptide docking methods. *Success rate (in %) of the Protein-peptide docking methods to achieve a predicted peptide pose quality based on the DockQ score.*

References

[1] L. Wang *et al.*, "Therapeutic peptides: current applications and future directions," *Signal Transduct Target Ther*, vol. 7, no. 1, p. 48, Feb 14 2022, doi: 10.1038/s41392-022-00904-4.
[2] P. Vlieghe, V. Lisowski, J. Martinez, and M. Khrestchatisky, "Synthetic therapeutic peptides: science and market," *Drug Discov Today*, vol. 15, no. 1-2, pp. 40-56, Jan 2010, doi: 10.1016/j.drudis.2009.10.009.
[3] O. Al Musaimi, D. AlShaer, B. G. de la Torre, and F. Albericio, "2024 FDA TIDES (Peptides and Oligonucleotides) Harvest," *Pharmaceuticals (Basel)*, vol. 18, no. 3, Feb 20 2025, doi: 10.3390/ph18030291.
[4] X. Xu and X. Zou, "PepPro: A Nonredundant Structure Data Set for Benchmarking Peptide-Protein Computational Docking," (in eng), *J Comput Chem*, vol. 41, no. 4, pp. 362-369, Feb 5 2020, doi: 10.1002/jcc.26114.

(94) Machine learning guided interaction analysis

Dr. Reuben Reed[1], Dr. Thomas Kavanagh[1], Dr. James Wilkinson[1]

[1] Fluidic Sciences Ltd, UK

reuben.reed@fluidic.com

Accurate characterization of molecular interactions is fundamental to understanding biological systems and developing therapeutic strategies. Conventional surface-based methods, however, often introduce immobilization artifacts, distorting native binding behavior, reducing flexibility, and making assay development time-consuming. Here, we present Microfluidic Diffusional Sizing (MDS), an innovative analytical technique that measures protein interactions directly in solution. By avoiding immobilization, MDS enables advanced assays that can simultaneously determine both affinity and calibration-free analyte concentration. These assays benefit from integrated Bayesian analysis, which provides a complete view of parameter space and streamlines experimental design for accurate results with minimal resources.

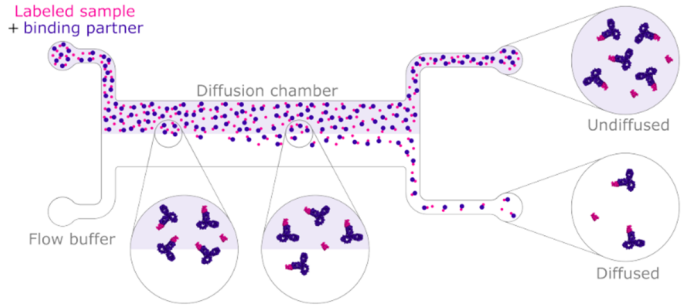


Figure 1: Operating principle of Microfluidic Diffusional Sizing (MDS).

(95) A map of hydrolytic triads informs the design of a triad-swapped plastic-degrading enzyme

Aleksandr Zlobin[1,2], Norbert Graefe[1], Georg Kuenze[1,2,3]

[1] Institute for Drug Discovery, Leipzig University, 04103 Leipzig, Germany
[2] Interdisciplinary Center for Bioinformatics, Leipzig University, 04107 Leipzig, Germany
[3] Center for Scalable Data Analytics and Artificial Intelligence, Leipzig University, 04105 Leipzig, Germany

aleksandr.zlobin@medizin.uni-leipzig.de

Recent advances in AI methods led to the creation of efficient de novo esterases. Their catalytic sites were borrowed from nature and then scaffolded into new folds. These sites were Ser-His-Asp triads, a canonical combination used by hydrolases. We noticed that, despite the task was to create an esterase, donor sites came from proteases. It is not clear whether spaces of proteolytic and esterolytic triads overlap, when a precise 3D geometry is considered, and, as a result, whether such choice is optimal. Here we performed an exhaustive classification of all Ser-His-Asp/Glu triads of proteolytic and esterolytic enzymes to show that these spaces are finite and well-defined, and that there is no overlap. We expect these findings to inform future de novo enzyme design strategies.

While performing the exploration of the triad space, we noticed triad-fold correspondence trends. Some folds tolerated multiple types of triads, for example, a fold that all known PET-degrading enzymes adopt. We wondered if a rational, knowledge-based triad swap is possible, to unlock a new potential starting point for directed evolution campaigns. We created an initial series of eight variants of PHL7 enzyme and showed that two of them retained esterolytic activity. Unsurprisingly, stability was compromised, and now we will focus on introducing additional stabilizing mutations with both computational and experimental methods. We will explore if the triad identity corresponds to substrate specificity, in this case, to plastics. Our results show that even such conserved regions as active sites are not necessarily immutable, if the function is considered, broadening the scope of strategies for enzyme redesign.

(96) Human AMBN -Derived Antimicrobial Peptides Targeting Biofilm Formation

Veronika Vetyskova[1,+], Petra Kasparova[2], Lucie Bednarova[1], Miroslav Hajek[1], Olga Matatkova[2], Jan Masak[2], Jiri Vondrasek[1,*], Kristyna Vydra Bousova[1,*]

[1] Institute of Organic Chemistry and Biochemistry of the Czech Academy of Sciences, Department of Synthetic Biology and Immunology, Prague, Czechia
[2] University of Chemistry and Technology, Department of Biotechnology, Prague, Czechia

[+] Currently at INM – Leibniz Institute for New Materials, Materials Synthetic Biology (PI: Prof. Dr. Wilfried Weber), Saarbrücken, Germany

[*] Correspondence: kristyna.bousova@uochb.cas.cz; jiri.vondrasek@uochb.cas.cz

Ameloblastin (AMBN) is a protein that belongs to a class of proteins known as intrinsically disordered proteins (IDPs). AMBN is one of the key proteins for proper dental enamel formation and, along with other enamel matrix proteins such as amelogenin, enamelin, and tuftelin, plays a critical role in dental biology. In humans, the AMBN gene gives rise to two isoforms, ISO I and ISO II, whose specific biological functions are still under investigation. This research focuses on exploring the antimicrobial capabilities of peptides de-rived from the ISO I variant.

During enamel development, AMBN is proteolytically processed by enamel-specific enzymes such as matrix metalloproteinase-20 (MMP-20) and kallikrein-4 (KLK-4). Studies have shown that peptides derived from enamel matrix proteins, including AMBN, exhibit antimicrobial properties, suggesting a dual role in both enamel development and innate immune defense. In this study, we investigated the antimicrobial potential of peptides de-rived from recombinant human AMBN ISO I, focusing on their activity against biofilm-forming bacteria.

Using mass spectrometry and computational analysis, we characterized the proteolytic cleavage patterns of AMBN by MMP-20 and KLK-4, which enabled us to identify peptide candidates with predicted antimicrobial activity using bioinformatics tools. These peptides were tested for their effectiveness against bacterial biofilms commonly found in dental plaque. The results demonstrated that the peptides significantly inhibited biofilm formation while maintaining low toxicity toward mammalian cells. This work highlights the multifunctionality of AMBN and contributes to the growing field of peptide-based therapeutics, particularly in the context of antibiotic resistance and implant-associated infections.

References:

[9] V. Vetyskova et al. Proteolytic profiles of two isoforms of human AMBN expressed in E. coli by MMP-20 and KLK-4 proteases. *Heliyon*, 10(2): e24564, 2024

(97) Machine Learning-Accelerated Quantum Mechanical Energy Function with Electron Density based Energy Decomposition

Maximilian Hoffmann[1], Aleksandr Kazimir[1], Toni Oestereich, Luisa Kaermer[2], Felipe Engelberger, Lukas Marche, Christina Lamers, Jens Meiler[1,2]

[1] Institute for Drug Discovery, University Leipzig, Germany
[2] Department of Chemistry, Vanderbilt University, Nashville, Tennessee 37240, US

Maximilian.Hoffmann@medizin.uni-leipzig.de, alexandr.kazimir@uni-leipzig.de

Accurate evaluation of the energy contribution in the enthalpy term in biomolecules is essentially for elucidating their structures and interactions, properties and functions. Traditional approaches, relying mostly on classical terms, heuristics and empirical parameters, struggle to comprehensively capture the intricate interplay of **non-covalent interactions** and crucial **quantum mechanical** (QM) effects. Routinely it requires to use the established QM methods **Density Functional Theory** (DFT) which are computationally prohibitive for large biomolecules, limiting their application mainly to smaller systems. [1] Further, for gaining insight into the intra molecular interactions a variety of QM **Energy Decomposition Schemes** (EDA) exist, but most of them are limited to a certain kind of interaction or geometric constraints (i.e. only operating on fragments or limited to equilibrium structures). [2] Among the variety of EDA schemes, **the Interacting Quantum Atoms** (IQA) scheme is an exceptionally attractive one due to the separation **into intra- and inter-atomic decomposition in an intuitive manner** based on topological analysis of **the electron density**. [3]

We address these limitations by introducing a novel machine learning (ML) framework incorporating a state-of-the-art **equivariant graph neural network**, designed to **predict QM-accurate energies from DFT calculations alongside the decomposed IQA terms** applicable for covalent as well as non-covalent interactions and **not limited to predefined geometrical constraints**. This innovative strategy not only enables the **rapid and accurate** prediction of the total energy for arbitrary biomolecular structures within the chemical space covered by our training data, but also offers an unprecedented ability to dissect these energies based in a physical meaningful and intuitive manner. Our method provides **intra-atomic energies and inter-atomic covalent and non-covalent interaction energies for every atom and atom pair within a defined distance**. This detailed energy breakdown transforms our approach into a powerful tool for molecular analysis and design. It provides **profound insights into the stabilizing and destabilizing energy contributions within a molecular system**, which is invaluable for applications such as **analyzing protein structures, understanding ligand-target binding mechanisms, and guiding rational drug design**.

References:

[1] Ginex, Tiziana, et al. "Quantum mechanical-based strategies in drug discovery: Finding the pace to new challenges in drug design." Current Opinion in Structural Biology 87 (2024): 102870.

[2] EU Phipps, Maximillian JS, et al. "Energy decomposition analysis approaches and their evaluation on prototypical protein–drug interaction patterns." Chemical society reviews 44.10 (2015): 3177-3211.

[3] Guevara-Vela, José Manuel, et al. "Interacting quantum atoms—a review." Molecules 25.17 (2020): 4028.

(98) The Protein Binder Printer: Fully Automated Closed-Loop AI-Driven Protein Binder Design

Diego Ruiz Espi[1], Gustavo Avelar Molina[1], Esperanza Rivera de Torre[1], Andreas Hald[2], Anne Ljungars[1], Timothy Patrick Jenkins[1]

[1] Section for Biologics Engineering, Department of Biotechnology and Biomedicine, Technical University of Denmark, Kongens Lyngby, Denmark
[2] Research and Development, Novo Nordisk AS, Malov, Denmark

direspi@dtu.dk

The discovery of protein binders, to be used as reagents, for diagnostic purposes, and as therapeutics is a complex, costly, and time-consuming process. To date, there has been a significant reliance on animal immunisation, followed by different binder selection/enrichment steps.

Here we propose to leverage recent advances in generative AI to shift the focus from binder discovery to binder design. Specifically, we will explore several approaches toward generative library generation, followed by in silico screening and in vitro evaluation and refinement.

To ensure a direct and meaningful feedback loop, we will establish a series of high-throughput automated in vitro assays. These assays will provide quantitative measurements that will be used for refinement and optimization in a closed loop. This will require sophisticated cloning strategies combined with rapid cell free expression approaches, followed by binding, stability, and solubility evaluations.

Together, this pipeline will present the foundation for a fast, truly closed loop, and fully automated binder design platform.

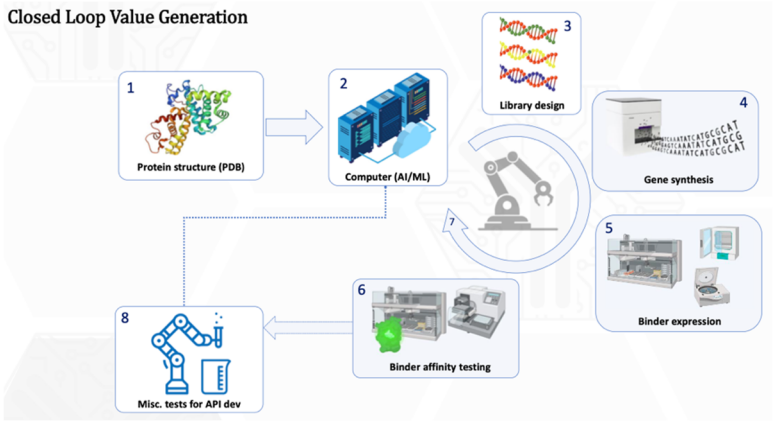


Figure 1: First Idea of the Closed-Loop Pipeline. *The Closed-Loop pipeline starts with the PDB file of the selcted target as input. It first generates protein binders against the target and evaluates them in silico. After several optimization rounds, the genes of the binders are synthesized, expressed and validated with FIDA. The positive hits undergo further in silico optiomization and in vitro validation until the required properties are met.*

(99) Computer-aided design of anti-VEGF proteins

Abibe Useini[1], Carsten Geist[2], Stefan Kalkhof[2, 3] and Georg Künze[1,4,5]

[1] Institute for Drug Discovery, Leipzig University, Leipzig, Germany
[2] Department of Preclinical Development and Validation, Fraunhofer Institute for Cell Therapy and Immunology—IZI, Leipzig, Germany
[3] Institute for Bioanalysis, Coburg University of Applied Sciences and Arts, Coburg, Germany
[4] Interdisciplinary Center for Bioinformatics, Leipzig University, 04107 Leipzig, Germany
[5] Center for Scalable Data Analytics and Artificial Intelligence (ScaDS.AI), Leipzig University, Leipzig, Germany

abibe.useini@bbz.uni-leipzig.de

Computational protein design can create therapeutics with improved efficacy, specificity, and reduced side effects [1]. Vascular endothelial growth factor A (VEGFA) is a key protein required for blood vessel growth, but its dysregulation contributes to diseases such as cancer and diabetic retinopathy [2]. While current VEGFA inhibitors are clinically effective, their limitations highlight the need for alternative approaches. Peptide- and miniprotein-based inhibitors provide advantages including reduced toxicity and enhanced stability.

In this study, we applied advanced design tools, particularly ProteinMPNN, Rosetta, and BindCraft to generate optimized miniprotein inhibitors of VEGFA. Building on the established templates Z-1-2 and Z-3B [3], we computationally predicted new sequences and validated them experimentally. The Z-1-2 miniprotein served as a promising starting point due to its strong affinity and anti-angiogenic activity, though further optimization is required. By efficiently exploring vast sequence space computationally and experimentally testing only the most promising candidates, we demonstrate a streamlined strategy for developing next-generation VEGFA inhibitors. Low micromolar concentrations of our best miniproteins designed so far achieve almost complete VEGF inhibition, with even higher efficacy being expected for the upcoming design rounds.

References:

[1] MR. Swift, BM. Weinstein. Arterial–Venous Specification During Development. *Circulation Research*, 104(5):576–88, 2009.

[2] S. Kazerounian, J. Lawler. Integration of pro- and anti-angiogenic signals by endothelial cells. *The Journal of Cell Communication and Signaling*, 12(1):173, 2018.

[3] A. Fedorova, K. Zobel, HS. Gill, A. Ogasawara, JE. Flores, JN. Tinianow, et al. The Development of Peptide-Based Tools for the Analysis of Angiogenesis. *Chemistry & Biology*, 18(7):839–45, 2011.

(100) *De novo* design of phosphorylation-induced protein switches for synthetic signalling in cells

Stephen Buckley[1‡], Yangyang Miao[1‡], Mubarak Idris[1], Pao-Wan Lee[2], Leo Scheller[1], Roland Riek[3], Sebastian J. Maerkl[2], Luciano A. Abriata[4*], Bruno E. Correia[1*]

[1] Laboratory of Protein Design and Immunoengineering, Institute of Bioengineering, École Polytechnique Fédérale de Lausanne and Swiss Institute of Bioinformatics, Lausanne, Switzerland

[2] Laboratory of Biological Network Characterization, Institute of Bioengineering, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

[3] Institute of Molecular Physical Science, ETH Zurich, Zurich, Switzerland

[4] Protein Production and Structure Core Facility, School of Life Sciences, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

* Corresponding authors. Email: bruno.correia@epfl.ch , luciano.abriata@epfl.ch

Cells, as the basic unit of life have evolved mechanisms in order to respond to endogenous and exogenous cues, critical for survival. Many of these cues propagate intracellular signalling cascades, allowing a cell to respond rapidly and robustly to upstream signalling input. Protein phosphorylation is at the core of many of these signalling cascades, as it acts as a mechanism to regulate protein function, relaying intracellular signals. This regulation relies on conformational dynamics, with many kinase substrates populating distinct structural and functional states in response to phosphorylation. While *de novo* protein design has made great advances, encoding such dynamic behaviour remains a major challenge. We address this by designing Phosphorylation-Induced Protein Switches (PIPS) that populate distinct conformational states in response to phosphorylation by serine and tyrosine kinases - PKA, Lck and Abl1. We demonstrate our ability to integrate these PIPS into numerous synthetic signalling networks, driving transcriptional responses, subcellular localisation and protein degradation.

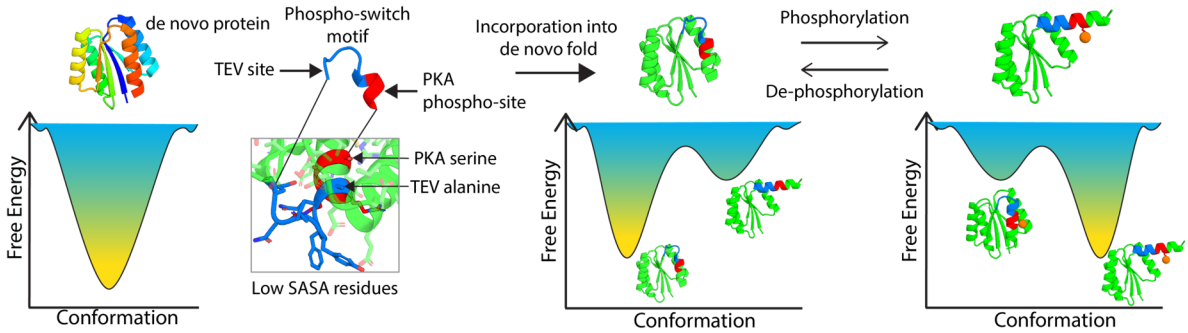


Figure 1: Schematic representation of phospho-switch mechanism. *Generally, De novo proteins adopt a single hyperstable conformation. By incoporating a Phospho-Switch motif – consisting of a protease cleavage site and a kinase phospho-site – into de novo scaffolds, two distinct conformational states which are populated in response to phosphorylation and de-phosphorylation can be introduced.*

(101) Bioengineering plant immune recognition of pathogen virulence proteins using *de novo* binding domains.

Gregory Knight[1], Angus H. Bucknell[2], Yuxuan Xi[3], Amber Gentle[2], Lauren S. Ryder[2], Xia Yan[2], Joseph L. Watson[4], Peter M. F. Emmrich[3,5], Nicholas J. Talbot[2], Mark J. Banfield[3], Adam R. Bentham[1]

[1] Centre for Programmable Biological Matter, Department of Biosciences, Durham University, DH1 3LE, United Kingdom.
[2] The Sainsbury Laboratory, University of East Anglia, Norwich Research Park, Norfolk, NR4 7UH, United Kingdom.
[3] Department of Biochemistry and Metabolism, John Innes Centre, Norwich Research Park, Norfolk, NR4 7UH, United Kingdom.
[4] Institute for Protein Design, University of Washington, Seattle, WA, United States of America.
[5] Norwich Institute for Sustainable Development, School of Global Development, University of East Anglia, Norwich, NR4 7TJ, United Kingdom.

Gregory.Knight@Durham.ac.uk

The global, annual yield loss from phytopathogenic disease is estimated to be 10-30%, and is rising each year[1]. This increase is climate change driven, exacerbated by industrial farming practices such as monoculture. Effector proteins are small virulence factors secreted by phytopathogens that function to promote infection and suppress the hosts immune response[2]. The recognition of effector proteins by plant NLRs, a type of intracellular immune receptor, is crucial to activation of effector triggered immunity (ETI), a response which culminates in localized cell death to limit the spread of infection[3]. A subset of NLRs have evolved to contain integrated domains (IDs), which enable recognition of their cognate effector proteins[4].

NLR IDs are diverse, with many different protein folds used as recognition modules by NLRs in nature. Previous work has shown IDs to be modular, with integration of different IDs facilitating new NLR recognition profiles through enabling binding of non-cognate effectors [5]. The development of AI-protein design has created the possibility of new to nature, *de novo* proteins. Here we have used the RFdiffusion[6] – ProteinMPNN[7] – AlphaFold2[8] pipeline to design *de novo* binders against phytopathogenic effector AVR-Pii. We demonstrate integration of these binders into the NLR pair Pikm-1/Pikm-2, alters their effector binding profile and enables recognition AVR-Pii.

The binders were initially tested for activity *in planta* in *Nicotiana benthamiana* before biochemical characterisation for shape and molecular weight. Finally, we characterised the structure of the complex of each binder with AVR-Pii to understand their interaction. This work offers promise for the deployment of this technology within rice as well as other cereal crops which will help us build resilient solutions to protect global food security.

References:

1. Ristaino, J. B. *et al.* The persistent threat of emerging plant disease pandemics to global food security. *Proc. Natl. Acad. Sci.* **118**, e2022239118 (2021).
2. Hogenhout, S. A., Van der Hoorn, R. A. L., Terauchi, R. & Kamoun, S. Emerging Concepts in Effector Biology of Plant-Associated Organisms. *Mol. Plant-Microbe Interactions®* **22**, 115–122 (2009).
3. Peng, Y., van Wersch, R. & Zhang, Y. Convergent and Divergent Signaling in PAMP-Triggered Immunity and Effector-Triggered Immunity. *Mol. Plant-Microbe Interactions®* **31**, 403–409 (2018).
4. Cesari, S., Bernoux, M., Moncuquet, P., Kroj, T. & Dodds, P. N. A novel conserved mechanism for plant NLR protein pairs: the 'integrated decoy' hypothesis. *Front. Plant Sci.* **5**, (2014).
5. Maidment, J. H. *et al*. Effector target-guided engineering of an integrated domain expands the disease resistance profile of a rice NLR immune receptor. *eLife* **12**, e81123.
6. Watson, J. L. *et al.* De novo design of protein structure and function with RFdiffusion. *Nature* **620**, 1089–1100 (2023).
7. Dauparas, J. *et al.* Robust deep learning–based protein sequence design using ProteinMPNN. *Science* **378**, 49–56 (2022).
8. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

(102) Efficient nanobody cross-binding discovery through antibody structure prediction and design

Max D. Overath[1], Suthimon Thumtecho[1], Esperanza Rivera-de-Torre[1] Melisa Bénard-Valle[1], Darian S. Wolff[1,2], Nils Hofmann[1], Anne Ljungars[1], Andreas H. Laustsen[1], Simon Olsson[5], Thomas J. Fryer,[6#], Timothy P. Jenkins[1#]

[1] Department of Biotechnology and Biomedicine, Technical University of Denmark, Kongens Lyngby, Denmark
[2] Novonesis, Kongens Lyngby, Denmark
[5] Department of Computer Science and Engineering, Chalmers University of Technology and University of Gothenburg, SE-41296 Gothenburg, Sweden
[6] IMES, Massachusetts Institute of Technology, Cambridge, MA, USA
[#] equal contribution
maxove@dtu.dk

Therapeutic antibodies are critical for treating various diseases, yet their discovery remains time- and resource-intensive. Successful development requires antibodies with high affinity and favorable biophysical properties, and in some applications also broad neutralization across related targets. Snakebite envenomation exemplifies this challenge, as promising next-generation treatments are expected to rely on broadly neutralizing antibodies that cross-bind venom toxins from multiple snake species. To address this, we developed a high-throughput computational-experimental pipeline that integrates antibody-antigen complex prediction and antibody design to accelerate the identification of cross-binding nanobodies against snake long-chain α-neurotoxins (LNTxs).

Yeast display and deep sequencing yielded a diverse dataset of cross- and mono-binding nanobody sequences against LNTxs. A central innovation of the pipeline was the integration of AlphaFold3 [1] to generate high-confidence nanobody-toxin complex models from this dataset, to identify nanobodies engaging conserved toxin epitopes. These predicted structures were then used to design optimized variants with a recently described structure-informed language model for affinity improvement [2] and CamSol for solubility enhancement. Final candidates were validated with low-cost DNA assembly and split-luciferase assays, enabling parallel testing of hundreds of designs and identifying several variants with improved solubility and affinity.

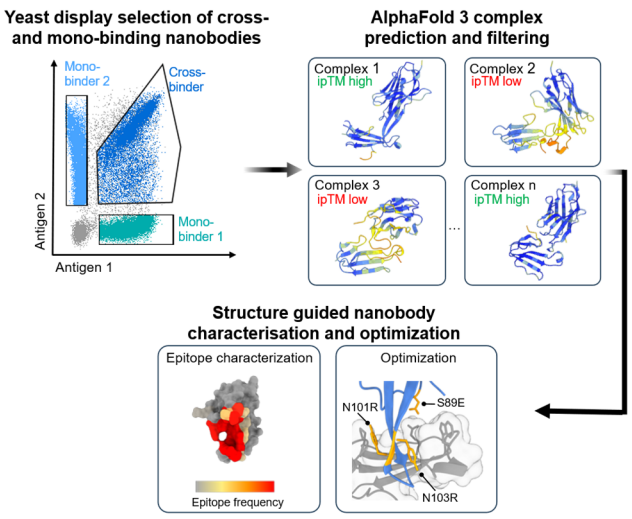


Figure 1: Overview of pipeline

This work illustrates how state-of-the-art structure prediction tools, combined with high-throughput screening, can accelerate antibody discovery pipelines, with applications extending beyond venom toxins to broadly applicable frameworks for antibody design.

References:

[1] Abramson et. al, Accurate structure prediction of biomolecular interactions with AlphaFold 3**,** *Nature* 630, 493–500 (2024)
[2] Shanker et. al, Unsupervised evolution of protein and antibody complexes with a structure-informed language model, *Science* 385,46-53(2024)

(103) Engineering structurally defined SlyB nanodiscs through de novo protein design

Jose L. Morales Hellin[1,2], Toon van Thillo[1], Koen Martens[2], Pol van Dorpe[2], Anastassia A. Vorobieva[1]

[1] VIB-VUB Center for Structural Biology, Brussels, Belgium
[2] Imec, Leuven, Belgium

jose.morales.hellin@vub.be

Lipid nanodiscs have emerged as powerful tools for membrane protein research, providing native-like lipid environments in a soluble format. However, current nanodisc systems often suffer from heterogeneous sizes, limiting their applications in structural biology and biotechnology. Here, we present a comprehensive approach to engineer modular lipid nanodiscs with defined composition and controlled oligomeric states using the bacterial outer membrane protein SlyB [1] as a scaffold.

SlyB naturally oligomerizes into nanodiscs that encapsulate small phospholipid bilayers, featuring a transmembrane helical hairpin topped with a water-soluble domain. However, these nanodiscs adopt heterogeneous sizes and oligomeric states in detergent. The goal of this project is to engineer SlyB variants with a stable and defined oligomeric state, focusing on the water-soluble domains. These regions offer greater possibilities for mutations relative to the membrane-spanning domains, thereby preserving essential protein function while enabling targeted structural modifications. For this we will adapt the generative deep-learning model RFDiffusion [2] for membrane protein design by conditioning it on coarse-grained representations of lipid bilayers. To bias designs toward peripheral extensions, we will introduce a lattice of virtual residues representing the bilayer or micelle surface, positioned using PPM3.0, a tool for modeling membrane protein orientation. We will further implement partial noising/denoising trajectories using RFDiffusion retrained on MemProtMD datasets to capture membrane protein-lipid interactions and local bilayer distortions. Our designs will be validated through structural prediction (AlphaFold2), followed by experimental testing including transmission electron microscopy, and cryo-electron microscopy analysis.

This work will try to produce a fully characterized lipid nanodisc with potential applications in membrane protein characterization (cryo-EM structure resolution) and therapeutics (molecule deliverability).

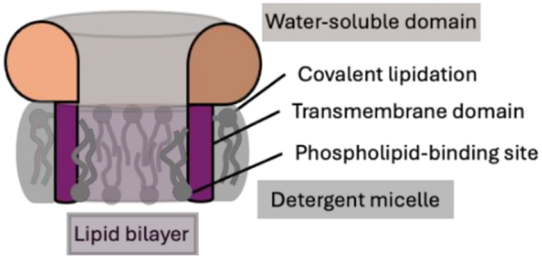


Figure 1: SlyB nanodisc molecular schematic architecture.

References:

[1] Janssens, A., Nguyen, V.S., Cecil, A.J. *et al.* SlyB encapsulates outer membrane proteins in stress-induced lipid nanodomains. *Nature* **626**, 617–625 (2024).
[2] Watson, J.L., Juergens, D., Bennett, N.R. *et al.* De novo design of protein structure and function with RFdiffusion. *Nature* **6**

(104) Improving the affinity of an EpCAM-binding peptide using *in silico* approaches for *in vivo* tumor imaging

Nada Badr [1,2*], Nicky Janssen[2], Mark Fonteyne[1,2], Marvin Nogaitzig[1], Taryn March[1], Shadhvi Bhairosingh[1], Alexander L. Vahrmeijer[1], Gerard J.P. van Westen[2], and Peter J. K. Kuppen[1]

[1]Department of Surgery, Leiden University Medical Center, Leiden, The Netherlands

[2]Drug Discovery and Safety, Leiden Academic Centre of Drug Research, Leiden, The Netherlands

Surgery is the primary cancer treatment, but is often limited by poor tumor visibility[1]. We aim to improve tumor visibility by developing tumor-binding peptide-based near-infrared fluorescent (NIRF) probes[2]. Novel computational tools enable the prediction of peptide-protein interactions, potentially supporting the design of target-specific probes. The epithelial cell adhesion molecule (EpCAM) is a transmembrane protein highly expressed in epithelial cancers, and its extracellular domain (EpEX) is a suitable imaging target[3,4]. This study computationally optimizes an EpCAM-binding peptide and evaluates this using *in vitro* and *in vivo* experiments. An *in* silico alanine scan on a known binder, EP114, was performed using ICM Molsoft, AutoDock CrankPep, AlphaFold2 (AF2), and RosettaFold2 (RF2). Platforms' scoring functions were used to evaluate the impact of alanine substitutions on binding. AF2 simulations indicated that serine at position 12 contributed less to binding than an alanine. This residue was therefore substituted with all naturally occurring amino acids and ranked for EpEX interaction. Methionine provided the best result, increasing the AF2 confidence score from 74% for EP114 to 83% for EP5. EP2 (control), EP5 (highest-scoring), and EP114 were synthesized and conjugated to fluorescein (FITC) or indocyanine green (ICG). EP5 showed improved affinity of 6.99 nM compared to EP114 with an affinity of 80.92 nM. Next, EpCAM-positive HT-29 and EpCAM-negative HT-1080 cells were incubated with the probes. EP5-FITC and EP5-ICG bound specifically to EpCAM in HT-29 and not to HT-1080, with a higher fluorescence signal than EP114 fluorescent conjugates. NIRF imaging in HT-29 tumor-bearing mice showed that after administration of 4 nmol of the ICG-labelled probes, EP5-ICG resulted in consistently higher tumor fluorescence compared to EP114-ICG and EP2-ICG. This study demonstrates that computational tools such as AF2, can predict peptide-protein cofolding and guide optimization. EP5 is a promising EpCAM-targeted NIRF probe for surgical resection.

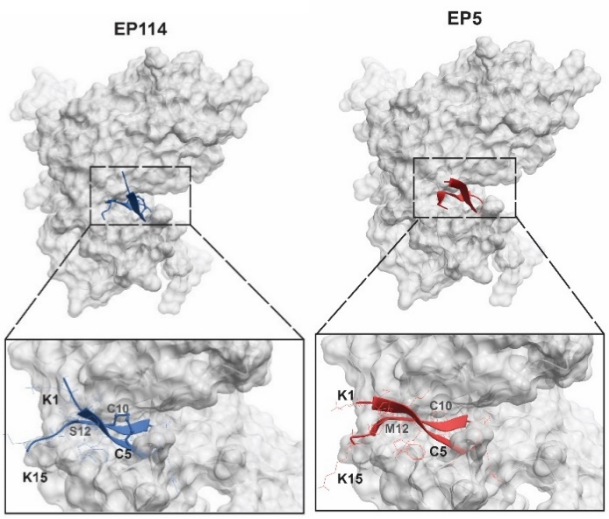


Figure 1. Visualization of the predicted protein-peptide complex formation using AF2. Interaction between EpEx (extracellular domain of EpCAM), shown in grey, EP114, the reference peptide, shown in **blue**, and the improved peptide, EP5, shown in **red**.

References:

1. DOI: 10.3322/caac.70011
2. DOI: 10.1038/nrclinonc.2013.123
3. DOI: 10.1038/ncomms5764
4. DOI: 10.1007/s00259-023-06407-w

(105) Structural determination of Y4R complexes with hexapeptides using a Rosetta-based computational pipeline

Mateusz Sklodowski[1], Fabian Liessmann[1], Paul Eisenhuth[1], Tim Pelczyk[2], Corinna Schüß[1,2], Albert O. Gattor[3], Annette Beck-Sickinger[2], Max Keller[3], Jens Meiler[1,4,5]

1 Institute for Drug Discovery, Leipzig University, Leipzig, Germany
2 Institute of Biochemistry, Leipzig University, Leipzig, Germany
3 Institute of Pharmacy, University of Regensburg, Regensburg
4 Center for Scalable Data Analytics and Artificial Intelligence, Leipzig, Germany
5 Center for Structural Biology, Vanderbilt University, Nashville, USA

mateusz.sklodowski@medizin.uni-leipzig.de

Peptides serve diverse functions in the human body, including regulatory and signaling roles. Their involvement in G-protein signaling cascades has been of major interest, as these pathways influence appetite, stress responses, blood pressure, and cardiovascular disease [1]. In recent years, cyclic peptides have attracted increasing attention in drug discovery [2]. Their vast combinatorial potential enables fine-tuning for target specificity, while their cyclic structure confers superior metabolic stability compared to linear peptides. Furthermore, solid-phase peptide synthesis facilitates partial automation of early development steps and helps reduce the cost of generating potential therapeutics [3].

In this study, we computationally investigated interactions between the Y4 receptor (Y4R), its native agonist pancreatic polypeptide (PP), and two agonistic cyclic synthetic hexapeptides, UR-AK95c and UR-AK86c [4]. We developed a Rosetta-based pipeline [5] to predict peptide and Y4R structures. Starting from the structures of both hexapeptides, we integrated them into the Rosetta framework and performed molecular screening against the receptor structure. The predictions were subsequently validated via experimental mutagenesis. By combining computational and experimental approaches, we identified key interactions between Y4R and its ligands. Notably, our findings align with previous small-molecule studies, suggesting that critical interaction patterns are preserved across different ligand classes.

References:

[1] Wang, Lei, et al. "Therapeutic peptides: current applications and future directions." Signal transduction and targeted therapy 7.1 (2022): 48.
[2] Bechtler, Clément, and Christina Lamers. "Macrocyclization strategies for cyclic peptides and peptidomimetics." RSC Medicinal Chemistry 12.8 (2021): 1325-1351.
[3] Jaradat, Da'san MM. "Thirteen decades of peptide synthesis: key developments in solid phase peptide synthesis and amide bond formation utilized in peptide ligation." Amino acids 50.1 (2018): 39-68.
[4] Konieczny, Adam, et al. "N-terminus to arginine side-chain cyclization of linear peptidic neuropeptide Y Y4 receptor ligands results in picomolar binding constants." Journal of Medicinal Chemistry 64.22 (2021): 16746-16769.
[5] Koehler Leman, J. et al. Better together: Elements of successful scientific software development in a distributed collaborative community. PLoS computational biology 16, e1007507; 10.1371/journal.pcbi.1007507 (2020).

(106) Computational Design of α/β Propeller Building Blocks for Protein Assemblies

Yueyan Wu[1], Aurora Pierucci[1], Arnout Voet[1]

[1] Biochemistry, Molecular and Structural Biology, Department of Chemistry, Leuven, Belgium

yueyan.wu@kuleuven.be, arnout.voet@kuleuven.be

Protein assemblies hold great potential in diverse applications. While symmetry is pivotal in artificial protein assembly design, perfect symmetry is often absent in natural proteins, hindering the development of assemblies based on them. Computational protein design emerges as a solution. Starting with ***Pizza*** [1], various perfectly symmetric artificial β propeller proteins were designed in our lab. Some of them are proven as building blocks for protein assemblies based on metal or ligand binding strategies. However, they are environment-sensitive and have low biocapacity, which limits their applications. To overcome this, we introduced α helices into the β propeller protein, offering a promising extension which may form self-assemblies based on the protein-protein interaction of α helices.

The preliminary α/β propeller building blocks were designed with a combination of knowledge-based methods and advanced deep-learning methods which were successfully validated by biochemical experiments and **X-ray diffraction**. Based on these, new building blocks will be designed, and future work will explore the design of protein assemblies, especially 2D arrays and 3D cages, for applications in enzyme scaffolds, biosensors, and drug delivery.

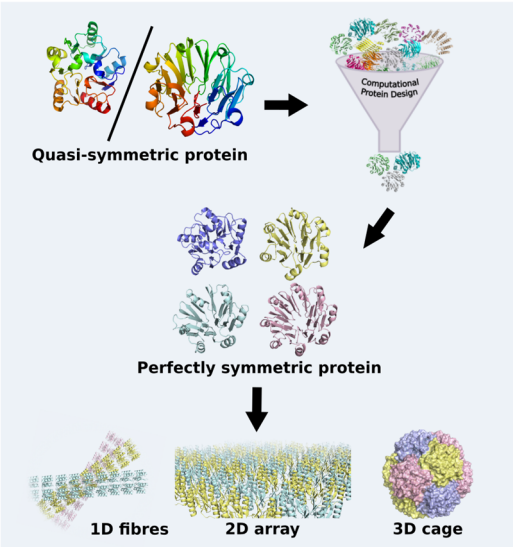


Figure 1: Overview of computational design of α/β Propeller Building Blocks. *Starting from natural quasi-symmetric proteins, perfectly symmetric building blocks are designed. By varying the symmertic elements and interfaces, these building blocks can be assembled into diverse higher-order structures, such as 1D fibres, 2D arrays, and 3D cages, with broad potential applications.*

References:

[1] A.R.D. Voet,H. Noguchi,C. Addy,D. Simoncini,D. Terada,S. Unzai,S. Park,K.Y.J. Zhang, & J.R.H. Tame, Computational design of a self-assembling symmetrical β-propeller protein, Proc. Natl. Acad. Sci. U.S.A. 111 (42) 15102-15107, https://doi.org/10.1073/pnas.1412768111 (2014).

(107) Closing the Loop: Fast Wet Lab Validation for AI-Guided Protein Design

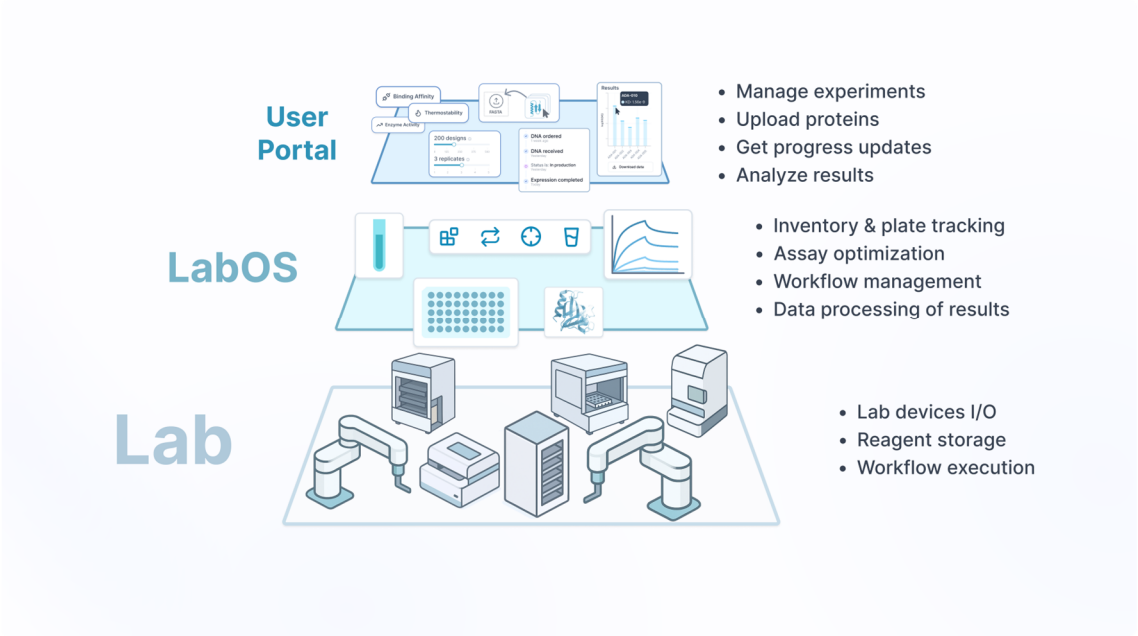
Julian Englert[1]

[1]Adaptyv

julian@adaptyvbio.com

Adaptyv accelerates data generation for training and validating AI models with a high-throughput lab that companies can access via our software interface and API. We empower protein-design teams to validate their AI models many times faster than before, without the need to run in-house wet labs. Overall, this year, over 30 companies started using Adaptyv to validate their protein designs - from some of the biggest pharmas to frontier AI labs to many seed-stage techbio startups. We've run hundreds of experiments, tested well over 10,000 proteins this year and are generating the data that validates the best AI models currently in development.

In this talk, we'll give you a behind-the-scenes look into the infrastructure that we've built to run a lab at high-throughput with a small team of engineers. We'll cover the software stack (what the customer sees vs what's going on internally), how we're thinking about lab automation and how we're improving biology protocols using the data we're generating.



(108) Crowdsourced protein design: The Adaptyv Protein Design Competition and BenchBB

Tudor-Stefan Cotet[1,2], Igor Krawczuk[2], Julian Englert[2]

[1] ETHZ, [2] Adaptyv

tcotet@student.ethz.ch, tudor@adaptyvbio.com

The 2024 **Adaptyv Protein Design Competition** [1] challenged participants to create proteins to bind the Epidermal Growth Factor Receptor (EGFR), an important cancer drug target, using a variety of computational and Machine Learning (ML) methods. This initiative aimed to foster innovation and collaboration between protein designers (through shared media like blog posts and threads) and to address the complexities of designing functional protein binders. The competition saw extensive participation, with over 1,800 designs submitted across two rounds. From these submissions, 601 proteins were selected for experimental characterization of expression and binding affinity, utilizing an automated Bio-Layer Interferometry (BLI) pipeline. The results were promising, yielding both the optimization of existing binders and the creation of *de novo* binders [2]. It also saw the rise of promising new design tools such as BindCraft [4], which won the first round. Our subsequent manuscript thoroughly analyses the changes in design strategies across the two rounds, why some methods were better, and how current *in silico* binding predictors are not reliable enough [2].

While the competition successfully demonstrated the power of crowdsourcing in protein design, it also underscored significant challenges within the field, particularly the absence of standardized experimental benchmarks. To address this gap, we propose the Bench-tested Binder Benchmark (**BenchBB**) [2][3], a curated set of protein-binding targets intended to facilitate the evaluation of computational binder design methods, along with a set of guidelines for experimentally validating hits.

We anticipate that this combination of open competition and standardized benchmarking will be a key driver of future innovation in protein design.

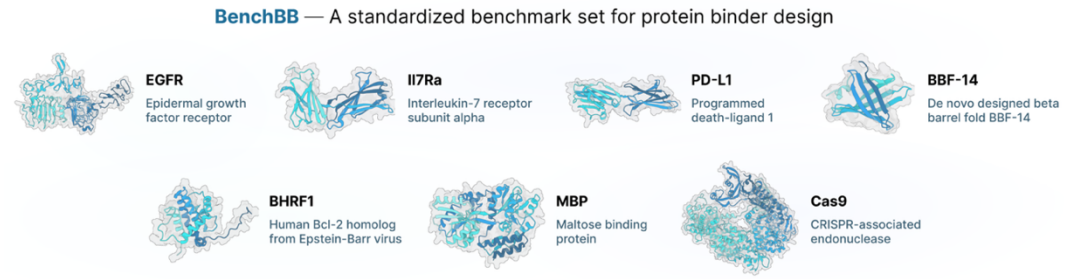


Figure 1: The BenchBB benchmark targets. *BenchBB is a curated set of 7 protein targets designed to provide an accessible, consistent, and practical benchmark for computational binder design methods.*

References:

[1] Adaptyv Protein Design Competition https://foundry.adaptyvbio.com/competition

[2] Cotet, T.-S., Krawczuk, I., Stocco, F., Ferruz, N., Gitter, A., et al. (2025) Crowdsourced Protein Design: Lessons From the Adaptyv EGFR Binder Competition. *bioRxiv*. 2025.04.17.648362. doi:10.1101/2025.04.17.648362.

[3] BenchBB blog post https://www.adaptyvbio.com/blog/benchbb

[4] Pacesa, M., Nickel, L., Schmidt, J., Pyatova, E., Schellhaas, C., et al. (2024) BindCraft: one-shot design of functional protein binders. *bioRxiv*. 2024.09.30.615802. doi:10.1101/2024.09.30.615802.

(109) *De novo* design of binders targeting monomeric α-synuclein

Yunxuan Zhang[1], Hannah L. Han[1], Danny D. Sahtoe[2], Tuomas P.J. Knowles[1]

[1] Centre for Misfolding Diseases, Yusuf Hamied Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge
[2] Hubrecht Institute for Development Biology and Stem Cell Research, Utrecht, Netherlands

yz922@cam.ac.uk, hlh40@cam.ac.uk

The aggregation of α-synuclein is central to the pathology of Parkinson's disease, yet strategies to directly engage its monomeric form remain limited. We are applying a *de novo* design pipeline integrating backbone generation, sequence design, and structure prediction to create binders against aggregation-prone regions of α-synuclein. Experimental characterisation of the initial set of designs identified three structurally different binders with affinities in the high nanomolar to low micromolar range that inhibited α-synuclein aggregation. To further refine the approach, we implemented a partial diffusion strategy, which markedly improved AlphaFold confidence scores and yielded an expanded set of candidates for testing. Together, these results establish a pipeline for generating functional binders against intrinsically disordered proteins and highlight opportunities for probing and modulating α-synuclein biology.

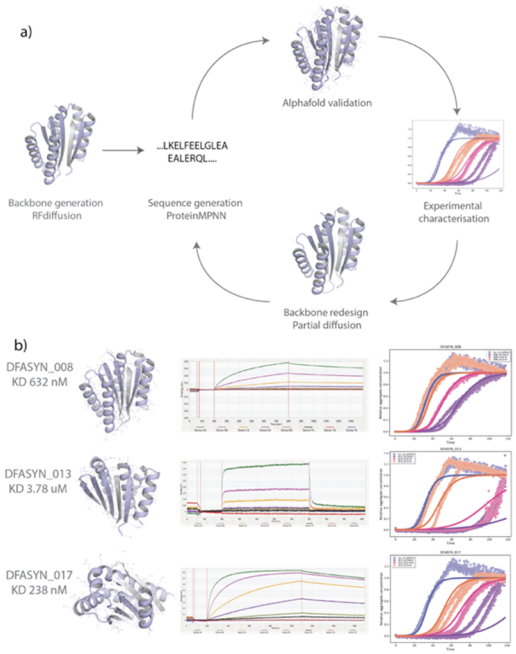


Figure 1: Computational design and experimental validation of de novo binders targeting monomeric α-synuclein. *(A) Schematic of the design pipeline combining RFdiffusion backbone generation, ProteinMPNN sequence design, and ColabFold validation. Incorporation of a partial diffusion strategy led to markedly improved AlphaFold confidence scores. (B) Experimental screening of the initial designs identified binders (left) with measurable affinity in the high nanomolar to low micromolar range, as assessed by bio-layer interferometry (middle). Functional aggregation assays demonstrated that selected binders inhibit α-synuclein fibrillization (right).*

(110) *De Novo* Design and Characterization of novel-function cytokines

Valeriia Hatskovska[1,2], Julia Skokowa[1], Birte Hernandez Alvarez[2], Mohammad El-Gamacy[1,2]

[1] Department of Internal Medicine II, University Hospital Tuebingen, Tuebingen, Germany.
[2] Department of Protein Evolution, Max Planck Institute for Biology – Tuebingen, Tuebingen, Germany.

Selective targeting of leukemic stem cells remains an outstanding challenge due to several reasons. Among these are their low abundance, dormant nature, and lack of specific markers. Trials to target these using recombinant cytokines have often failed owing to the poor stability, difficulties in heterologous expression, and pleiotropic activity of these proteins. To address these challenges, we use a fragment-based approach to design novel-function cytokines (novokines) capable of co-binding and thus co-targeting different receptor types. This approach greatly enhances the specificity of targeting leukemic stem cells, and yields low-molecular weight proteins that are hyperstable, easy to produce, and associate the target receptors in defined orientations.
Here I report our results on creating proteins co-engaging two key hematopoietic receptors, IL-3Rα and G-CSFR, with high affinity. Five candidates were expressed with high yields, purified as monomers, and thermostable. Two showed pico- to low-nanomolar affinities and co-engaged receptors simultaneously at nanomolar concentrations. Functional assays confirmed specific receptor co-activation and signaling comparable to natural cytokines. These proteins validate a modular design approach for generating compact, high-avidity ligands applicable to diverse therapeutic targets.

(111) Computational Design of in-house Ene-reductase Panel for Asymmetric Hydrogenation

Yuqi Yu, Andrea Pettenuzzo, Lucy Ward, Eric Lang, Michael Nunn, Nathaniel Gordon, Alba Diaz-Rodriguez

Early Chemical Development, Pharmaceutical Sciences, R&D AstraZeneca, Cambridge Biomedical Campus, Cambridge, UK

yuqi.yu@astrazeneca.com

Asymmetric hydrogenation of activated alkenes is essential in the production of pharmaceuticals, agrochemicals, and fine chemicals. Traditionally, this process relies on costly metal catalysts that face supply issues and lack selectivity. In contrast, biocatalysts like Ene-reductases (ERED) can catalyse reductions under ambient conditions with high enantioselectivity, offering a sustainable and efficient method for producing valuable chiral compounds. In this study, we present the design of an *in-house* diverse ERED panel using a combination of in silico tools. We collected sequences from literature, the UniProt, and metagenomic databases, resulting in a pool of 50,000 initial sequences. After clustering and analysing critical catalytic motifs, we selected 80 enzymes. Additionally, 16 sequences were artificially generated using ancestral reconstruction and generative AI methods, such as ProteinMPNN and RFdiffusion. The synthesized 96 genes were expressed in *E. coli*, achieving a 70% soluble expression rate confirmed by SDS-PAGE analysis. Screening against two internal active pharmaceutical ingredients yielded significantly better hit rates than the commercial panels.

(112) Hybrid Modelling to Support Biologics Formulations in Pharmaceutical Development

Massimo G. Totaro[1], Michael Brunsteiner[1]

[1] Biomolecular modelling Unit, RCPE GmbH, Graz, Austria

massimo.totaro@rcpe.at

Biomolecules such as **monoclonal antibodies** and **mRNA**-based therapeutics are at the forefront of modern pharmacology, offering transformative potential across a wide range of diseases. Given their growing economic importance, the complexity of their b**ehaviour in formulation environments**, where stability, viscosity, aggregation, and long-term storage risks can make or break a drug's viability and directly impact **therapeutic efficacy** and **cost**. Understanding conformational landscapes, aggregation propensities, and interaction profiles under formulation-relevant conditions is essential to overcoming these hurdles[1]. As the pharmaceutical industry moves toward **digital twin technologies**, bio-modelling emerges as a key enabler for predictive formulation science.

In this context, we aim to develop a hybrid modelling strategy that combines established physics-based tools with state-of-the-art AI-based methods. These approaches offer complementary strengths: physics-based simulations provide **atomistic accuracy**, while AI models accelerate **conformational sampling** and **generalize** across biomolecular families. A significant challenge lies in reliably modelling biomolecules under varying conditions (temperature, pH, ionic strength) where small changes can lead to large shifts in behaviour. By targeting key pharmacological properties such as **stability**, **viscosity**, **aggregation**, and **storage resilience**, this project lays the groundwork for a robust and efficient modelling framework that supports formulation decisions and industrial-scale digital twin initiatives.

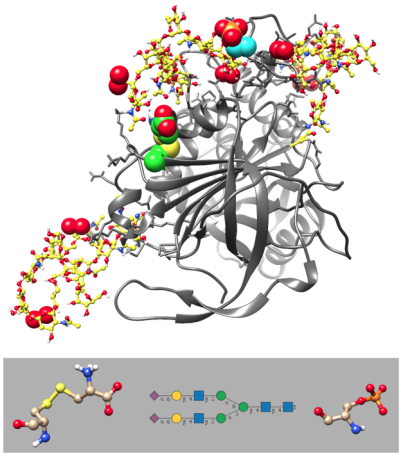


Figure 1: Post-Translational Modifications in Complex Biomolecules. *Structural visualization of a biomolecule with diverse post-translational modifications (PTMs), which critically influence its stability, interactions, and pharmacological behavior. Highlighted PTMs illustrate how chemical variation across sites can alter formulation and storage outcomes.*

References:

[1] Park, E., & Izadi, S. (2024). Molecular surface descriptors to predict antibody developability: sensitivity to parameters, structure models, and conformational sampling. mAbs, 16(1). *https://doi.org/10.1080/19420862.2024.2362788*

(113) Codon usage shapes translation pace and protein folding

Martin Špendl[1], Gregor Gunčar[2]

[1] University of Ljubljana, Faculty of Computer and Information Science, Ljubljana, Slovenia
[2] University of Ljubljana, Faculty of Chemistry and Chemical Technology, Ljubljana, Slovenia

gregor.guncar@fkkt.uni-lj.si

Translation and protein folding are coupled processes: changes in elongation pace influence when nascent segments exit the ribosomal tunnel and begin to acquire structure, making codon usage an important modulator of folding outcomes. Using ribosome-traffic simulations at codon resolution, we estimated local translation speeds across *E. coli* coding sequences and compared these kinetic profiles with protein structures while accounting for the latency imposed by the exit tunnel. Rather than isolated rare codons, we observe short clusters of slower-translated codons and a recurring positional pattern within a window comparable to the ribosome footprint, consistent with a role in coordinating when specific structural elements encounter the cellular folding environment. These features are more prominent in larger, often multidomain proteins with multiple hydrophobic cores. Within such slow regions, synonymous usage tends to favor slower codons for a given amino acid, suggesting that codon choice modulates, rather than overrides, an amino-acid–intrinsic bias in elongation rate. Together, these findings support using translation-pace "pause maps" as gentle constraints in sequence design and codon optimization to preserve native folding schedules without altering amino-acid composition.

(114) Development of Peptidic-Binders for GPCRs

Marija Sorokina[1], Annette Beck-Sickinger [2], Christina Lamers[1], Jens Meiler[1,3,4]

[1] Leipzig University, Medical Faculty, Institute for Drug Discovery, Leipzig, Germany
[2] Leipzig University, Faculty of Life Sciences, Institute of Biochemistry, Leipzig, Germany
[3] Center for Scalable Data Analytics and Artificial Intelligence, Leipzig, Germany
[4] Center for Structural Biology, Vanderbilt University, Nashville, USA

marija.sorokina@medizin.uni-leipzig.de

G protein-coupled receptors (GPCRs) with around 800 members, represent the largest family of human membrane proteins. Located on the cell-membrane they transduce extracellular signals into physiological effects, acting as cell's "control panel". Therefore, GPCRs are implicated in many diseases, such as type 2 diabetes mellitus, obesity, depression, cancer, Alzheimer's disease, *etc*. Signaling pathway of GPCRs is initiated via binding of a ligand to inactive protein conformation. Their endogenous ligands include odors, hormones, neurotransmitters, chemokines, *etc.*, varying from photons, amines, carbohydrates, lipids, peptides to proteins. Ligand binding triggers conformational change and brings GPCR in its active state.

In our research we are aiming to develop peptidic and small-molecule binders that would modulate the GPCR activation. One of the examples is a complement component 5a Receptor 1 (C5aR1) which is a part of the complement system, the innate immune response. Activation of C5aR1 by its ligand, the 74-amino-acid linear peptide anaphylatoxin C5a, triggers recruitment of immune cells, release of pro-inflammatory cytokines, and amplification of the inflammatory response, making C5aR1 a key player in conditions such as rheumatoid arthritis, autoimmune diseases, and neurodegenerative disorders.

A number of antagonists have been developed over the past three decades; however, only one, a small molecule marketed as Tavneos (Avacopan) [1], which targets an allosteric binding pocket outside the transmembrane helices of C5aR1, has been approved for the treatment of a rare autoimmune condition. A few linear peptides, including C089 [2], as well as cyclic peptides such as PMX53 [2], which are derivatives of the natural ligand C5a and target the orthosteric binding pocket, have been developed; however, they failed in clinical studies due to insufficient clinical efficacy and challenges related to oral bioavailability and metabolic stability.

The understanding of the reasons underlying the insufficient efficacy of peptides, as well as the successful development of potent peptidic antagonists, is most likely hampered by a poor understanding of the C5aR1 activation mechanism, which is driven by C5a binding through a three-site binding model. The goal is to develop an extended derivative of an existing C5aR1 peptidic antagonist that mimics the two-site binding of C5a, rather than the one-site binding characteristic of currently available peptidic antagonists [3]. In addition, derivatives of the cyclic peptide PMX53, which contains non-proteinogenic amino acids, will be modeled and synthesized to gain deeper insights into its antagonistic properties and to provide information for the future development of peptidomimetics targeting C5aR1.

References:

[1] D. R. W. Jayne, *et al*., *J Am Soc Nephrol.* 28(9):2756-2767 (2017)
[2] A. K. Wong, *et al*., *Journal of Medicinal Chemistry*, 1, 3417-3425 (1998)
[3] A. C. Dumitru, *et al*., *Commun Biol.* 18;3(1):786 (2020)

(115) Determinants of activity in computationally designed retro-aldol enzymes

Markus Braun[1], Adrian Tripp[1], Gustav Oberdorfer[1]

[1] Graz University of Technology, Institute of Biochemistry, Graz, Austria

markus.braun@tugraz.at

Enzymes power biocatalysis and medicine, yet the scope of applications of natural enzymes is constrained by their native reaction repertoire and by their natural origins. Experimental optimization strategies can broaden the scope of natural enzymes but are resource intensive and not always effective. Computational enzyme design promises access to enzymes that catalyze reactions beyond the scope of their natural counterparts, without the need for expensive experimental optimization. However, computationally designed enzymes struggle to meet industrial demands for activity, hampering adoption of the technology in industry. Despite recent progress on the activity of designed enzymes, computational metrics used to optimize and rank designed enzymes remain largely uncorrelated with experimental activity. Here, we explored origins of the catalytic rates diverging over four orders of magnitude in a set of designed retro aldol enzymes using computational methods. Molecular Dynamics simulations revealed increased flexibility and lacking catalytic positioning in the active sites of the designed retro-aldolases compared to the active site of the enzyme on which the designs were based. Predictions of the active sites in two distinct reaction states revealed misplaced catalytic interactions in several designs and metrics to evaluate the predicted catalytic interactions exhibited stronger correlations with activity than structural similarity metrics such as RMSD. These results support the use of interaction-centric metrics for design optimization and ranking, paving the way towards robust computational enzyme design campaigns.