CONFERENCE BOOKLET

ADVANCES IN PROTEIN FOLDING, EVOLUTION, AND DESIGN 2022

BAYREUTH
6-8 APRIL 2022

Conference Booklet

ABSTRACTS

TABLE OF CONTENTS

1. INVITED TALKS **3**

2. SELECTED TALKS **20**

3. POSTER PRESENTATIONS **35**

*speakers appear underlined in multi-author abstracts

INVITED TALKS

Protein structure prediction in a post-AlphaFold2 world

Mohammed Alquraishi

AlphaFold2 burst on the life sciences stage with the remarkable claim that protein structure prediction has been solved. In this talk I will argue that in some fundamental sense the core scientific problem of static structure prediction is finished, but that further maturation is necessary before AlphaFold2 and similar systems can materially inform biological questions beyond those of structure determination itself. I will outline some of these necessary developments and highlight one in particular: the prediction of structure from individual protein sequences. I will describe present challenges and opportunities, and our efforts to tackle them by combining advances in protein language modeling with end-to-end differentiable structure prediction, presenting new results on the prediction of de novo designed proteins and proteins in the twilight zone of sequence space. Time permitting, I will end by speculating on what abundant availability of structural information might mean for the future of biology.

Computational design of signaling membrane receptors

Patrick Barth

The ability to dynamically switch between distinct conformations and transduce long-range signals represents a hallmark of membrane receptor functions. Understanding the molecular underpinnings of these critical activities remains however challenging as subtle differences in protein sequence often give rise to profound changes in signaling response with no obvious connection to their structure. We developed a computational approach for predicting and designing allosteric signaling functions. Using the method, we designed various G protein-coupled receptors (GPCRs) with novel allosteric properties, potent and selective GPCR-based biosensors and GPCR oligomers with biased signaling activities that agreed well with our predictions. Combining allostery and *de novo* design techniques, we created stabilized functional variants of challenging GPCRs and characterized their structure in active signaling states. To further demonstrate the generality of our approach, we *de novo* assembled and engineered single-pass receptor sensor chimera with potent signaling functions that enhanced CAR-T cell anti-tumor functions. Overall, the studies reveal the existence of distinct classes of allosteric determinants that define an unforeseen molecular mechanism of regulation and evolution of membrane receptor signaling functions and disease-related dysfunctions. Our approach should prove useful for engineering a wide range of biosensors with programmable sensing-signaling behaviors for basic and therapeutic applications.

Designing proteins for assembly on templates: virus-like particles and surface coatings

Renko de Vries

Physical Chemistry and Soft Matter, Wageningen University, Wageningen, the Netherlands

Many interesting natural nanomaterials assemble from proteins plus other components. The other components may be a single biomacromolecules such as a nucleic acids, in the case of viral capsids, or it may be a large biomolecular surface, for example a cell surface in the case of bacterial S-layers. Related to these cases of templated assembly are the many natural proteins that interact with solid surfaces, such as proteins involved in biomineralization or ice-binding proteins. Taking inspiration from these natural examples, in our group, we design proteins that form nanomaterials by co-assembly with other components. In this talk we focus on co-assembly of designed proteins with nucleic acids and with solid surfaces.

First we argue that, as in nature, in order to get the materials properties we want, we may need to construct functional sequences that combine structurally diverse sequence elements: simple repeat motifs inspired by natural structural polypeptides such as silks, elastin, collagen etc, that have simple secondary structures, computationally designed sequences with precisely defined folds, and finally small functional peptide motifs such as solid-binding peptides. We show how each of these types of sequence-blocks can play a role to obtain specific protein nanomaterials properties in multi-block protein designs.

First, for co-assembly with double stranded DNA, we show how rod-shaped virus-like particles may be obtained via cooperative templated assembly for widely differing sequences: on the one hand sequences based on simple repeats of motifs found in natural structural polypeptides such as silks, elastin and collagen, and on the other hand from sequences obtained by computational redesign of natural TALE proteins. We discuss the relative advantages and disadvantages of both types of artificial virus-like particles.

Next, for co-assembly with solid surfaces, we focus on designs for proteins that render solid surfaces such as silica, gold and plastic antifouling and amenable for easy further functionalization, with applications in for example biosensing. During this work we also investigated the relative merits of flexible solid-binding peptides and structured solid binding proteins, and this set us on course to design ice-binding proteins for which we will also present some preliminary results.

Cellular machineries devoted to Rubisco – Nature's predominant CO_2 fixing enzyme.

Manajit Hayer-Hartl

Max Planck Institute of Biochemistry, Department of Cellular Biochemistry, Martinsried, Germany

Photosynthesis is a fundamental process in biology as it converts solar energy into chemical energy and thus, directly or indirectly, fuels all life on earth. The chemical energy is used to fix atmospheric CO_2 and produce reduced carbon compounds in the Calvin-Benson-Bassham cycle. The key enzyme for this process in all photosynthetic organisms is ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco), which is responsible for the conversion of an estimated amount of 10^{11} tons CO_2 per annum into organic material. It is the most abundant enzyme in nature, owing in part to its low catalytic turnover rate and limited specificity for CO_2 versus O_2. Additional complexity comes from the fact that the multistep catalytic reaction of Rubisco is prone to processing errors. As a result, tightly binding 'misfire' products are produced that inhibit catalysis and need to be removed by the AAA+ protein Rubisco activase.

Recent forecasts suggest that global food production will need to rise more than 25 % by 2050 to meet the ever increasing demand. Engineering a catalytically more efficient Rubisco enzyme in higher plants could contribute to reaching that goal. However, the complex nature of Rubisco's folding and assembly pathway has made these efforts exceedingly challenging. In my talk I will review recent progress in understanding the complex chaperone machineries that are necessary for the efficient biogenesis of Rubisco and its functional maintenance.

1. *Aigner, H., Wilson, H.R., Bracher, A., Calisse, L., Bhat, J.Y., Hartl, F.U. and Hayer-Hartl, M. (2017). Plant Rubisco assembly in E. coli with five chloroplast chaperones including BSD2. Science 358(6368), 1272-1278. (doi: 10.1126/science.aap9221).*
2. *Bracher, A., Whitney S.M., Hartl, F.U. and Hayer-Hartl, M. (2017). Biogenesis and metabolic maintenance of Rubisco. Annu. Rev. Plant Biol. 68, 29-60. (doi: 10.1146/annurev-arplant-043015-111633).*
3. *Bhat, J.Y., Miličić, G., Thieulin-Pardo, G., Bracher, A., Maxwell, A., Ciniawsky, S., Mueller-Cajar, O., Engen, J.R., Hartl, F.U., Wendler, P. and Hayer-Hartl, M. (2017). Mechanism of enzyme repair by the AAA+ chaperone Rubisco activase. Mol. Cell 67(5), 744-756. (doi: 10.1016/j.molcel.2017.07.004).*
4. *Wilson, R.H. and Hayer-Hartl, M. (2018). The complex chaperone dependence of Rubisco biogenesis. Biochemistry 57(23), 3210-3216. (doi: 10.1021/acs.biochem.8b00132).*
5. *Hayer-Hartl, M. and Hartl, F.U. (2020). Chaperone machineries of Rubisco – the most abundant enzyme. Trends Biochem. Sci. 45(9), 748-763. (doi: 10.1016/j.tibs.2020.05.001).*
6. *Flecken, M., Wang, H., Popilka, L., Hartl, F.U., Bracher, A. and Hayer-Hartl, M. (2020). Dual role of a Rubisco activase in metabolic repair and carboxysome organization. Cell 183(2), 457-473. (doi: 10.1016/j.cell.2020.09.010).*

Lessons from outside the borders of biological protein sequence space

Klara Hlouchova[1], Vyacheslav Tretyachenko[1], Kosuke Fujishima[2], Alma Carolina Sanchez Rocha[1], Valerio Guido Giacobelli[1]

[1] Department of Cell Biology, Faculty of Science, Charles University, Biocev, Prague, Czech Republic. [2] Earth-Life Science Institute, Tokyo Institute of Technology, Tokyo, 1528550, Japan

Natural proteins represent numerous but tiny structure/function islands in a vast ocean of possible protein sequences, most of which has not been explored by either biological evolution or research. Recent studies have suggested this uncharted sequence space might possess surprisingly high structural propensity. However, the properties of such structures and the factors affecting their occurrence remain enigmatic, hiding important lessons for protein evolution and protein design.

We analyzed two highly combinatorial protein libraries representing proxies of the available sequence space in two different evolutionary periods. The first library was composed of the entire canonical alphabet of 20 amino acids while the second one used a subset of only 10 residues (A,S,D,G,L,I,P,T,E,V) that represent a consensus view of plausibly available amino acids through prebiotic chemistry. Our study shows that compact conformations resistant to proteolysis are abundant in both libraries while the early alphabet is more soluble and refoldable inherently, without the aid of chaperones. We hypothesize that the structure (either in forms of oligomers or globular/molten globule) of early alphabet is assisted by the cell-like environment (such as salts and in/organic cofactors). A parallel bioinformatic analysis of all available cofactor-binding proteins shows that some cofactors can be coordinated solely by early amino acids, supporting this hypothesis. In addition, this is further supported by an exemplary top-down reduction of an RNA-binding protein.

Our work indicates that structured conformations were readily available already to early protein alphabets, possibly capitalizing on interactions that are either absent or extremely rare in today's biology. On the contrary, today's canonical alphabet represents a greater solubility challenge and proteins undergo significant optimization on the path to their viability.

Evolution of specificity and entrenchment at the dawn of Form I Rubiscos

Georg Hochberg

Many protein complexes depend on accessory subunits with elusive functions to assemble into functional oligomers. Such accessory subunits are often evolutionary novelties that elaborate a simpler, ancestral state. The drivers of sophistication and the cause for dependence on novel interaction partners are mostly unknown. A prominent example of a protein that depends on an interaction partner is Form I ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco)[4,5], the key enzyme in the conversion of carbon dioxide into biomass[6,7]. During its evolution, Rubisco started to interact with and depend on a small subunit (SSU), whose function remains enigmatic. Here, we use ancestral sequence reconstruction and biochemical characterizations to elucidate how this obligate interaction evolved. Immediately upon recruitment, the SSU improved Rubisco's catalysis and opened evolutionary paths to increased substrate specificity. Afterwards, a single substitution was able to cause Rubiscos complete dependence on the SSU by causing self-assembly of the octameric Rubisco complex into insoluble fibers in the absence of the SSU. Our work reveals the evolutionary and biochemical causes responsible for both the initial recruitment and subsequent entrenchment of a new protein-protein interaction.

Interplay of hydration, water mobility, and proton transfer in cytochrome C oxidase

Rene F. Gorriz[2], Vincent Stegmaier[2], Marco Reidelbach[2], Senta Volkenandt[1,2], Tahereh Ghane[2], and Petra Imhof[1,2]

[1]Friedrich-Alexander (FAU) Universität Erlangen-Nürnberg, Erlangen, Germany [2]Freie Universität Berlin, Berlin, Germany

Cytochrome c Odixase (CcO) is a protein in the respiratory chain that uses the energy from oxygen reduction to water to pump protons through the membrane. Proton uptake from the inner side of the membrane to the chemical redox centre takes place through two channels, named D or K, after an important Asp or Lys residue, respectively.

Our simulations show that the protonation state of the two channels has an impact on the hydration level within the two channels [1] and of the communication within and between the two channels [2]. For the D-channel, the hydration level is lower when the proton has already reached E286 at the end of the channel as can be explained by the hydrogen-bonded network pointing from E286 to the asparagine gate (formed by N139 and N121), favouring a “closed” conformation [1]. The thus prevented water passage also blocks the most favourable pathway [3] for proton transfer in the D-channel.

In the K-channel, the hydration level depends critically on the position of the excess proton, suggesting that the proton drags its own hydration sphere with it. The probabilities for proton transfer depend on the hydration level and the directionality of the hydrogen-bonded networks which both are coupled to the conformation of K362 [4]. When protonated, K362 is pre-dominantly in an “up” conformation [2] that allows proton transfer only in the upper part of the channel [4], serving as a barrier for passage of an extra proton or back leakage.

The interplay of hydration, water mobilita, and proton transfer in the two channels can thus be regarded as an auto-regulation, allowing proton passage only when required and preventing it, once the proton has reached the upper part of the channel and is therefore close to the redox centre.

[1] T. Ghane, R.F. Gorriz, S. Wrzalek, S. Volkenandt, F. Delatieh, M. Reidelbach, and P. Imhof, J. Membr. Biol. 251 299 (2018).
[2] M. Bagherpoor Helabad, T. Ghane, M. Reidelbach, A.L. Woelke, E.W. Knapp, and P. Imhof, Biophys. J. 113 817 (2017).
[3] M. Reidelbach and P. Imhof, Biochim. Biophys. Acta General Subjects 1864 129614 (2020).
[4] V. Stegmaier, R.F. Gorriz, and P. Imhof, Processes 9 265 (2021).

Computational challenges of graph theory and machine learning to analyze protein structures and properties

Ina Koch

The talk will give insights into former and recent developments to model protein structures as graphs at multiple scales. We consider graph models at amino acid level, at secondary structure level and at quaternary structure level for different requirements in the respective applications, such as prediction, comparison and classification of proteins and their properties. The talk addresses two applications, first, the modularization of protein complexes and their assembly based on complex graphs. For modularization, we applied network theoretic decomposition algorithms. We will introduce the generation of protein complex assembly using hierarchical agglomerative clustering techniques. In the second part of the talk, we consider the influence of protein structural and sequence features on redox modifiability of proteins and specific cysteines therein using statistical and machine learning methods. We will briefly introduce briefly machine learning basics, in particular Extra Trees, Support Vector Machines and Random Forests before we explain their application to redox modifiability.

Interrogating a Protein's Energy Landscape: From Protein Folding to Coronavirus

Susan Marqusee MD/PhD

Depts. of Molecular & Cell Biology and Chemistry UC Berkeley, Berkeley CA

The amino acid sequence of a protein encodes more than the native three-dimensional structure; it encodes the entire energy landscape – an ensemble of conformations whose energetics and dynamics are finely tuned for folding, binding and activity. Small variations in the sequence and environment modulate this landscape and can have effects that range from undetectable to pathological, even when the protein's folded structure is unchanged. I will describe results from different experimental approaches that allow us to peak into these hidden regions of the energy landscape. I will present our work towards an atomic-level comparison of the co-translational folding trajectory (where the ribosome, the rate of synthesis and codon selection has the potential to modulate this process) and in vitro refolding trajectory of a protein using a combination of biophysical approaches and our work using similar approaches to identify a novel conformation of the SARS- CoV-2 Spike protein.

De novo design of novel proteins structures.

Nobuyasu Koga

The structural diversity of proteins underlies their functional variety. We have developed principles for designing protein structures from scratch, based on a theory for protein folding, the consistency principle: Proteins fold into the free energy minimum structures stabilized by consistent local and non-local interactions. The developed design principles describe favorable backbone geometries, lengths of secondary structures and loop torsion patterns, for design target topologies, which indicates the importance of backbone structures for folding, not the detail of amino acid sequences. Using the developed principles, we have succeeded in various protein structures, including novel topologies not observed in nature, with atomic-level accuracy. Moreover, most of the designed proteins exhibited high thermal stability: the melting temperatures were more than 100 °C. The super-stable designed protein structures should be used as scaffolds for engineering functions. Here, I will introduce the developed design principles and showcase the created various proteins structures, with fundamental understandings of protein folding and evolution. Furthermore, I will talk about an application of the designed proteins.

Predicting distal activity-enhancing mutations in tryptophan synthase

Sílvia Osuna[1,2*], Miguel Ángel Maria-Solano[1], Javier Iglesias-Fernández[1]

[1] *Institut de Química Computacional i Catàlisi (IQCC) and Departament de Química, Universitat de Girona, Girona.* [2] *ICREA, Barcelona. *silvia.osuna@udg.edu*

Enzymes exist as an ensemble of conformational states, whose populations can be shifted by substrate binding, allosteric interactions, but also by introducing mutations to their sequence. Tuning the populations of the enzyme conformational states through mutation enables evolution towards novel activity.[1] A common feature observed in many laboratory-evolved enzymes, is the introduction of remote mutations from the catalytic center, which often have a profound effect in the enzyme catalytic activity. [2] As it happens in allosterically regulated enzymes, distal mutations regulate the enzyme activity by stabilizing pre-existing catalytically important conformational states.

In this talk, our new computational tools based on inter-residue correlations from microsecond time-scale Molecular Dynamics (MD) simulations and enhanced sampling techniques are applied in Tryptophan synthase (TrpS) complex. TrpS is composed of TrpA and TrpB subunits, which allosterically activate each other and have no activity when isolated. [3,4] We show how distal mutations introduced in TrpS resuscitate the allosterically-driven conformational regulation and alter the populations and rates of exchange between multiple conformational states, which are essential for the multistep reaction pathway of the enzyme.[3] The exploration of the conformational landscape of TrpS is key for identifying conformationally-relevant amino acid residues of TrpB distal from the active site.[4] We predict positions crucial for shifting the inefficient conformational ensemble of the isolated TrpB to a productive ensemble through intra-subunit allosteric effects. The experimental validation of the new conformationally-driven TrpB design demonstrates its superior stand-alone activity in the absence of TrpA, comparable to those enhancements obtained after multiple rounds of experimental laboratory evolution. Our work evidences that the current challenge of distal active site prediction for enhanced function in computational enzyme design can be ultimately addressed.

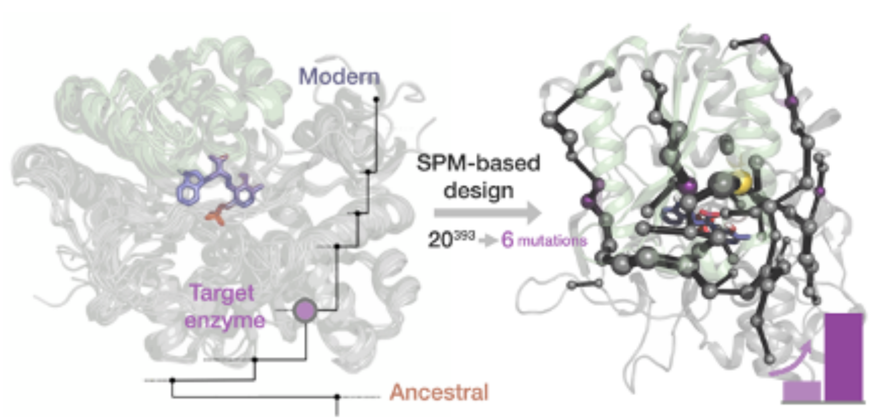


Figure 1. *Scheme of the computational protocol used for the rational design of conformationally-driven stand-alone TrpB variants. Ancestral sequence reconstruction is combined with the correlation-based tool SPM.*

---

[1] Maria-Solano, M. A.; Serrano-Hervás, E.; Romero-Rivera, A.; Iglesias-Fernández, J.; Osuna, S. Role of conformational dynamics for the evolution of novel enzyme function, Chem. Commun. ***2018****, 54, 6622-6634.*
[2] Osuna, S. The challenge of predicting distal active site mutations in computational enzyme design, WIREs Comput Mol Sci. ***2020****, e1502.*
[3] Maria-Solano, M. A.; Iglesias-Fernández, J.; Osuna, S. Deciphering the Allosterically Driven Conformational Ensemble in Tryptophan Synthase Evolution, J. Am. Chem. Soc. ***2019****, 141, 33, 13049-13056.*
[4] Maria-Solano, M. A.; Kinateder, T.; Iglesias-Fernández, J.; Sterner, R.; Osuna, S. In Silico Identification and Experimental Validation of Distal Activity-Enhancing Mutations in Tryptophan Synthase, ACS Catal. ***2021****, 11, 13733-13743.*

Ancestral proteins as scaffolds for enzyme engineering

Jose M. Sanchez-Ruiz

Plausible approximations to sequences of ancestral proteins (i.e., proteins from extict organisms) can be derived from the known sequences of their modern descendants using suitable analyses. The possibility of ancestral sequence reconstruction was proposed by Linus Pauling and Emile Zuckerkandl in 1963. While this was only a theoretical possibility in the mid-twentieth century, ancestral sequence reconstruction has recently become a standard approach in molecular evolution studies, as a result of advances in phylogenetics and bioinformatics, together with the availability of very large sequence databases. In the last 25 years, proteins encoded by reconstructed ancestral sequences (i.e., "resurrected" ancestral proteins, in the common jargon of the field) have proven to be very useful tools to address important problems in molecular evolution. Furthermore, we and others have shown that resurrected ancestral proteins may display properties that are desirable in scaffolds for enzyme engineering. These properties include high stability, as well as substrate and catalytic promiscuity, and will be described in some detail in relation with the generation of *de novo* enzymes, a major unsolved problem in protein engineering. Finally, the unusual properties and biotechnological potential of ancestral TIM-barrels will be discussed.

Can we determine which enzyme designs will be catalytic without experimentation?

Joanna Slusky

I will present the recent development of a machine learning model to distinguish between catalytic and non-catalytic metalloprotein sites.

Functional protein complexes from symmetric designer proteins

Arnout Voet

Katholieke Universiteit Leuven

While symmetric protein assemblies are omnipresent in nature, it is not observed at the monomeric level. However, pseudo-symmetry is frequently observed in tandem-repeating proteins. This originates from duplication and fusion events during the course of protein evolution. This evolutionary process can be computational reverse engineered. As such our lab created several symmetric proteins such as the Pizza, Tako and Cake proteins. While these proteins were originally top-down developed as evolutionary probes, they also served as building blocks for the bottom-up functionalisation leading to self assembling particles, catalytic activity and binding to ligands.

Further functionalisation however was limited as their design did not allow the incorporation of modular/flexible loops which could be used to bind proteins similar as the CDR of antibodies, or form enzymatic substrate binding pockets.

Therefore we computationally designed a novel building block inspired by the KELCH-motif. These Self-Assembling Kelch (SAKe) proteins are highly stable and can be decorated with loops varying in length and amino acid composition. Furthermore by rationnaly introducing metal binding sites the proteins are able to form larger assemblies such as 2D layers or as 3D cages irrespective of the loop composition. These results validate the SAKe proteins as a highly versatile building blocks for applications such as diagnostics, bio-therapeutics and scaffolding or enzymatic cascade scaffolding.

Development and application of computational methods for reliable and accessible protein design

Christopher W Wood

School of Biological Sciences, Univeristy of Edinburgh.

Atomically accurate design of novel protein structures has become a reality over the past few decades[1,2]. The field has advanced from producing structural curiosities to regularly creating proteins with useful functions[3–5]. However, protein design remains the domain of highly specialised research groups, mainly due to the computational expertise and resources required to utilise current design methods, and the high failure rate of designs taken on to experimental characterisation[6].

Our lab focusses on developing user-friendly tools to assist with all stages of computational protein design. We have developed a highly performant algorithm for fixed-backbone sequence design, which requires a fraction of the compute of current physics-based methods[7]. Novel sequences designed by our algorithm accurately fold to the target structures in computational folding experiments, and we are in the process of experimentally validating designed sequences for a range of folds. However, generating sequences is only the start of the design process: designs must be computationally screened to identify those that are likely to have the properties required for a particular function. To this end, we have developed DE-STRESS, a user-friendly web application for evaluating models of designed and engineered proteins[8]. DE-STRESS calculates a range of metrics for each design and provides information to the user to help contextualise the results. We have utilised these metrics to automatically identify designed antibodies that are likely to be expressed and soluble. Ultimately, our aim in developing these tools is to increase the efficiency of the protein-design process, making it more accessible and reliable as a technique.

1. Woolfson, D. N. A Brief History of De Novo Protein Design: Minimal, Rational, and Computational. J. Mol. Biol. 433, 167160 (2021).
2. Pan, X. & Kortemme, T. Recent advances in de novo protein design: Principles, methods, and applications. J. Biol. Chem. 296, (2021).
3. Sesterhenn, F. et al. De novo protein design enables the precise induction of RSV-neutralizing antibodies. Science 368, (2020).
4. Quijano-Rubio, A. et al. De novo design of modular and tunable protein biosensors. Nature 591, 482–487 (2021).
5. Herud-Sikimić, O. et al. A biosensor for the direct visualization of auxin. Nature 1–5 (2021) doi:10.1038/s41586-021-03425-2.
6. Huang, P.-S., Boyken, S. E. & Baker, D. The coming of age of de novo protein design. Nature 537, 320–327 (2016).
7. Castorina, L. V., Petrenas, R., Subr, K. & Wood, C. W. PDBench: Evaluating Computational Methods for Protein Sequence Design. ArXiv210907925 Cs Q-Bio (2021).
8. Stam, M. J. & Wood, C. W. DE-STRESS: a user-friendly web application for the evaluation of protein designs. Protein Eng. Des. Sel. 34, gzab029 (2021)
.

Engineering lanthanide-dependent artificial metalloenzymes

Cathleen Zeymer

Department of Chemistry, Technical University of Munich (TUM), Germany.
cathleen.zeymer@tum.de

The chemical and pharmaceutical industry is under increasing pressure to replace traditional chemical catalysis with sustainable biocatalytic approaches. However, it remains a major challenge to develop novel enzymes for chemical reactions beyond nature's synthetic repertoire. A successful strategy is the development of artificial metalloenzymes, which are designed rationally to combine the catalytic properties of an abiotic metal cofactor with the chiral environment of a protein scaffold that provides stereoselectivity. These systems are genetically encodable and therefore amenable to optimization by directed evolution. This technique mimics natural selection in the laboratory through iterative cycles of mutagenesis and screening. We recently established the formation of specific metal-protein complexes from highly stable, computationally designed protein scaffolds. The metal ions are incorporated by dative anchoring, which exploits direct coordination by natural amino acids of the protein. We now work on turning these *de novo* metalloproteins into a biocatalytic platform for synthetically valuable reactions, such as stereoselective carbon-carbon bond formations. Specifically, we developed a *de novo* TIM barrel scaffold with femtomolar affinity for lanthanides, where metal binding can be observed by sensitizing the element-specific luminescence. Lanthanide ions are not only potent Lewis acid catalysts; they also promote photoredox chemistry. Our ongoing work thus focuses on the development of artificial lanthanide enzymes.

SELECTED TALKS

Protein Fold Diversification

Claudia Alvarez Carreño[1,*], Rohan Gupta[2], Anton Petrov[1,2], and Loren Dean Williams[1,2]

[1]NASA Center for the Origin of Life, Georgia Institute of Technology
[2]School of Chemistry and Biochemistry, Georgia Institute of Technology

Introduction: Emergence of the earliest folding-competent proteins remains a central question in biological sciences. The prospect of a *de novo* discovery of protein domains by randomly sampling the sequence space appears essentially impossible. In ancestral biology, the ribosome was the nursery of protein folding; the earliest coded proteins were polymerized on the ribosome and for the ribosome.

To understand early fold diversification, we studied the 3D architecture of universal ribosomal proteins (rProteins), which include some of the oldest, simplest, and best-preserved folds in biology.

Fold Transformations: Fold change is the result of events at the gene level that change the type or sequential order of secondary structural elements in the protein's 3D structure.

Circular permutation: uL33. The best-known example of fold change is circular permutation. In circularly permuted protein domains, the 3D structure remains essentially unperturbed, but the N-and C-terminal ends are swapped. A circular permutation obscuring universality of ribosomal protein uL33 has been identified (Klein et al. 2004, Kovacs et al. 2018). Circular permutation results from duplication and fusion of an ancestral gene and adaptation and loss of the terminal elements of the tandem repeat.

Fold change: SH3/OB. We also studied the relationship between the two globular domains of uL2: SH3 and OB. We showed that SH3 and OB domains evolved by duplication and adaptation, which is the same mechanism that is responsible for circular permutations (Alvarez-Carreño et al. 2021).

The Domain Repertoire of the Translation System: The universal core of the ribosome contains a diversity of protein folds, many of which are considered to have originated independently. Only eight folds are recurrent in universal rProteins, the remaining 23 folds appear only once in universal rProteins.

Common evolutionary origin of folds with different 3D structure. Beyond the SH3/OB case, we have identified cross-fold sequence similarities between universal rProteins and other extant globular domains in ECOD. These similarities provide information on early fold diversification.

Alvarez-Carreño C et al. (2021) Mol. Biol. Evol., 38, 5134-5143. Klein DJ, Moore PB and Steitz TA (2004) J. Mol. Biol., 340, 141-177. Kovacs NA et al. (2018) J. Mol. Evol., 86, 581–592.

Acknowledgements: *This research was supported by the NASA Astrobiology Program under the NASA Center for the Origin of Life (80NSSC18K1139) and the NASA Postdoctoral Program, administered by Universities Space Research Association under contract with NASA.*

De novo Enzymes by Design and Evolution

H. Adrian Bunzel[a,b], J. L. Ross Anderson[a], Donald Hilvert[c], Vickery L. Arcus[d], Marc W. van der Kamp[a,b], Adrian J. Mulholland[b]

a School of Biochemistry, University of Bristol, UK. b Centre for Computational Chemistry, University of Bristol, UK. [c]Laboratory of Organic Chemistry, ETH Zurich, CH. [d]School of Science, University of Waikato, NZ.

Novel enzymes can be created by computational design and directed evolution. Understanding how evolution improves designed enzymes promises invaluable insights into the fundamental principles of catalysis and will allow deriving design rules to create better enzymes in the future.[1]

We recently employed directed evolution to improve a computationally designed enzyme catalyzing an elementary proton transfer reaction by four orders of magnitude.[2] Here, we present molecular dynamics simulations[3] revealing that evolution resulted in the closure of solvent-exposed loops and better packing of the active site. The enhanced transition state recognition gave rise to a dynamical network involving large parts of the protein, which tightens the protein ensemble in the transition state and results in a negative activation heat capacity signaled by a curved activity-temperature profile. Comparable effects are observed in the evolutionary trajectory of a related designer enzyme in which directed evolution introduced a novel oxyanion stabilizing residue for catalysis.[4] Here, evolution gradually altered the protein conformational ensemble on various timescales to populate catalytically superior states.[5] Transition-state stabilization was perfected by evolutionary fine-tuning of extended backbone conformational changes and meticulous organization of the catalytic residues. Our results have implications for understanding enzyme evolution and suggest that selectively targeting conformational dynamics by design and evolution will expedite the creation of new enzymes.

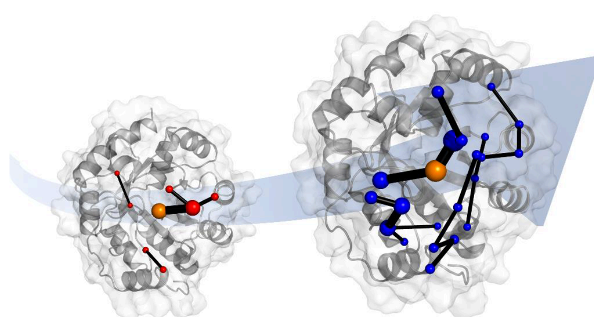


Fig 1: *Emergence of dynamical networks during the evolution of a designer enzyme.*

[1] HA Bunzel, *et al., Curr. Opin. Struct. Biol*. 2021, **67**, 212-218
[2] HA Bunzel, *et al., J. Am. Chem. Soc*. 2019, **141**, 11745-11748.
[3] HA Bunzel et al., Nat. Chem. 2021, ***13****, 1017-1022.*
[4] H Kries, JS Bloch*, HA Bunzel*, et al., ACS Catal. 2020,* ***10****, 4460–4464.*
[5] R Otten, RAP Pádua*, HA Bunzel*, et al., Science 2020,* ***370****, 1442-1446.*

Experimental reconstruction of rapid enzyme evolution: the case of the hydroxyatrazine ethylaminohydrolase AtzB

Markus Busch, Caroline Hiefinger, Reinhard Sterner

Institute of Biophysics and Physical Biochemistry, Regensburg Center for Biochemistry, University of Regensburg, D-93040 Regensburg, Germany

Background: We wish to understand how new enzymes with high substrate specificity and catalytic efficiency have emerged over the course of natural evolution. The underlying processes can be best investigated for enzymes that have evolved only recently. This is most accurately realized in enzymes which have acquired the ability to metabolize anthropogenic (xenobiotic) substances after their relatively recent introduction into the environment, mostly in the form of pesticides and industrial waste. An interesting example is provided by the proteins AtzA, AtzB and AtzC, which decompose the herbicide atrazine and enable their host organisms to use the breakdown products as carbon and nitrogen sources. Similarities with respect to amino acid sequence, three-dimensional structure, and reaction mechanism suggest that the Atz enzymes have evolved by gene duplication and subsequent specialization from different members of the amidohydrolase superfamily for which prominent representatives are guanine or cytosine deaminases. The goal of this work was to identify likely amidohydrolase precursors of the Atz enzymes, and to specifically reconstruct AtzB evolution in the laboratory.

Results: Steady-state kinetics showed that AtzB, in addition to its high native hydroxyatrazine ethylaminohydrolase activity (kcat/KM = 2.4×10^5 $M^{-1}s^{-1}$), has low promiscuous guanine deaminase (GuaD) activity (kcat/KM = 5.6 $M^{-1}s^{-1}$). Importantly, we also found that the two closest known AtzB-homologues $AtzB_{Hom1}$ and $AtzB_{Hom2}$, which previously lacked functional annotation, are natural guanine deaminases (kcat/KM = 1.8×10^3 $M^{-1}s^{-1}$). Based on these findings, $AtzB_{Hom1}$ and $AtzB_{Hom2}$ as well as other close AtzB homologues were used as templates for the improvement of the promiscuous guanine deaminase activity of AtzB. For this purpose, residues determining hydroxyatrazine ethylamine-specificity in AtzB that were identified by sequence logos and mutated to the corresponding guanine specificity-determining residues in a stepwise manner. Eventually, an AtzB variant containing five mutations emerged as the best guanine deaminase, with a decent kcat/KM = 4.1×10^3 $M^{-1}s^{-1}$, corresponding to an approximately 740-fold improvement over wild-type AtzB. At the same time this variant still possesses a residual hydroxyatrazine ethylaminohydrolase activity with a kcat/KM = 2.5×10^2 $M^{-1}s^{-1}$, corresponding to a 980-fold decrease compared to wild-type AtzB. In an inversion of this set of experiments, specificity-determining residues of the guanine deaminases $AtzB_{Hom1}$ and $AtzB_{Hom2}$ were exchanged by the corresponding residues of AtzB. While wild-type $AtzB_{Hom1}$ and $AtzB_{Hom2}$ exhibit only weak hydroxyatrazine ethylaminohydrolase activity (kcat/KM< 65 $M^{-1}s^{-1}$), it could be increased by only two residue exchanges to 8.5×10^2 $M^{-1}s^{-1}$ and 2.6×10^3 $M^{-1}s^{-1}$, respectively.

Significance: The recent emergence of novel metabolic functions in response to the introduction of man-made chemicals into nature showcases the remarkable adaptability of enzymatic systems. Our findings provide insight into mechanisms, constraints, and epistatic effects of this process by generating highly active enzymes with novel substrate specificities from existing enzymes through a limited number of mutations. The results further our understanding of the very recent acquisition of new enzymatic functions which may play a key role in mitigating the ecological effects of anthropogenic environmental toxins.

De novo design of a cell-penetrating high-affinity ligand for microtubule-based transport

Jessica A. Cross[1,2,*], Guto G. Rhys[3,*], William M. Dawson[1*], Harry F. Thompson[1,2], Sooruban Shanmugaratnam[3], Nigel J. Savery[2,4], Mark P. Dodding[2], Birte Höcker[3,*], and Derek N. Woolfson[1,2,4,]

[1]School of Chemistry, University of Bristol, Cantock's Close, Bristol BS8 1TS, UK. [2]School of Biochemistry, Medical Sciences Building, University of Bristol, University Walk, Bristol BS8 1TD, UK. [3]Department of Biochemistry, University of Bayreuth, Universitätsstrasse 30, 95447 Bayreuth, Germany. [4]BrisSynBio, Life Sciences Building, University of Bristol, Bristol, BS8 1TQ, UK

Increasingly, it is possible to design peptide and protein assemblies *de novo* from first principles or computationally. This provides new routes to functional synthetic polypeptides, including designs to target and bind proteins of interest. Protein-protein interactions (PPIs) are central to most, if not all, biological processes. Therefore, disrupting existing PPIs or generating new ones to monitor or intervene in such processes is a major endeavour in peptide and protein design and engineering. Much of this work has been developed *in vitro*. Therefore, a challenge is to deliver *de novo* polypeptides efficiently to sites of action within cells. Cell-penetrating peptides can enter cells *via* direct or passive translocation across biological membranes. Here, we describe the design, characterization, intracellular delivery, and subcellular localisation of a *de novo* synthetic peptide system.

This comprises a dual-function basic peptide (apCC-Di-B), programmed both for cell penetration and target binding, and a complementary acidic peptide (apCC-Di-A), that can be fused to proteins of interest and introduced into cells using synthetic DNA. The designs are characterized *in vitro* using biophysical methods and X-ray crystallography. The basic, arginine-rich peptide is shown to enter mammalian cells and to bind its acidic partner. This is exploited by subcellular targeting of the basic peptide to label specific organelles. Finally, the system is employed to deliver a functional peptide to a subcellular compartment, demonstrated by recruitment of an endogenous motor protein to relocate lysosomes within the cell.

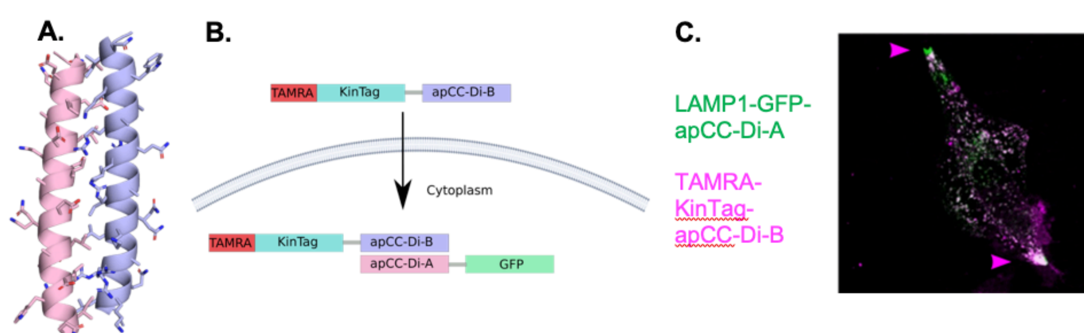


Figure 1. *Design of a de novo cell penetrating peptide to deliver functional cargo to a subcellular compartment. A. X-ray crystal structure of the designed heterodimer apCC-Di-AB. B. Cartoon illustration of apCC-Di-B as a cell penetrating peptide for delivery of a functional peptide, in this case KinTag: a sequence shown to recruit the endogenous motor protein kinesin-1 for organelle transport.1 C. KinTag recruited to lysosomes by apCC-Di-AB dimerization promotes their transport to the cell periphery.*

1 Cross, J. A., Chegkazi, M. S., Steiner, R. A., Woolfson, D. N. & Dodding, M. P. Fragment-linking peptide design yields a high-affinity ligand for microtubule-based transport. Cell Chemical Biology ***28****, 1347-1355.e1345, doi:10.1016/j.chembiol.2021.03.010 (2021)*

Designing stable metalloproteins using deep learning

Simon L. Dürr, Ursula Röthlisberger

*Laboratory of Computational Biochemistry and Chemistry, ISIC, Ecole Polytechnique Federale de Lausanne, Switzerland. * ursula.roethlisberger@epfl.ch*

30 to 40 % of proteins are estimated to depend on at least one metal ion for their biological function[1]. Despite their important biological function the computational design of metalloproteins remains an arduous task due to the inaccuracy of force fields for metals (especially for important transition metals such as zinc) and the computational cost of QM calculations for biologically relevant systems. Some computational successes include the design of a zinc-mediated PPI which due to the inaccuracy of the scoring function exhibited a vacant site and a small cavity close to it and had low esterase activity[2] or our redesign of GB1 which introduced a metal site with one vacancy but dimerized in solution as head-to-tail dimer via the Zn^{2+} binding site[3].

While designing simple binding sites rationally is something that has been achieved for multiple different folds, computational design of functional metallosites with defined first and second shell coordination such as required for enzymatic activity has not been achieved yet[4]. In this work, we present applications of deep learning towards designing stable metalloproteins using 3D convolutional neural networks. Similar networks have been used for predicting binding pockets in proteins[5], for identification of masked residues[6] and for fixed backbone protein design[7]. In this work, we present new results for the prediction of masked residues with metals as an explicit input channel and evaluate the performance with respect to preorganized active sites of metalloenzymes. A comparison with deep mutational scanning results, ΔΔG predictions and Zn^{2+} binding assays shows that the network can filter out false-positive ΔΔG predictions efficiently and captures important hydrogen bond networks inside the active site of natural metalloenzymes and thus might be useful to optimize the second shell around metal active sites. Secondly, we present a network that, given a protein environment, can predict whether the environment contains a zinc site and outputs a probability density over the input space which can be used to maximize the probability of metal binding by varying the environment through rotamer optimization. Several applications for the model related to protein structure prediction, drug discovery and protein design will be presented. Both models are implemented in EVOLVE - a new genetic algorithm based toolbox for protein research and can be used to perform multiobjective optimization to create new metal sites in proteins.

---

(1) Andreini, C.; Bertini, I.; Cavallaro, G.; Holliday, G. L.; Thornton, J. M. Metal Ions in Biological Catalysis: From Enzyme
Databases to General Principles. J. Biol. Inorg. Chem. JBIC Publ. Soc. Biol. Inorg. Chem. 2008, 13 (8), 1205–1218. https://doi.org/10.1007/s00775-008-0404-5.
(2) Der, B. S.; Edwards, D. R.; Kuhlman, B. Catalysis by a De Novo Zinc-Mediated Protein Interface: Implications for Natural
Enzyme Evolution and Rational Enzyme Engineering. Biochemistry 2012, 51 (18), 3933–3940. https://doi.org/10.1021/bi201881p.
(3) Bozkurt, E.; Perez, M. A. S.; Hovius, R.; Browning, N. J.; Rothlisberger, U. Genetic Algorithm Based Design and

Experimental Characterization of a Highly Thermostable Metalloprotein. J Am Chem Soc 2018, 140 (13), 4517–4521. https://doi.org/10.1021/jacs.7b10660.
(4) Guffy, S. L.; Der, B. S.; Kuhlman, B. Probing the Minimal Determinants of Zinc Binding with Computational Protein Design. Protein Eng. Des. Sel. 2016, 29 (8), 327–338. https://doi.org/10.1093/protein/gzw026.
(5) Skalic, M.; Varela-Rial, A.; Jiménez, J.; Martínez-Rosell, G.; De Fabritiis, G. LigVoxel: Inpainting Binding Pockets Using
3D-Convolutional Neural Networks. Bioinformatics 2019, 35 (2), 243–250. https://doi.org/10.1093/bioinformatics/bty583.
(6) Shroff, R.; Cole, A. W.; Diaz, D. J.; Morrow, B. R.; Donnell, I.; Annapareddy, A.; Gollihar, J.; Ellington, A. D.; Thyer, R. Discovery of Novel Gain-of-Function Mutations Guided by Structure-Based Deep Learning. ACS Synth. Biol. 2020. https://doi.org/10.1021/acssynbio.0c00345.
(7) Anand-Achim, N.; Eguchi, R. R.; Mathews, I. I.; Perez, C. P.; Derry, A.; Altman, R. B.; Huang, P.-S. Protein Sequence Design with a Learned Potential. bioRxiv 2021, 2020.01.06.895466. https://doi.org/10.1101/2020.01.06.895466

Cofactor protection as a mechanism for the emergence of primordial enzymes

Luis I. Gutierrez-Rus[1], Valeria A. Risso[1], Gloria Gamiz-Arco[1], Beatriz Ibarra-Molero[1] and Jose M. Sanchez-Ruiz[1]

[1]*Department of Physical Chemistry, University of Granada, 18071 Granada Spain*

Many of the crucial biochemical reactions that sustain life are performed by enzymes that contain cofactors directly involved in their catalytic activities. Most of the LUCA reconstructions reveal proteomes with an abundance of cofactor dependant enzymes [1-3]. Therefore, cofactors are believed to have evolved together with an early prebiotic metabolism playing a key role in the emergence of the first metabolic networks in the origin of life [3-5]. However, it remains unclear which were the evolutionary driving forces that led cofactors to associate with proteins and to the emergence of some primordial catalytic enzymes.

Recently, we described the unexpected binding of the iron protoporphyrin cofactor heme to a resurrected ancestral TIM-barrel protein in a well-defined buried and flexible region of the structure [6]. In this work we show that heme, which is an extremely promiscuous and versatile redox cofactor, provides the ancestral TIM-barrel with a substantial catalytic redox power. Free unbound heme displays peroxidase catalytic activity against the typical phenolic substrate o-dianisidine in the presence of hydrogen peroxide. However, during catalysis the extremely low solubility of the tetrapyrrole structure in aqueous media and the oxidative damage induced by the hydrogen peroxide inactivate the heme cofactor and the catalytic activity is depleted. On the other hand, heme binding to the ancestral TIM-barrel leads to the protection of the cofactor against aggregation and inactivation. Despite lacking a typical peroxidase active site, the heme containing TIM-barrel is able to catalyse peroxidase reactions with a higher number of turnovers than free heme.

Therefore, our results suggest that this protective association between the heme cofactor and the ancestral TIM-barrel protein scaffold may represent a plausible general model of cofactor protection that could have play a key role as a driving force in the emergence of primordial cofactor dependant enzymes.

[1] Ouzounis, C. et al. A minimal estimate for the gene content of the last universal common ancestor—exobiology from a terrestrial perspective. Res Microbiol 157, 57-68 (2006)
[2] Weiss, M. et al. The physiology and habitat of the last universal common ancestor. Nat Microbiol 1, 16116 (2016)
[3] Kirschning, A. The coenzyme/protein pair and the molecular evolution of life. Nat Prod Rep 38, 993-1010 (2021)
[4] Kirschning, A. Coenzymes and their role in the evolution of life. Angew Chem Int Ed 60, 6242-6269 (2021)
[5] Xavier, J. et al. Autocatalytic chemical networks at the origin of metabolism. Proc R Soc B 287, 20192377 (2020)
[6] Gamiz-Arco, G., Gutierrez Rus, L. et al. Heme-binding enables allosteric modulation in an ancient TIM-barrel glycosidase. Nat Commun 12, 380 (2021)

Expanding the repertoire of *de novo* protein assemblies: secretion optimized and polymorphic assemblies

Alena Khmelinskaia[1,2,3], Jing Yang (John) Wang[1,2,4] Neil P. King[1,2]

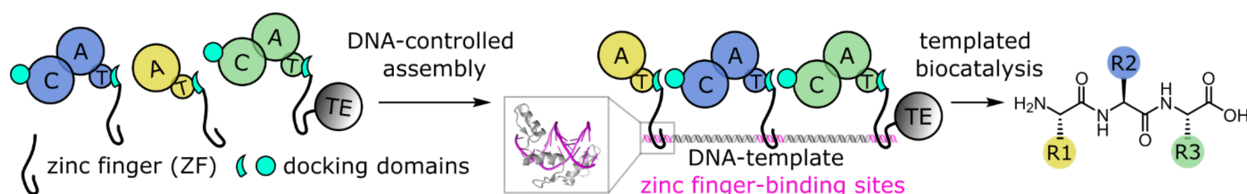
[1]Department of Biochemistry, University of Washington, Seattle, WA, USA. [2]Institute for Protein Design, University of Washington, Seattle, WA, USA. [3]Life and Medical Sciences Institute, University of Bonn, Bonn, Germany. [4]Graduate Program in Molecular and Cellular Biology, University of Washington, Seattle, WA, USA.

Protein assemblies have long been targeted by the nanomaterial community for their various functions in nature, such as shielding macromolecules from the surrounding environment and providing spatial control over biochemical reactions. Recently, computational methods have been developed for designing novel protein assemblies with atomic-level accuracy, yet several aspects of current methods limit the structural and functional space that can be explored. For example, the designed hydrophobic interfaces that are essential for successful assembly are often interpreted by cells as transmembrane segments, resulting in inefficient secretion. Furthermore, the underlying perfect symmetry limits the size and types of architectures that can be designed. Here, I will discuss approaches we have developed to overcome these limitations. First, I will describe the "degreaser", a new computational protocol developed to not only improve secretion of natural proteins and existing protein materials, but also streamline the design of novel SOAPs (secretion optimized assembling proteins). Next, I will discuss how local instabilities introduce structural flexibility in protein scaffolds, breaking symmetry and vastly expanding the repertoire of possible target architectures.

Engineering DNA-templated nonribosomal peptide synthetases

Hajo Kries[1]*

[1]*Leibniz-HKI, Beutenbergstr. 11a, Jena/D* *hajo.Kries@Leibniz-HKI.de



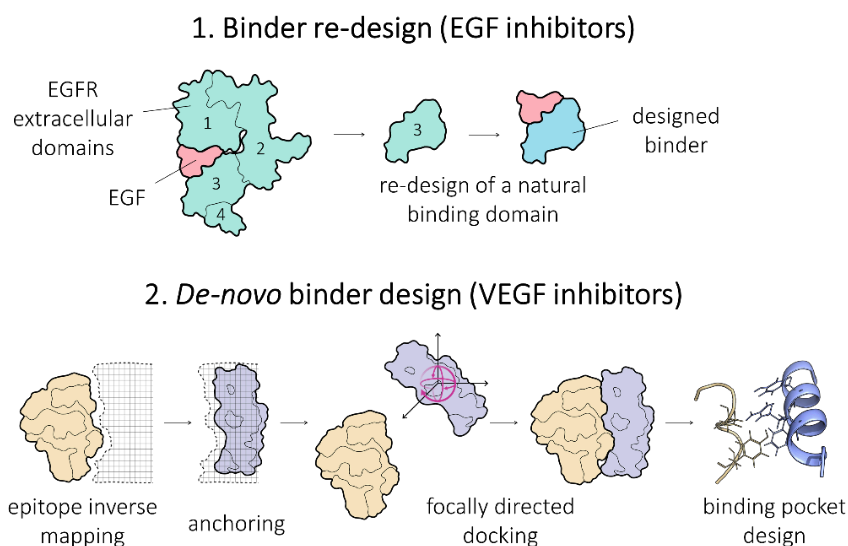
Nonribosomal peptide synthetases (NRPSs) protect microorganisms against environmental threats by producing siderophores or antibiotics, for instance, and are predisposed for biosynthetic engineering because of their modular molecular structure. We have explored several strategies for the redesign of NRPS specificity. Notable examples are the incorporation of a clickable amino acid through targeted binding pocket mutagenesis [1] or specificity transfer through swapping of small protein fragments [2, 3]. Incorporation of clickable amino acids has further enabled a strategy for high-throughput sorting of mutagenized NRPSs displayed on yeast [4]. Here, we demonstrate the addition of DNA templates to nonribosomal peptide synthetases to facilitate NRPS reprogramming [5]. We have split the NRPS for the cyclic decapeptide gramicidin S into modules. Up to four modules were later reassembled on a DNA template using DNA binding domains with high specificity and affinity, and loosely binding intermodular docking domains. The complex nonribosomal machinery showed astonishing tolerance for structural variations when the DNA spacers between modules were altered in length. In the future, DNA programmable NRPS might allow to write the sequences of natural product-like peptides into short DNA templates to speed up NRPS design.

[1] H. Kries et al.; *Angew. Chem. Int. Ed. Engl.* ***2014***, 53, 10105.
[2] H. Kries et al.; Chem. Biol. ***2015****, 22, 640.*
[3] A. Stanisic et al.; Chem. Sci ***2019****, 10, 10395.*
[4] D. L. Niquille et al.; Nat. Chem. ***2018****, 10, 282.*
[5] H.-M. Huang et al.; Cell Chem. Biol. ***2021****, 28, 221-227.e7.*

Protein design of growth factor inhibitors

Kateryna Maksymenko[1], Patrick Müller2, Julia Skokowa3, Andrei Lupas1, Mohammad ElGamacy1,3

[1]Department of Protein Evolution, Max Planck Institute for Biology Tübingen, Germany. [2]University of Konstanz, Germany. [3]Division of Translational Oncology, University Hospital Tübingen, Germany



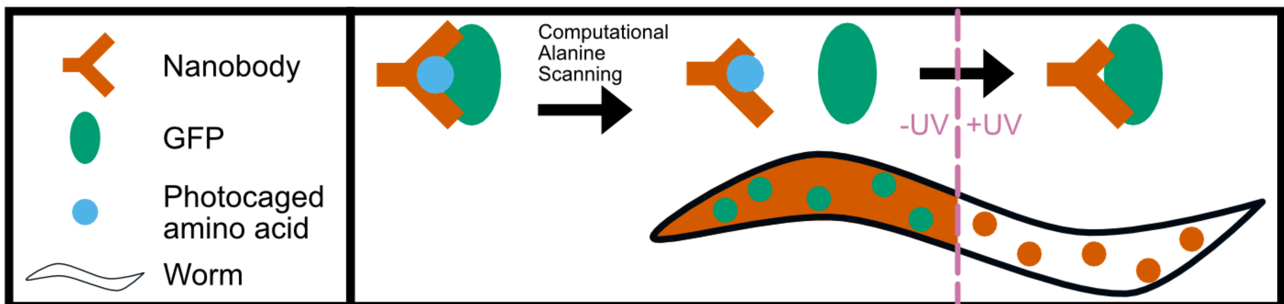
Growth factors are signaling molecules coordinating the complex functionality of multicellular organisms during development and homeostasis. Since aberrant expression of growth factors can cause diverse disorders such as cancer, autoimmune and cardiovascular diseases, growth factors and their receptors are central targets for therapeutic modulation. One of the options to manipulate signaling interactions is to use protein-based binders that are highly specific and able to target various molecular surfaces. Here, we present two different strategies of computational protein design to obtain inhibitors against growth factors which are key modulators of tumor progression. The first approach requires the structure of a native growth factor:growth factor receptor complex and aims to re-engineer a natural binding domain to make it more soluble, more stable, or more affine. In contrast, the second approach relies only on the structure of a target epitope and takes advantage of new software for massive-scale docking of a target site against a protein structure database to select the high shape complementary scaffolds. Adopting the first approach, we designed inhibitors of epidermal growth factor (EGF) using a single domain of EGF receptor as a template. Experimental evaluation of only two designed candidates revealed that both of them are solubly expressed, stable, and bind EGF with nanomolar affinities (i.e. 5-fold stronger than a native domain). Furthemore, we showed that one design inhibits EGF-induced proliferation of epidermoid carcinoma cells with IC50 of 0.5 nM. Using the second strategy, we designed inhibitors of vascular endothelial growth factor (VEGF) based on two different scaffolds. The binding affinities of the designs (16 candidates) to VEGF range from nano- to micromolar levels. X-ray structure determination of one of the candidates showed atomic-level agreement with the design model. Moreover, the best designs showed the ability to inhibit proliferation of VEGF-dependent cells. Thus, our results demonstrate the feasibility of the rational and generalizable approaches to design high-affinity protein binders against predefined conformational motifs.

Engineering photoinducible GFP-binding nanobodies for *Caenorhabditis elegans* using genetic code expansion and computational alanine scanning

Jack M. O'Shea, Angeliki Goutou, Jack Brydon, Cyrus R. Sethna, Christopher W. Wood, Sebastian Greiss.

University of Edinburgh

Pre-Print: https://www.biorxiv.org/content/10.1101/2021.04.16.440193v1



Manipulation of protein-protein interactions can yield key information about protein function *in vivo*. Photoinduction promises the most spatiotemporally precise manipulation possible, so tools for *in vivo* photoinducible protein-protein binding are in high demand. A route to such tools are photocaged amino acids, which have sterically interfering groups that can be removed by 365nm illumination, thus going from a non-binding state to a binding state. Photocaged amino acids can be incorporated into proteins *in vivo* using genetic code expansion. Here, we present a general method for engineering photoinducible nanobody/antigen binding using genetic code expansion and computational alanine scanning, demonstrated in a living animal, the nematode worm *Caenorhabditis elegans*.

To achieve a photoinducible protein-protein interaction, we set out to design photoinducible nanobodies using photocaged amino acids that are active in *C. elegans*. Nanobodies are fragments of camelid heavy chain only antibodies that are increasingly used as *in vivo* tools to bind proteins of interest to control their function. Photoinducible nanobody/antigen binding via photocaged amino acids would add precise spatiotemporal control over existing nanobody based tools used in *C. elegans*.

We used genetic code expansion to introduce, during translation, a photocaged amino acid at a critical site in the nanobody's GFP binding domain. The photocaged nanobody fused with mCherry was expressed alongside nuclear GFP in living *C. elegans*. Thus, a nuclear mCherry signal represents nanobody/GFP binding and a cytoplasmic signal represents no binding. We observed significant GFP

binding by the photocaged nanobody pre-illumination, which was contrary to our expectations based on *in vitro* K_d analysis of the photocaged nanobody.

To design photocaged nanobodies that display no GFP binding pre-illumination and strong binding post-illumination, we used the computational alanine scanning (CAS) program, BUDE Alanine Scan (BAlaS). As input data, we took the final frame of twenty 20ns MD simulations on the crystal structure of the nanobody/GFP complex. We introduced single alanine mutations that BAlaS predicted to reduce GFP binding into the photocaged nanobody, anticipating that this would achieve photoinducible GFP binding. For two out of the three mutant photocaged nanobodies, on testing in *C. elegans* we observed no GFP binding pre-illumination, and significant GFP binding post-illumination.

We engineered a photoinducible nanobody by applying punitive mutations identified by CAS. We repeated this method with a different anti-GFP nanobody, again encountering the same initial failure to break binding and again achieving photoinducible binding via CAS. This work highlights the importance of *in vivo* validation of *in vitro* and *in silico* predictions for applications in complex cellular environments. Our method shows the power of simple, fast computational analysis in tweaking protein designs, and represents a novel application of CAS.

Physicochemical characterization of ancestral triosephosphate isomerases obtained by different reconstruction methods.

Pérez-Niño, Jorge Alejandro[1]; Guerra Borrego, Yasel[1,2], Rodríguez Romero, Adela[3]; Fernández-Velasco, D. Alejandro[1].

[1]Laboratorio de Fisicoquímica e Ingeniería de Proteínas, Facultad de Medicina, Universidad Nacional Autónoma de México; CDMX, México.
[2]Centro de Investigación para la Salud en América Latina, Facultad de Ciencias Exactas y Naturales, Pontificia Universidad Católica del Ecuador; Quito, Ecuador.
[3]Instituto de Química; Universidad Nacional Autónoma de México; CDMX, México.

Ancestral sequence reconstruction (ASR) uses phylogenetic information and evolutionary models to infer ancestral protein sequences that are then synthetized, expressed and characterized in the laboratory. ASR has been used to study the evolution of several facets of proteins, such as substrate specificity and promiscuity, allostery and thermostability. Being a probability-based methodology, there is ambiguity in the amino acid identity of sequences inferred by ASR. The robustness of the phenotype to the uncertainty of the reconstruction process is therefore central to the validity of the conclusions obtained by ASR.

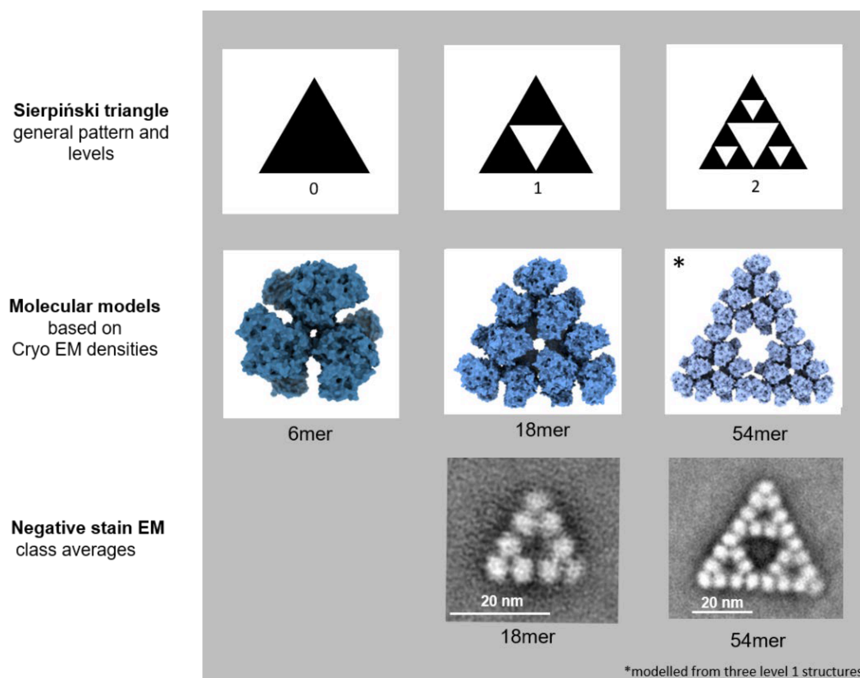
In this work we characterized the function, stability and structure of four different amino acid sequences of the enzyme triosephosphate isomerase (TIM) from a particular eukaryotic node, the last common ancestor of the Opisthokonta lineage. These sequences were inferred using either maximum likelihood (ML) or Bayesian methods as well as the "worst possible case" sequence from each method. The four TIMs were dimeric, all of them showed diffusion-controlled catalytic efficiency. Regarding stability, the TIMs inferred by the ML method showed a melting temperature 10 ^{0}C higher than those obtained from Bayesian inference. Dilution experiments showed a lower dissociation constant for ML TIM. High resolution structures showed that in all cases, the TIM barrel fold of the monomers as well as the association mode between them is very similar. Notably, the interface region is also conserved. Significant differences were only found in the interaction between helices that may explain the observed differences in thermostability and dissociation constant. Our results indicate that the functional and structural properties of TIM were robust to sequence uncertainty whereas stability was not.

Supported by PAPIIT IV200322 and IN210519 and CONACYT fellowship CVU 817384.
The authors acknowledge Isabel Velazquez, Miguel Costas and Jessica Salazar, from Laboratorio de Biofisicoquímica, Facultad de Química, UNAM; and Georgina Espinoza Pérez from Laboratorio Nacional de estructura de Macromoleculas, UNAM

Discovery of a naturally evolved protein fractal

Franziska Sendker[1], Thomas Heimerl[2], Stefan Bohn[3], Jan Schuller[4], Georg Hochberg[1,2,4]

[1]Max Planck Institute for Terrestrial Microbiology, [2]Philipps-University Marburg, [3]Helmholtz Zentrum Munich, [4]Zentrum für Synthetische Mikrobiologie



Fractals are complicated yet fascinating patterns that are self-similar across multiple length scales. Divided into parts they result in a nearly identical reduced–size copy of the whole; one famous example is the Sierpiński triangle. Fractals are prominent in mathematics and science but also frequently employed in aesthetics and art. Engineering fractal assemblies on a molecular level has become a popular while still challenging target e.g. using small molecules that assemble on chiral surfaces. In nature molecular fractals have never been observed.

Here we report the discovery of a citrate synthase that self-assembles into complexes closely resembling the Sierpiński triangle. To understand how these complexes are possible we have solved a cryo EM structure of the first fractal order which is built from 18 monomers: the structure relies on a symmetry violation within hexameric subcomplexes that make up the fundamental building blocks of the fractal pattern. This assembly can be dissociated via a physiological shift in pH or the presence of high concentrations of the enzyme's substrates. We show that the resulting hexameric building blocks are more catalytically efficient compared to the fractal-like complexes, which indicates a form of allosteric regulation associated with the assembly state. Using ancestral sequence reconstruction, we show that this assembly evolved recently from simpler dihedral hexamers and further identified a single historical substitution that drove this process.

Our findings expand the space of known protein geometries to fractal-like complexes of different sizes and imply that these assemblies may be surprisingly easy to evolve

---

1. *Martínez-Martínez, Md., González-Pedrajo, B., Dreyfus, G., Soto-Urzúa, L., and Martínez-Morales, LJ. (2019). Phasin PhaP1 is involved in polyhydroxybutyrate granules morphology and in controlling early biopolymer accumulation in Azospirillum brasilense Sp7. AMB Expr, 9, 155. https://doi.org/10.1186/s13568-019-0876-4*
2. *Ushimaru, K., Motoda, Y., Numata, K., & Tsuge, T. (2014). Phasin proteins activate Aeromonas caviae polyhydroxyalkanoate (PHA) synthase but not Ralstonia eutropha PHA synthase. Applied and environmental microbiology, 80(9), 2867–2873. https://doi.org/10.1128/AEM.04179-13*

POSTER PRESENTATIONS

In silico study of interactions between DNA and phasins of *A. brasilense* Sp7

Yovani Aguilar-Carrillo, Lucía Soto-Urzua, María de los Ángeles Martínez-Martínez and Luis Javier Martínez-Morales*

Centro de Investigaciones en Ciencias Microbiológicas, Instituto de Ciencias, Benemérita Universidad Autónoma de Puebla, Puebla, México.

Introduction. Phasins (PhaP) are amphiphilic proteins that participate in regulation of polyhydroxybutyrate (PHB) granules in terms of their size and number. PHB is a biopolymer of interest as an industrial plastic due to its biodegradability also because it is originated from renewable resources. Martínez *et al*., 2019 reported, using bioinformatic tools, that *A. brasilense* Sp7 contains 6 genes with a Phasin_2 domain. These genes were named PhaP1Abs to PhaP6Abs. Studies on phasins have suggested possible interactions with other Granule-Associated Proteins (GAPs) such as PHB synthase (PhaC), PHB depolymerases (PhaZ) and regulators (PhaR or PhaM), according to the producer microorganism (Ushimaru *et al*.,2014). Although phasins are the main proteins covering the PHB granule there are not much information about phasins of *A. brasilense* Sp7 so, this study will allow the acquisition of new knowledge.

Methodology. 3D models of PhaP4, PhaP5 and PhaP6 of *A. brasilense* Sp7 were obtained in the I-TASSER server using as a template the crystal structure of *Aeromonas hydrophila* PhaPAhy (PDB number 5IP0). Later, molecular docking was performed using HDOCK server to determine the binding sites between the upstream region (233 bp) of *pha*C gene (encoding PHB synthase) and phasins (PhaP4, PhaP5 and PhaP6), analyzed as receptor and ligand respectively. Then BIOVIA Discovery Studio was used to determine the type of interaction between phasins and DNA. Finally, MEME suited was used to determine the promoter regions -10 and -35 and HDOCK models that joined one of these promoter regions were chosen.

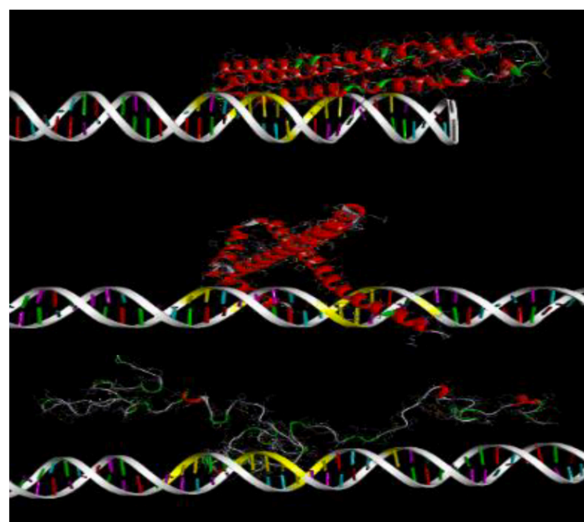


Figure 1. Interactions between phasins and upstream DNAregion of thephaCgene (233bp)by means of BIOVIA Discovery Studio in yellow the regions of the DNA where there is interaction are shown. PhaP4 (Top), PhaP5 (Middle) and PhaP6

Results. The sequences of the promoter regions obtained of *pha*C gene from MEME were ACCGCAC (-10) and CCTAAA (-35). Phasin models of PhaP4, PhaP5 and PhaP6 that showed an interaction with the *pha*C promoter were chosen (Figure 1). PhaP4

interacted with *pha*C promoter near to the region -10. However, PhaP5 and PhaP6 interacted with the box -35. Main type of interactions between DNA and phasins were salt bridge, conventional hydrogen bond, attractive charge and Pi-Alkyl interactions.

Discussion. Only one phasin showed an exact match to the promoter region -35 this was the case of PhaP5 while PhaP4 and PhaP6 joined near of the -10 and -35 regions respectively, this suggests that they could intervene in production of PHB by having the possibility of increasing or decreasing it, given these interactions with DNA during *pha*C transcription, considering the strength of the interactions that are formed between phasins and DNA.

Conclusion. Binding of phasins to *pha*C promoter regions could be a key point in the PHB metabolism of *A. brasilense* Sp7.

Production optimization of heterogeneous proteins employing translation engineering for tuberculosis diagnosis

Angel H. Alvarez-Herrera[1], Abel Gutiérrez-Ortega[1]

[1]Centro de Investigación y Asistencia en Tecnología y diseño del Estado de Jalisco. Av. Normalistas 800. Colinas de la Normal C.P. 44270. Guadalajara, Jal. México.
E-mail: aalvarez@ciatej.mx

Synthetic biology and bioengineering have made it possible biotechnologically to generate new molecules with bioactive characteristics to combat infectious diseases. Cell factories are a player in the sustainable production of biologics, representing hundreds of billions of USD/year in global markets [1]. A common aspect for the development of these cell factories is the continuous search for higher productivity through genetic engineering methods. The E. coli expression system remains a widespread tool for obtaining recombinant heterologous proteins of both bacterial and animal origin through biotechnological improvements [2]. Some gene sequences tend to contain higher cytosine/guanine content than the E. coli genome with up to about 65% G + C [3], contributing to the biosynthesis of proteins with altered sequences or to the formation of inclusion bodies. Rational design of synthetic genes to produce monomeric or hybrid proteins through codon usage optimization allows for protein overexpression in E. coli [4]. In recent years there has been much interest in generating biotechnological products with immunotherapeutic and diagnostic potential. Two improved gene sequences for augmented biosynthesis of biomolecules with importance for the diagnosis of tuberculosis (TB) were obtained by synthetic gene design. Proteins ESAT6 and CFP10 from M. tuberculosis with good antigenic characteristics were produced as a single recombinant fusion protein (rESAT6:CFP10) using E. coli BL21 expression system, as well as the whole molecule of mammal gamma-interferon-gamma (IFN-γ), considered a common biomarker of disease in TB suspects. The rESAT6:CFP10 protein showed a sustained conformational folding and antigenicity in whole blood cell cultures for the diagnosis of TB [5]. Similarly, recombinant IFN-γ was obtained as an abundant soluble protein in bacterial cell lysates, being strongly recognized by specific anti-IFN-γ antibodies of a commercial anti-TB ELISA test, which means that this synthetic molecule would have good physicochemical characteristics as a control reagent for in-vitro diagnostic tests against TB. Thus, gene engineering for optimized translation allows cost-effective increased protein yields of recognized biologicals for TB assessment.

---

[1] A.M. Davy, H.F. Kildegaard, M.R. Andersen. Cell. Syst., 4 (2017) 262-275.
[2] A. Kulshrestha, A. Gupta, N. Verma, S.K. Sharma, A.K. Tyagi, V.K. Chaudhary. Protein Expr. Purif., 44 (2005) 75-85.
[3] S.G.E. Andersson, P.M. Sharp. Microbiology, 142 (1996) 915-925.
[4] G.W. Hatfield, D.A. Roth. 13 (2007) Biotechnol. Ann. Rev., 27-42.
[5] A. Gutiérrez-Ortega, D.A. Moreno, S.A. Ferrari, H. Espinosa-Andrews, E.P. Ortíz, F. Milián-Suazo, A.H. Alvarez. Int. J. Biol. Macrom., 171 (2021) 82-88.

Discovery of a novel GDP-glucose 4-epimerase in archaea reveals an uncommon promiscuity

Carlos Josué Alvarez Quispe, Matthieu Da Costa, Koen Beerens, Tom Desmet

Centre for Synthetic Biology (CSB)-Ghent University

Nucleotide sugar epimerases form a very interesting group of enzymes, as they can invert the configuration of a specific hydroxyl group through a single reaction and without prior activation or protection steps. Within this group, UDP-galactose 4-epimerase (Gal4E, EC 5.1.3.2) is by far one of the best studied members due to its essential role in the Leloir pathway in which it interconverts UDP-galactose and UDP-glucose. Gal4E deficiency is responsible for galactosemia, a hereditary disease, highlighting its vital importance. Although Gal4E was widely studied throughout all domains of life, ranging from eukaryotes to archaea, its biochemical characterization was often limited to UDP-hexoses, neglecting the possibility that Gal4E might be promiscuous towards other NDP-sugars and derivatives thereof. In this study, we identified a novel Gal4E displaying an unprecedented specificity on guanosine diphosphate (GDP) sugars. Indeed, a detailed biochemical investigation performed on Gal4E from *Pyrococcus horikoshii* (phGal4E) revealed that it is a GDP-glucose 4-epimerase. In addition, we confirmed that it accepts a variety of other NDP-sugars including L-sugars moieties, such as GDP-L-Gal/Glc as well as their 6-deoxysugars counterparts GDP-L-fucose and GDP-L-quinovose, respectively.

Keywords
Epimerase, NDP-sugar, Deoxysugar, L-sugar, Discovery

TBvar3D: Mycobacterium tuberculosis resistance variants mapped on protein structures

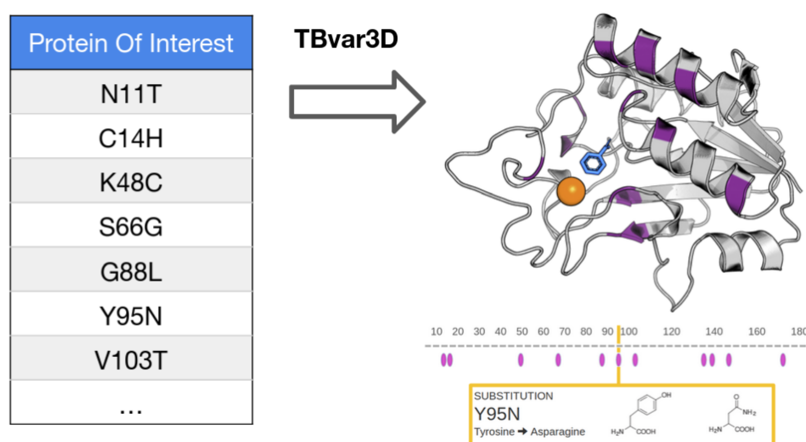
Erblin Asllanaj[1,2], Gabriel Studer[1,2], Andrew Waterhouse[1,2], Xavier Robin[1,2], Gerardo Tauriello[1,2] Rosalba Lepore[3], Torsten Schwede*[1,2]

[1] *Biozentrum, University of Basel, Basel, 4056, Switzerland.*
[2] *SIB Swiss Institute of Bioinformatics, Biozentrum, University of Basel, Basel, 4056, Switzerland.*
[3] *Department of Biomedicine, Molecular Immune Regulation, University of Basel and University Hospital Basel, Basel, 4056, Switzerland*

The antibiotic resistance crisis is one of the greatest worldwide challenges to medical research and society. New tools for reliable analysis and investigation of potential resistance variants are necessary to find new diagnostic markers of resistance and to further our understanding of antibiotic resistance mechanisms.

With TBvar3D, we present a web platform for the analysis of protein variants of *Mycobacterium tuberculosis* (MTB) in the context of protein structure information and antibiotic resistance variants. An easy-to-use interface integrates the extensive WHO MTB resistance variant catalogue with manually curated protein structures which were created with the biological mechanisms of the antibiotic and known mechanisms of resistance in mind. Antibiotic resistance variants can now directly be explored in a structural context and users can submit custom MTB variants to be analysed and visualised by TBvar3D.

The web site is available at https://swissmodel.expasy.org/var3d/.



Experimental characterisation of *de novo* proteins and their unevolved random-sequence counterparts

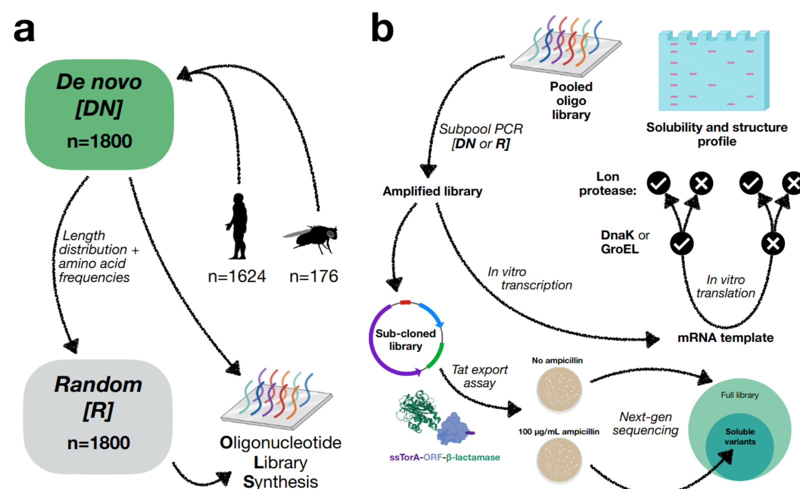
Brennen Heames[1], Filip Buchel[3,4*], Margaux Aubel[1*], Vyacheslav Tretyachenko[3], Andreas Lange[1], Erich Bornberg-Bauer[1,2], Klara Hlouchova[3,5]

[1]*Institute for Evolution and Biodiversity, University of Muenster, Germany.*
[2]*Department of Protein Evolution, MPI for Developmental Biology, Tuebingen, Germany.*
[3]*Department of Cell Biology, Charles University, BIOCEV, Prague, Czech Republic.*
[4]*Department of Biochemistry, Charles University, Prague, Czech Republic.*
[5]*Institute of Organic Chemistry and Biochemistry, Czech Academy of Sciences, Prague, Czech Republic.*
**presenting authors*

De novo gene emergence, increasingly being recognised as a biologically relevant pro- cess, provides a route for new proteins to be formed from previously non-coding DNA. Proteins born in this way are considered as random sequences, and typically assumed to lack defined structure. While it remains unclear how likely a *de novo* protein is to as- sume a soluble and stable tertiary structure, intersecting evidence from random-sequence and *de novo* designed proteins suggests that native-like biophysical properties are abun- dant in sequence space. In addition, laboratory evolution has shown random proteins to be highly evolvable with respect to diverse properties.

Using sets of putative *de novo* proteins previously identified in human and fly, we here characterise a library of these sequences *in vitro* to assess their solubility and structure propensity. Bioinformatic comparison to a library of random proteins with no evolutionary history suggests that *de novo* proteins may have remarkably similar bulk properties (such as secondary structure and disorder content) to unevolved random sequences of a given length and amino acid composition. However, upon expression *in vitro*, *de novo* proteins exhibit higher intrinsic solubility which can be further induced by the DnaK chaperone system.

We therefore suggest that, while random sequence proteins are a useful proxy for *de novo* proteins in terms of structure propensity, *de novo* proteins could be better integrated in the cellular system than their random counterparts. Their greater solubility indicates that *de novo* proteins have been shaped by their brief evolutionary histories, and further implies that *de novo* proteins may be more evolvable than synthetic random sequences.



Irreversibility of thermal unfolding and residual structure in de novo TIM barrels.

Octavio Martín Avilés Sánchez, Daniel Alejandro Fernández Velasco, Laura Dominguez Dueñas & Dulce C. Guzmán.

Universidad Nacional Autónoma de México, Facultad de Química & Facultad de Medicina.

Irreversibility in the in vitro unfolding/refolding is a very common phenomenon in proteins larger than 200 residues such as the $(\beta/\alpha)_8$ TIM barrel. Irreversibility decreases the yield of correctly folded recombinant protein and hampers the determination of thermodynamic parameters such as ΔH, ΔCp and ΔG. It has been reported that residual structure is a strategy followed by bacterial proteins to increase their thermostability, but in some cases, it also makes them more prone to aggregation and/or irreversible temperature-induced unfolding. Recently, a collection of designed TIM barrels (deNovoTIMs) was described (Romero S. et al., 2021). Some deNovoTIMs show reversible temperature-induced unfolding whereas others show irreversibility. In order to determine the role of residual structure in the stability and reversibility of deNovoTIMs, in this work, we analyzed the unfolding behavior of deNovoTIMs using molecular dynamics simulations. It was found that wereas α-helices begin to unfold first (starting with the first alpha helix found at the barrel closure), β-strands maintain their secondary structure throughout the simulation, this is achieved by relying on a hydrophobic network involving aliphatic residues (ILV) that form the core of the protein.

Engineering specific binding pockets for modular peptide binders to generate an alternative for reagent antibodies

Merve Ayyildiz[1], Florian J. Gisdon[1], Josef P. Kynast[1] and Birte Höcker[1 1]

Biochemistry III, University of Bayreuth, Germany

Current biomedical research and diagnostics critically depend on detection agents for specific recognition and quantification of protein molecules. Due to shortcomings of state-of-the-art commercial reagent antibodies like low specificity or cost-efficiency, we aim for creating an alternative recognition system based on a regularised armadillo repeat protein scaffold. Its modular architecture can be exploited for the specific binding of linear epitopes, potentially targeting various peptides and proteins simply by knowing their amino acid sequence [1,2].

To make such a powerful system universally applicable, individual modules must be engineered to recognise specific amino acids. This creates a versatile toolkit to easily assemble specific binders for a wide range of applications. The project is conducted in collaboration with researchers from the University of Zürich and Aston University. We developed a framework to computationally predict, graft and test binding pocket modules that is complemented by library synthesis, directed evolution and experimental screening. Here, we highlight our efforts and progress in computationally engineering interaction sites of designed armadillo repeat modules, and discuss challenges in the implementation for this potentially disruptive technology.

---

[1] Reichen et al. (2014) J. Struct. Biol. 185: 147–162,
[2] Gisdon et al. (2022) Biol Chem, doi: 10.1515/hsz-2021-0384

α-to-π helical transitions are associated to low local flexibility profiles in thermosensitive TRP channels

Daniel Balleza and Elizabeth P. Zamudio

Instituto Tecnológico de Veracruz. Unidad de Investigación y Desarrollo de Alimentos. Av. Miguel Angel de Quevedo 2779, Formando Hogar, 91897 Veracruz, Ver. MEXICO

The transient receptor potential (TRP) superfamily of cation channels is divided into several subfamilies. The TRPV (*vanilloid*) subfamily is composed by proteins which TRPV2 is also activated by heat (>52°C) as well as 2-aminoethoxydiphenyl borate (2-APB). On the other hand, TRPV3 and TRPV4 have been found to be activated by warm temperatures (~34– 38°C and ~27–35°C, respectively). Notwithstanding these differences in their activation temperatures or ligand dependence, the gating mechanisms of these channels exhibit conformational transition pathways in common. In contrast to the mechanism exhibited by Kv channels, where at the middle of segment S6 there is a highly conserved Gly residue and a tandem Pro-kinked motif which fulfill the role of a gating hinge allowing the tilting/swiveling/rotation of the C-terminal S6 segment, these thermo-TRP channels adopt their conductive conformations through α-to-π undergoes a closed-to-open gating transition in response to various physical and chemical stimuli including heat. TRPV1, the prototypical member of these thermo-TRP channels, is activated under physiological conditions by noxious high temperatures (>42°C), and by pungent chemicals including capsaicin - the active ingredient in hot chili peppers. Like TRPV1, π helical transitions in their corresponding S6 segments. In this work, we examine the side-chain flexibility for these segments according to normalized α-to-π B-factors and we describe that these crucial π helical transitions are associated to patterns of local rigidity. In addition, we also studied the tendency of these dynamic regions to be locally disordered, finding that they instead are highly ordered. Finally, both flexibility profiles and disorder predictors show a subtle but clear pattern: whereas TRPV1 and TRPV2 are highly rigid and ordered, TRPV3 and TRPV4 are more flexible and disordered for these sequences at the S6. This correlation is very interesting in terms of the corresponding activation temperatures for such channels and deserve more investigation.

---

[1] Zubcevic L, Lee SY. The role of π-helices in TRP channel gating. Curr Opin Struct Biol. 2019. 58:314-323. doi: 10.1016/j.sbi.2019.06.011.
[2] Romero-Romero S, Gomez Lagunas F, Balleza D. Side-chain flexibility and coupling between the S4-S5 linker and the TRP domain in thermo-sensitive TRP channels: Insights from protein modeling.
[3] Ishida T, Kinoshita K. PrDOS: prediction of disordered protein regions from amino acid sequence. Nucleic Acids Res. 2007. 35(Web Server issue):W460-4. doi: 10.1093/nar/gkm363.

Yersinia pseudotuberculosis OppA interacts with lactate dehydrogenase of rabbit muscle.

Miriam Becerril-Ramírez, Lucía Soto-Urzua, María de los Ángeles Martínez-Martínez and Luis Javier Martínez-Morales*

Centro de Investigaciones en Ciencias Microbiológicas, Instituto de Ciencias. Benemérita Universidad Autónoma de Puebla, Puebla, México.

In the periplasmic space of *Yersinia spp.*, none of the conventional cytoplasmic molecular chaperones such as GroEL and DnaK have been found. However, some periplasmic substrate- binding proteins (SBP), such as *E. coli* MalE (maltose binding protein), have demonstrated to have a chaperone-like function (1). *Yersinia* spp. OppA, a SBP of the OppABCDE operon, has a chaperone-like activity. *Yersinia* spp. OppA has a flexible hinge conformation, which provides a wide range opening angle. It has been proposed that R41 and D42 amino acid residues are directly involved in the chaperone-like activity (2).

It is well known about chaperones complexity, and how deregulation of its mechanism leads to proteotoxicity. However, it has not been understood how this process occurs in the periplasmic space of Gram-negative bacteria. In this study, *in silico* analysis were conducted in order to elucidate the chaperone-like activity of the *Y. pseudotuberculosis* OppA in response to heat stress. The aim of this study was to determine the link affinity with the active site of the enzyme and see the nature of le link.

Methods. Three dimensional structures of *Yersinia pestis* OppA (PDB code: 2Z23) and rabbit muscle Lactate dehydrogenase (LDH) (PDB code 6P6U) were used to perform an *in silico* protein–protein interaction using HDOCK and GRAMM-X servers. Then we selected and extracted 5 ligands with the UCSF-Chimera program. Molecular dockings were made using AutoDock Vina. All non-protein residues were identified and removed through the Chimera 1.7 software, so they did not affect docking calculations. Ligands were extracted from the most hydrophobic areas of LDH. The Avogadro software was used for auto-optimization of ligand bond angles. In Autodock software, solvent molecules were removed from the active site of the enzyme. Hydrogen atoms were added. For the grid box, a general selection of the OppA was made. To select the best ligand, it was considered the relationship between bond stability, which was seen by the number of H bridges and the binding energy (Kcal/mol).

Results. A) The conformation 1 of ligand 3 obtained the highest affinity with the OppA enzyme with a binding energy of -9 (Figure 1). **B**) It has at least 7 hydrogen bonds and forms interactions with 3 of the 4 amino acids residues of interest. **C**) Hydrogen bonds tend to form with Thr, Arg and Ala.

Vina score	Cavity size	Center			Size		
		x	y	z	x	y	z
-9	1945	15	44	5	29	23	23
-6	208	-2	53	22	23	23	23
-5.9	361	8	46	25	23	23	23
-5.9	207	-3	28	-3	23	23	23
-5.2	179	-10	50	8	23	23	23

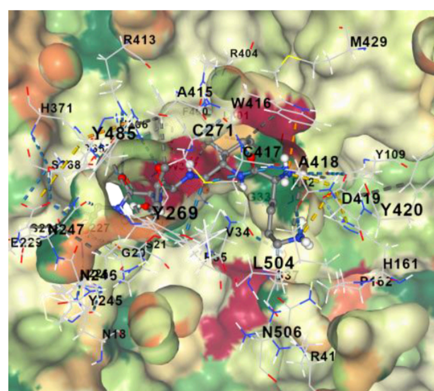


Fig1.- Left: Autodock vina calculations. Best binding energy conformation of ligand 3. Right: How the ligand looks like in its site position within OppA

---

1. *Matsuzaki, M., Kiso, Y., Yamamoto, I., & Satoh, T. (1998). Isolation of a periplasmic molecular chaperone- like protein of Rhodobacter sphaeroides f. sp. denitrificans that are homologous to the dipeptide transport protein DppA of Escherichia coli. Journal of Bacteriology, 180 (10), 2718–2722.*
2. *Escobar Garduño, E., Scior, T., Soto Urzúa, L. et al. (2020) Identification of residues for chaperone-like activity of OppA protein in Yersinia pseudotuberculosis. AMB Expr 10, 153. Springer Open.*

Functionalization of a *de novo* TIM barrel

Julian Beck, Noelia Ferruz, Veronika Waas, Constance Le Gac, Birte Höcker*

University of Bayreuth, Germany

Recent achievements in *de novo* protein design show that it can be increasingly used to address issues in biomedicine and biological engineering. A major challenge hereby is still the *de novo* design of ligand-binding sites, as it requires high accuracy in sampling and scoring. Here we explore options to build a universal phosphate-binding site and introduce it into the scaffold of a *de novo* TIM barrel (Huang et al., 2016; Kordes et al., 2021; Romero-Romero et al., 2021). On the one hand, the *de novo* TIM barrel provides a history-free scaffold and thus a blank slate for our designs. On the other hand, the ability to bind phosphate is present in many natural TIM barrels and thus provides a great example to learn from nature. A large group of TIM barrels share a common binding site for phosphate moieties in their substrates/cofactors/ligands. Therefore, a so-called Standard Phosphate Binding (SPB) motif has been described, built by a short helix and two loops. A database analysis via SCOPe provided us with over 100 different TIM-barrel domains with 30 different phosphate ligands. To define parameters for the ligand binding site, we analysed the angles between the secondary structural elements of the binding-sites and their amino acid sequences. These motifs are now used in our design approach. We start at the sequence level with the insertion of amino acids based on the identified motifs. After filtering PSIPRED predictions for the desired helix pattern (Wiese, Shanmugaratnam and Höcker, 2021), we utilize AlphaFold (Jumper *et al.*, 2021) to reach the structure level. The resulting structures were filtered based on the helix angles of the natural dataset. Subsequently, MD simulations and Rosetta energy calculations for the helix residues were used to identify promising designs for experimental characterization. These designs are currently under ongoing experimental characterization, but initial DSC measurements show stabilization by phosphate, which may indicate phosphate binding. Here we report the binding site analysis and computational workflow of the designs as well as first experimental insights.

Huang, P. S. et al. (2016) 'De novo design of a four-fold symmetric TIM-barrel protein with atomic-level accuracy', Nature Chemical Biology, 12(1), pp. 29–34. doi: 10.1038/nchembio.1966.
Jumper, J. et al. (2021) 'Highly accurate protein structure prediction with AlphaFold', Nature, 596(7873), pp. 583–589. doi: 10.1038/s41586-021-03819-2.
Kordes, S. et al. (2021) ' A newly introduced salt bridge cluster improves structural and biophysical properties of de novo TIM barrels ', Protein Science. doi: 10.1002/pro.4249. Romero-Romero, S. et al. (2021) 'The Stability Landscape of de novo TIM Barrels Explored by a Modular Design Approach', Journal of Molecular Biology, 433(18). doi: 10.1016/j.jmb.2021.167153.
Wiese, J. G., Shanmugaratnam, S. and Höcker, B. (2021) 'Extension of a de novo TIM barrel with a rationally designed secondary structure element', Protein Science, 30(5), pp. 982–989. doi: 10.1002/pro.4064.

Mimicking nature's evolutionary mechanism: design of cobalamin-binding chimeras from different protein folds

Alexander E. Braun,[1†] Johanna-Sophie Koch,[1†] Sergio Romero-Romero,[1] Saacnicteh Toledo-Patino,[1,2] Birte Höcker[1]

[1] *Department of Biochemistry, University of Bayreuth, Germany.*
[2] *Current address: Okinawa Institute of Science and Technology Graduate University, Japan.* [†] *These authors contributed equally to the work and both will present the work.*
Poster correspondence: sergio.romero-romero@uni-bayreuth.de

Proteins are one of the most important biomolecules, performing key cellular functions relevant to our modern life. For this, proteins evolved by reusing stably folded peptides and repurposing functional proteins. Studies in molecular evolution have identified a number of processes such as recombination, mutation, deletion, circular permutation, and duplication as the forces that shape the abundance and diversity of the protein universe. Mimicking some of these evolutionary tricks found in nature and combining them with a rational design approach based on biophysical principles results in protein chimeragenesis as a powerful tool for the creation of new macromolecules.

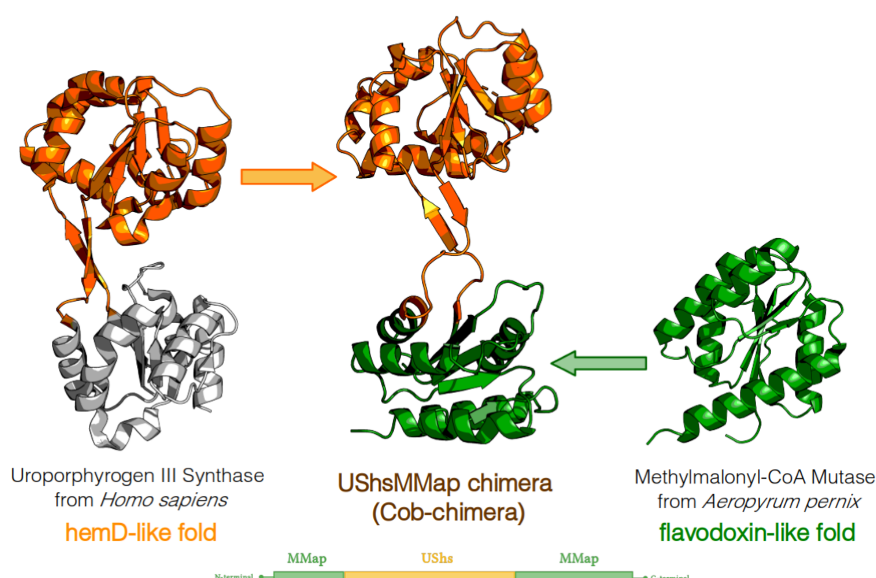


Figure 1. Design of a cobalamin-binding chimera by domain recombination of two distantly related folds.

By swapping sequence elements among related proteins, this approach aims to build up chimeric proteins with new behaviours and functions, paving the way to the construction of new biocatalysts. Inspired by synthetic chemistry, where transition metal catalysts can be utilized for performing more

sophisticated reactions, we aimed to work with cobalamin, a versatile molecule used for a myriad of reactions like methyl transfer, isomerization, ribonucleotide reduction, and dehalogenation. Using a chimeragenesis-based approach, a cobalamin-binding chimera was built by replacing the flavodoxin-like domain in the hemD-like fold of Uroporphyrinogen III synthase from Homo sapiens with the homologous cobalamin-binding domain of Methylmalonyl-CoA-mutase from Aeropyrum pernix.

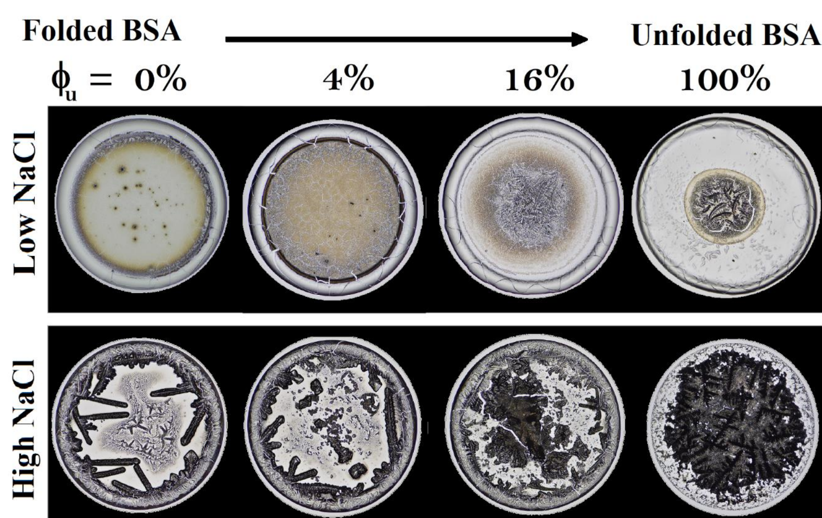
The resulting chimera UShsMMap01 (figure 1), hereinafter called Cob01, was biophysically and structurally characterized, showing a preliminary and weak binding for cobalamin. Using this starting point, a collection of chimeras was engineered and changes implemented based on structural consideration in order to improve the binding properties, yielding the variants Cob02-Cob10. These Cob-chimeras displayed broad changes in their structural and thermodynamic properties, as well as binding parameters. Cobalamin binding for the proteins was proved by the co-elution of the complex in size exclusion chromatography, titration experiments monitored by intrinsic fluorescence, and changes in thermostability followed by differential scanning calorimetry. Furthermore, crystal structures for some variants were obtained in order to elucidate the specific binding mode of the Cob-chimeras with cobalamin. This work provides crucial information towards the design of synthetic enzymes that use cobalamin, the most complex cofactor on Earth.

Pattern formation in dry droplets: evidence of folded and unfolded BSA

Yojana J. P. Carreón[1], O. Díaz-Hernández[2], José M. Saniger[1], and Jorge González-Gutiérrez[2].

[1]Instituto de Ciencias Aplicadas y Tecnología, Universidad Nacional Autónoma de México, Av. Universidad 3000, Coyoacán CDMX 04510, México.
[2]Facultad de Ciencias en Física y Matemáticas Universidad Autónoma de Chiapas, Tuxtla Gutiérrez, Chiapas, México

Protein aggregation is a biophysical phenomenon in which unfolded proteins interact and might form amorphous aggregates. The detection of unfolded conformational states of proteins may help to recognize different diseases. Texture analysis of dried droplets has been shown a useful method to detect different analytes in relevant biological solutions. Here, we carry out texture analysis of dried droplets to reveal unfolded relative concentrations of Bovine Serum Albumin (BSA) in mixed solutions with NaCl. Dried droplets with folded BSA are formed by a well-defined coffee ring and crystal patterns all over the droplet. At low relative concentrations of unfolded protein (above 2%) are produced small amorphous aggregates in the interior of deposits. At high unfolded relative BSA concentrations (above 16%) emerges an "eye-like pattern" surrounded by a uniform coating. Imaging data analysis such as the radial intensity profile, the mean pixel intensity, and entropy allowed us to characterize and differentiate patterns formed at different relative concentrations of unfolded protein. We prove that it is possible to achieve 100% accuracy in identifying 4% of unfolded BSA contained in a binary mixture of proteins in saline solution.



Delay in the aggregation of crystallin gamma S by inhibitors

Castañeda Gutierrez Kimberly*; Fernández Silva Arline; Ramírez Bello Vanesa; Amero Tello Carlos

Laboratorio de bioquímica y resonancia magnética nuclear, Centro de Investigaciones Químicas, Instituto de Investigación en ciencias básicas y aplicadas. Universidad Autónoma del Estado de Morelos, México. kimberly.castanedagut@uaem.edu.mx

Cataracts are defined as the partial or total opacity of the crystallin lens, which gradually causes vision impairment. This pathology is among the leading causes of blindness worldwide and mainly affects the population over 65 years. So far, the only existing treatment is surgical removal of the lens, which is an invasive and expensive process.

The human eye has some proteins called crystallins. These make up 90% of the total proteins in the lens and these are divided into 2 families, the α and the βɣ. Within the βɣ family is the human gamma S crystallin (HγS) which makes up approximately 9% of the total crystallins. It is found in the cortex of the lens and its function is structural.

Cataracts are caused by aggregation of these proteins. And because the crystallin lacks the necessary machinery for protein synthesis and repair, this damage accumulates. The destabilization in the native conformation of these proteins is due to modifications in the amino acids that compose them caused by various factors. Among the external factors that produce these changes is ultraviolet radiation (UV).

In this work we study the effect of exposure to UVB radiation for short periods of time and the ability of 5 molecules to delay the aggregation of crystallin Gamma S. UV visible light absorption, Fluorescence and dynamic light scattering techniques were used to characterize this effect.

TIMED & PDBench: Developing and Evaluating Computational Methods for Protein Sequence Design

Leonardo V. Castorina, Rokas Petrenas, Kartic Subr and Christopher W. Wood

University of Edinburgh
Pre-print: https://arxiv.org/abs/2109.07925

Proteins perform critical processes in all living systems: converting solar energy into chemical energy, replicating DNA, as the basis of highly performant materials, sensing and much more. While an incredible range of functionality has been sampled in nature, it accounts for a tiny fraction of the possible protein universe. If we could tap into this pool of unexplored protein structures, we could search for novel proteins with useful properties that we could apply to tackle the environmental and medical challenges facing humanity. This is the purpose of *de novo* protein design.

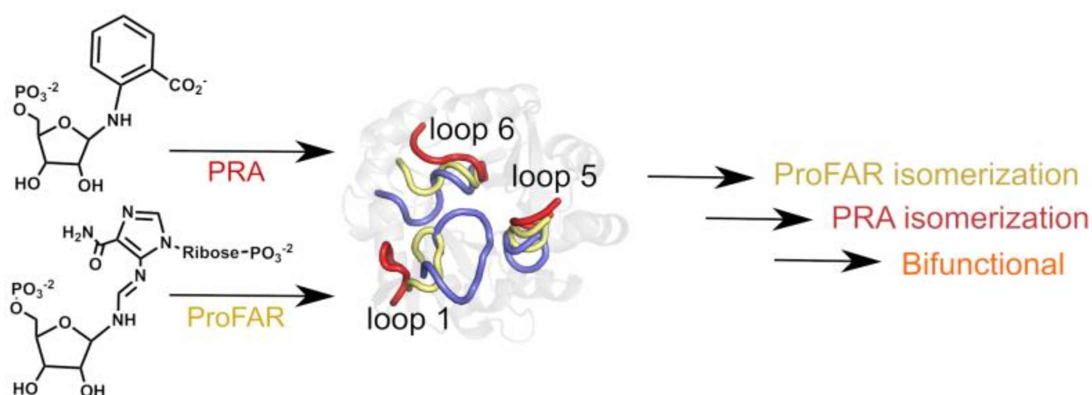
Protein sequence design is an important aspect of *de novo* protein design, and aims at identifying a sequence of residues required to obtain a target 3D shape. Recently, many successful methods to do this have been developed, including deep-learning methods, that frame it as a classification problem and have demonstrated high performance. Beyond their reported improvement in performance, their primary advantage over physics-based methods is that the computational burden is shifted from the user to the developers, thereby increasing accessibility to the design method. Despite this trend, the tools for assessment and comparison of such models remain quite generic. The goal of this paper is to both address the timely problem of evaluation and to shine a spotlight, within the Machine Learning community, on specific assessment criteria that will accelerate impact.

Here, we release the first open source tools for the 3D voxelisation of protein structures and for protein sequence design using Convolutional Neural Networks. We also present PDBench, a carefully curated benchmark set of proteins and collection of standard tests to assess the performance of deep learning based methods. Our robust benchmark provides biological insight into the behaviour of sequence-design methods, which is essential for evaluating their performance and practical utility. We compare five existing models with two novel models for protein sequence design. Finally, we test the designs produced by these models with AlphaFold2, a state-of-the-art structure-prediction algorithm, to determine whether they are likely to fold into the intended 3-Dimensional shapes.

Complex Loop Dynamics Underpin Activity, Specificity and Evolvability in the $(\beta\alpha)_8$ Barrel Enzymes of Histidine and Tryptophan Biosynthesis

Marina Corbella,[1,] Adrian Romero-Rivera,[1,] Antonietta Parracino,[1] Wayne M. Patrick[2] and Shina Caroline Lynn Kamerlin[1]

1. *Department of Chemistry – BMC, Uppsala University, BMC Box 576, S-751 23 Uppsala, Sweden.*
2. *2. Centre for Biodiscovery, School of Biological Sciences, Victoria University of Wellington, Wellington 6012, New Zealand.*



Enzymes are conformationally dynamic, and their dynamical properties play an important role in regulating their specificity and evolvability. In this context, substantial attention has been paid to the role of ligand-gated conformational changes in enzyme catalysis; however, such studies have focused on tremendously proficient enzymes such as triosephosphate isomerase and orotidine 5′-monophosphate decarboxylase, where the rapid (µs timescale) motion of a single loop dominates the transition between catalytically inactive and active conformations. In contrast, the $(\beta\alpha)8$-barrels of tryptophan and histidine biosynthesis, such as the specialist isomerase enzymes HisA and TrpF, and the bifunctional isomerase PriA, are decorated by multiple long loops that undergo conformational transitions on the ms (or slower) timescale.

Studying the interdependent motions of multiple slow loops, and their role in catalysis, poses a significant computational challenge. This work combines conventional and enhanced molecular dynamics simulations with empirical valence bond simulations to provide rich detail of the conformational behavior of the catalytic loops in HisA, PriA and TrpF, and the role of their plasticity in

facilitating bifunctionality in PriA and evolved HisA variants. In addition, we demonstrate that, similar to other enzymes activated by ligand-gated conformational changes, loops 3 and 4 of HisA and PriA act as gripper loops, facilitating the isomerization of the large bulky substrate ProFAR, albeit now on much slower timescales. This hints at convergent evolution on these different $(\beta\alpha)_8$-barrel scaffolds. Finally, our work highlights the potential of engineering loop dynamics as a powerful tool to artificially manipulate the diverse catalytic repertoire of TIM-barrel proteins.

Unravelling the Roles of Loop dynamics and Chemistry in Protein Tyrosine Phosphatase catalysis

Protein tyrosine phosphatases (PTPs) are a family of cellular signaling enzymes whose dysfunction has been implicated in many human diseases, including various cancers, diabetes, and obesity. All studied members of the PTP family share a conserved mechanism, similar transition states and highly conserved active sites. However, their catalytic rates vary by several orders of magnitude and are strictly regulated by the host organisms' requirements. Unlike the very large majority of enzymes, these rate differences appear to be driven by differences in the motion of a large loop (named the WPD-loop), which contains a catalytically essential aspartic acid. Taken together, this would suggest that evolution has found a way to modulate different PTPs catalytic rates through fine tuning the loop dynamics of each enzyme.

Our work provides a detailed computational study on the loop dynamics of human protein tyrosine phosphatase 1B (PTP1B) and YopH from Yersinia pestis, two well characterized PTPs whose loop dynamics are known experimentally to be highly coupled to their catalytic rates. Our study combines detailed structural analysis, enhanced sampling molecular dynamics simulations of their loop dynamics, and empirical valence bond simulations of the chemical step of catalysis. We provide deep insight into how loop dynamics and chemistry regulate these enzymes observed catalytic rates and identify the key residues and allosteric pathways responsible for modulating the catalytic rate as well as the conformations of the WPD-loop. Curiously, our simulations of wild-type YopH further identify a catalytically incompetent hyper-open conformation for its WPD-loop, sampled as a rare event. This has previously only been experimentally observed in YopH-based chimeras. The impact of different loop sequences on the WPD-loop and its neighboring loops on the modulation of loop dynamics, as revealed in this work, may provide a facile means for the family of PTP enzymes to respond to environmental changes and regulate their catalytic activities.

We have since further built on this work through combined computational/experimental studies to generate point mutants and chimeras of the WPD-loop, [2-3] enabling us to unravel in greater detail how different residues and sections of the loop work in unison to modulate the WPD-loop dynamics and ultimately the catalytic rate.

1. ***R. M. Crean*** *et al., J. Am. Chem. Soc.* ***2021****, 143, 10, 3830–45.*
2. *R. Shen,* ***R. M. Crean****, et al., JACS Au* ***2021****, 1, 5, 646–659.*
3. *R. Shen,* ***R. M. Crean****, et al., ChemRxiv.* ***2021****, DOI: 10.26434/chemrxiv-2021-8f6mc.*

De novo design of a cell-penetrating high-affinity ligand for microtubule-based transport

Jessica A. Cross[1,2,], Guto G. Rhys[3,*], William M. Dawson[1*], Harry F. Thompson[1,2], Sooruban Shanmugaratnam[3], Nigel J. Savery[2,4], Mark P. Dodding[2], Birte Höcker[3,*], and Derek N. Woolfson[1,2,4,]

See p. 23 (selected speaker)

Structural resolution of switchable states of a *de novo* peptide assembly

William M. Dawson[1*], Eric J. M. Lang[1,2*], Guto G. Rhys[1,3*], Kathryn L. Shelley[1,4*], Christopher Williams[1,2], R. Leo Brady4, Matthew P. Crump[1,2], Adrian J. Mulholland[1,2] & Derek N. Woolfson[1,2,4]

1. School of Chemistry, University of Bristol, Bristol, UK.
2. BrisSynBio, University of Bristol, Bristol, UK.
3. Department of Biochemistry, University of Bayreuth, Bayreuth, Germany.
4. School of Biochemistry, University of Bristol, Bristol, UK.
** Contributed equally*

The *de novo* design of dynamic multi-state proteins has been a long-standing challenge in protein design. Whilst peptide and protein switches triggered by changes in pH and metal binding have been designed, none of these systems have high- resolution data of both states under the same condition.

Here we describe the successful de novo design and structural resolution of a coiled-coil peptide that assembles into multiple, distinct conformational states.[1] Usually in coiled-coil designs, amphipathic a helices associate to form compact structures with consolidated hydrophobic cores. However, recent rational and computational designs have delivered open a-helical barrels with functionalisable cavities. By placing glycine judiciously in the helical interfaces of an a-helical barrel, we obtain both open and compact states in a single protein crystal (Figure 1). Molecular dynamics simulations indicate a free-energy landscape with multiple and interconverting states. Together, these findings suggest a frustrated system in which steric interactions that maintain the open barrel and the hydrophobic effect that drives complete collapse are traded- off. Indeed, addition of a hydrophobic co-solvent that can bind within the barrel affects the switch between the states both in silico and experimentally.

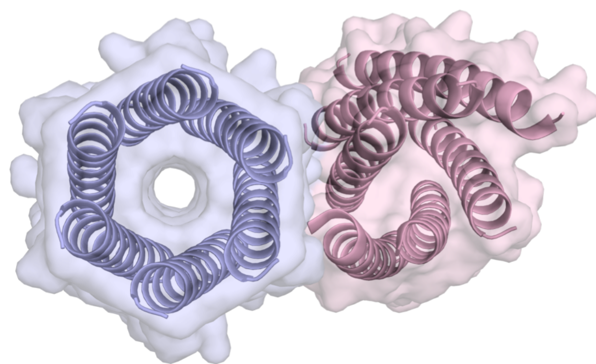


Figure 1. X-ray crystal structure of two states of a switchable peptide. This de novo designed peptide can access an open, a-helical barrel structure (left) or a collapsed, a-helical bundle (right). PDB: 6ZT1.

Our findings should help pave the way to designing allosteric synthetic systems that emulate these natural proteins. In turn, this could allow responsive binding, catalysis and sensing to be built into *de novo* protein frameworks.

---

1. Dawson W.M., Lang E.J.M., Rhys G.G., Shelley K.L., Williams C., Brady R.L., Crump M.P., Mulholland A.J. & Woolfson D.N., 2021, Nature Communications, 12, 1530

Prevalent association to the bacterial cell envelope of prokaryotic expansins revealed by bioinformatics analysis

Andrés de Sandozequi, Juan José Salazar-Cortés, Irán Tapia-Vázquez and Claudia Martínez- Anaya*

Departamento de Ingeniería Celular y Biocatálisis. Instituto de Biotecnología, UNAM. Av. Universidad 2001. Chamilpa. 62210, Cuernavaca, Morelos. Mexico

Expansins are a group of proteins from diverse organisms from bacteria to plants. Although expansins show structural conservation, their biological roles seem to differ among kingdoms. In plants these proteins remodel the cell wall during plant growth and other processes. Contrarily, determination of bacterial expansin activity has proven difficult, although genetic evidence of bacterial mutants indicates that expansins participate in bacteria-plant interactions. Nevertheless, a large proportion of expansin genes are found in the genomes of free-living bacteria, suggesting roles that are independent of the interaction with living plants. Here we analysed all available sequences of prokaryotic expansins for correlations between surface electric charge, extra protein modules, and sequence motifs for association with the bacteria exterior after export. Additionally, information on the fate of protein after translocation across the membrane also point to bacterial cell association of expansins through six different mechanisms, such as attachment of a lipid molecule for membrane anchoring in diderm species or covalent linking to the peptidoglycan layer in monoderms such as the Bacilliales. Finally, we analysed expansin Exl1 from *Pectobacterium brasiliense*, a representative of an expansin without sequence information for cell envelope localization and found that it does bind *P. brasiliense* and *Escherichia coli* cells (both diderm species). Our results have implications for expansin function in the context of bacteria-plant interactions, but also for free-living species in which expansins might affect cell-cell or cell- substrate interaction properties and indicate the need to re-examine the roles currently considered for these proteins.

Insights into distribution and cataloguing of archaeal PPIase

Anchal*, Vineeta Kaushik and Dr. Manisha Goel

Department of Biophysics, University of Delhi, South Campus, New Delhi.
**Email id:* deshwal.anchal@gmail.com

Cis–trans isomerization of peptidyl-prolyl bonds is one of the rate-limiting steps in protein folding. Peptidyl prolyl cis–trans isomerases (PPIases, EC 5.2.1.8) catalyse this step and thereby accelerate protein folding. Currently known PPIases are classified into three families: Cyclophilins, FK506-binding proteins (FKBPs) and Parvulins. PPIase function has been studied in eukaryotes and bacteria, but very little is known about their role in archaeal organisms. Eukaryotes and bacteria often have multiple copies of various PPIase homologs e.g. *Saccharomyces cerevisiae* have 8 cyclophilins, 4 FKBPs, and 2 parvulins copies whereas in humans have 18 cyclophilins and 16 FKBPs copies.

The current study is an attempt to understand the mechanistic working of the "proteostasis" machinery found in archaeal organisms vs. those from other two domains (eukaryotes and prokaryotes). We compiled the distribution of various PPIase homologs in the five archaeal phyla. Subsequently we studied the phylogenetic evolution, conservation of functional amino acids residues, and structural variation in native and substrate/inhibitor bound states of these proteins to throw light on their function.

Fragment and Binding Mode Conservation with the Fuzzle database: A Case Study with Periplasmic Binding Proteins.

Surbhi Dhingra,† Julian Krumm,† Noelia Ferruz,† Florian Michel, Birte Höcker

Department of Biochemistry, University of Bayreuth, Bayreuth 95447, Germany.
†These authors contributed equally to the work and will present the work.
Poster communication: surbhi.dhingra@uni-bayreuth.de / noelia.ferruz-capapey@uni-bayreuth.de

In recent years, protein evolutionary studies have shifted from protein domains considered as the basic evolutionary units to sub-domain sized fragments as the basis of diverse modern protein architecture. A number of studies have identified these conserved fragments in the form of ancestral peptide fragments (Alva et al., 2015, Kolodny et al., 2021). Similarly, in 2020, Ferruz et al. (2020) documented more than 1000 of these sub-domain sized fragments in a database called Fuzzle, which in its newest version incorporated ligand information (Ferruz et al., 2021) (https://fuzzle.uni-bayreuth.de/2.0). The data available in this resource facilitates identification and analysis of evolutionarily conserved sections in proteins, while construction of chimeric proteins by re-assembling two of these fragments can be accomplished with the Protlego software (Ferruz et al., 2021).

To study the conservation of ligand binding modes within conserved fragments, we have used the above-mentioned resources and applied them to the subset of Periplasmic Binding Proteins (PBPs). Earlier research has suggested that PBPs evolved from the Flavodoxin-like fold (c.23.1) (Fukami-Kobayashi et al., 1999). In this study we highlight the extent of these evolutionary relationships while shedding light onto more such SCOP domains that share weak sequence links with the PBPs. We also looked into the conservation of binding modes within the shared fragment space of PBPs. This study highlights the ease to study these evolutionary relationships and lays the groundwork for a large-scale evolutionarily analysis of binding-mode conservation.

1. *Alva, V., Söding, J., & Lupas, A. N. (2015). A vocabulary of ancient peptides at the origin of folded proteins. eLife, 4, 1-19. https://elifesciences.org/articles/09410*
2. *Ferruz, N., Lobos, F., Lemm, D., Toledo-Patino, S., Farías-Rico, J. A., Schmidt, S., & Höcker, B. (2020). Identification and Analysis of Natural Building Blocks for Evolution-Guided Fragment-Based Protein Design. J Mol Bio, 13, 3898-3914. https://pubmed.ncbi.nlm.nih.gov/32330481/*
3. *Ferruz, N., Michel, F., Lobos, F., Schmidt, S., & Höcker, B. (2021, August 18). Fuzzle 2.0: Ligand Binding in Natural Protein Building Blocks. Front. Mol. Biosci. https://www.frontiersin.org/articles/10.3389/fmolb.2021.715972/full#B7*
4. *Ferruz, N., Noske, J., & Höcker, B. (2021, April 26). Protlego: A Python package for the analysis and design of chimeric proteins. Bioinformatics. https://pubmed.ncbi.nlm.nih.gov/33901273/*
5. *Fukami-Kobayashi, K., Tateno, Y., & Nishikawa, K. (1999, Feb 12). Domain dislocation: a change of core structure in periplasmic binding proteins in their evolutionary history. J Mol Biol. https://pubmed.ncbi.nlm.nih.gov/9931266/*
6. *Kolodny, R., Nepomnyachiy, S., Tawfik, D. S., & Ben-Tal, N. (2021). Bridging Themes: Short Protein Segments Found in Different Architectures. Molecular Biology and Evolution, 38(6), 2191–2208. https://academic.oup.com/mbe/article/38/6/2191/6120801?login=true*

Design and characterization of NanECheY chimeras by fragment recombination using Fuzzle and Protlego

Surbhi Dhingra,† Sergio Romero-Romero,† Lucas Carnell, Sabrina Wischt, Birte Höcker

Department of Biochemistry, University of Bayreuth, Bayreuth 95447, Germany.
†These authors contributed equally to the work and both will present the work.
surbhi.dhingra@uni-bayreuth.de / sergio.romero-romero@uni-bayreuth.de

An impressive number of modern proteins have been generated through mechanisms such as recombination, duplication, and accretion of smaller protein fragments. These subdomain-sized fragments are believed to be peptidal ancestors of the proteins that exist today (Romero-Romero et al., 2021). In an effort to locate all fragments shared within the known protein sequence space, a database called Fuzzle (Ferruz et al., 2020) was developed that maps these evolutionary links within the dimensions of the SCOP domain classification. Given this data, a tool for chimeric protein design called Protlego (Ferruz et al., 2021) was then proposed that uses these evolutionary links to construct new proteins. Protlego is a high-throughput chimera design tool that uses fragment recombination as a method for generating new proteins.

Here, we would like to shed some light on high-throughput chimeric protein design and engineering using one of the chimeric proteins constructed with the tools Fuzzle and Protlego. The chimera, NanECheY, is formed from the recombination of protein domains coming from SCOP superfamily c.1 and c.23. The model is an outcome of a single recombination event between the proteins NanE and CheY. Its in silico analysis showed a canonical $(\beta/\alpha)_8$ structure. NanECheY chimeras were experimentally characterized by biophysical and structural methods, giving extra information about the determinants that modulate successful fragment recombination towards new stable and functional proteins.

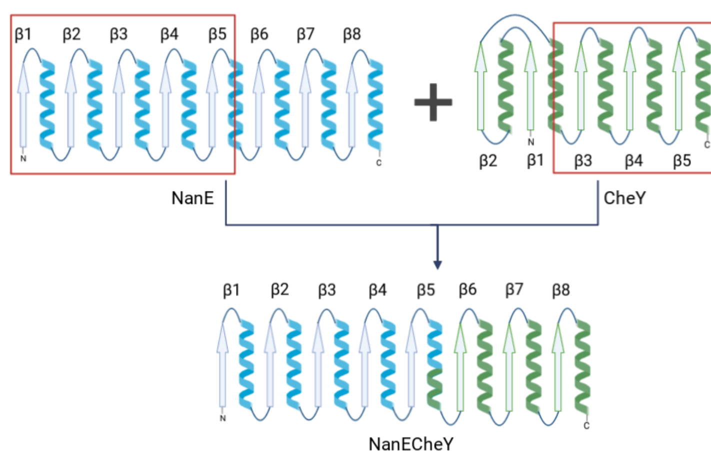


Figure 1. *Topographic representation of NanECheY chimera and its parent proteins, NanE (blue) and CheY (green), respectively.*

1. *Romero-Romero S., Kordes S., Michel F., Höcker B. (2021). Evolution, folding, and design of TIM barrels and related proteins. Current Opinion in Structural Biology, 68, 94-104.*
2. *Ferruz, N., Lobos, F., Lemm, D., Toledo-Patino, S., Farías-Rico, J. A., Schmidt, S., & Höcker, B. (2020). Identification and Analysis of Natural Building Blocks for Evolution-Guided Fragment-Based Protein Design. Journal of Molecular Biology, 432(13), 3898–3914.*
3. *Ferruz, N., Noske, J., & Höcker, B. (2021). Protlego: A Python package for the analysis and design of chimeric proteins. Bioinformatics (Oxford, England), 37(19), 3182–3189.*

Binding PhaP1, PhaP2 and PhaP3 to regulatory region phaC gene of *Azospirillum brasilense*

Yahev Domínguez-Antunez, Lucía Soto-Urzua, María de los Ángeles Martínez-Martínez and Luis Javier Martínez-Morales*

Centro de Investigaciones en Ciencias Microbiológicas, Instituto de Ciencias, Benemérita Universidad Autónoma de Puebla, Puebla, México.

Introduction. Phasins are amphiphilic proteins involved in the regulation, number, and size of polyhydroxybutyrate (PHB) granules. PHB is a polyester of 3-hydroxybutyrate. *Azospirillum brasilense* Sp7 accumulates high quaintities of bioplastic PHB. In bacteria, PHB granules are covered by granule associated proteins (GAP) such as PHB polymerase (PhbC), PHB depolymerase (PhbZ), regulators proteins (like PhaR or PhaM) and structural proteins, known as phasins (PhaP). Studies on the *A. brasilense* Sp7 genome, have shown that it contains 6 genes with a Phasin_2 domain. The aim of this study is to analyse the interaction between phasins (PhaP1, PhaP2 and PhaP3) and promoter región of *phb*C of *A. brasilense* Sp7.

Methodology. Three-dimensional protein structures of PhaP1, PhaP2 and PhaP3 were obtained by the I-TASSER server. The promoter predicted region of *pha*C was obtained at NCBI database by selecting the 233 upstream nucleotides of *pha*C. Then, HDOCK server and BIOVIA discover studio visualizer were used to analyse the interaction between phasins (ligands) and promoter region of *phb*C (receptor). The interactions between phasins and the predicted promoter of *phb*C were analyzed by observing the types of interactions they shown. Also, it was observed how stable or unstable the binding of the proteins to the upstream region of the *pha*C gene can be.

Results. It was observed that each phasin (PhaP1, PhaP2 and PhaP3) binds to the *pha*C upstream región (Figure 1). The interactions between phasins and *phb*C predicted promoter were mainly non-covalent bonds and salt interactions.

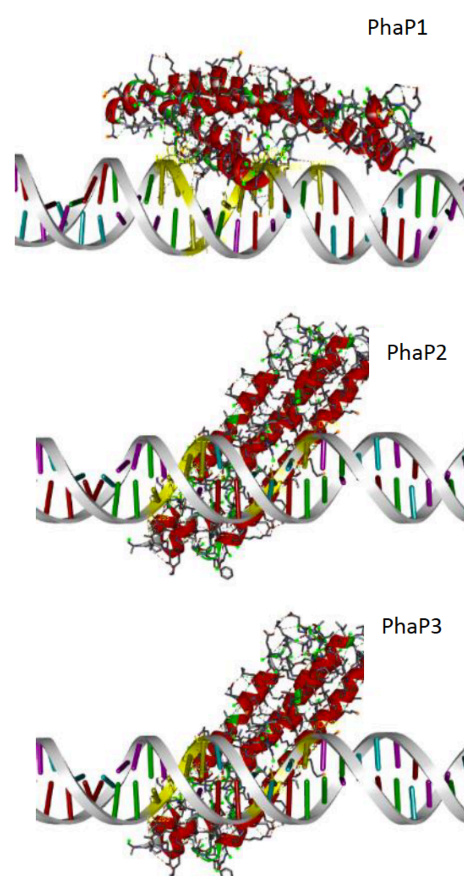


Figure 1. Interactions between A. brasilense Sp7 phasins and phbC promoter region.

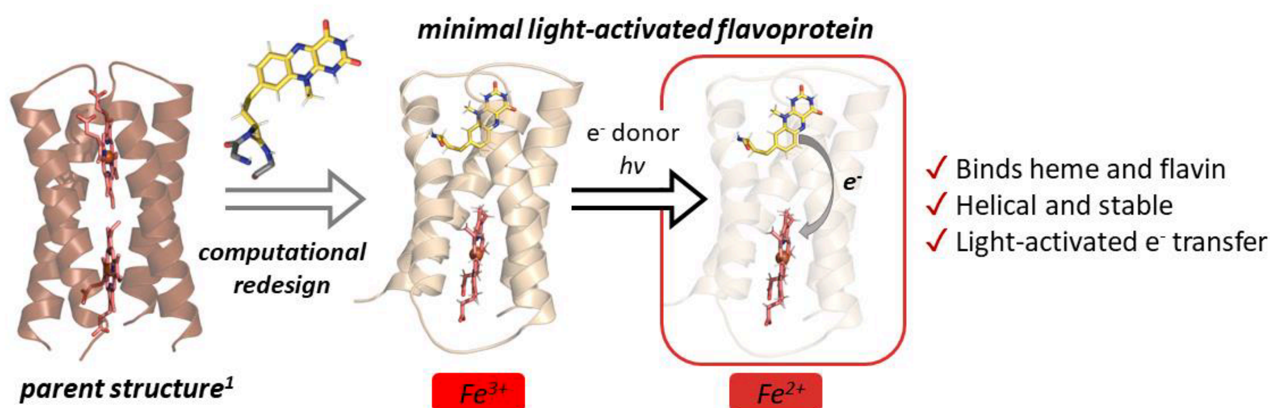
Conclusion. Bioinformatic analyzes showed posible interactions between phasins and *phb*C. However, verifying it experimentally will be the key point to corroborate the data obtained so far.

Martínez-Martínez, M., González-Pedrajo, B., Dreyfus, G., Soto-Urzúa, L., & Martínez-Morales, L. J. (2019). Phasin PhaP1 is involved in polyhydroxybutyrate granules morphology and in controlling early biopolymer accumulation in Azospirillum brasilense Sp7. AMB Express, 9(1), 155. https://doi.org/10.1186/s13568-019-0876-4.

Computational design of photoactive *de novo* flavoproteins

Paulina Dubiel[1], Marc van der Kamp[1], J. L. Ross Anderson[1]

[1]*Department of Biochemistry, University of Bristol, BS8 1TD, Bristol, United Kingdom*
Corresponding email: paulina.dubiel@bristol.ac.uk



Electron transfer between redox centres is at the core of biochemical processes essential for life, such as respiration and photosynthesis. However, studying these processes can be challenging due to complex nature of bioenergetic proteins, in which functionality is often obscured by the protein's evolutionary legacy. This can be addressed by designing proteins *de novo*, with redox active cofactors, such as hemes or flavins, placed within a simple protein scaffold. Flavins are of particular interest as they are capable of one and two electron transfer reactions, and can be photoactivated, enabling light-induced electron transfer and catalysis. Here, we engineer a flavin-binding site into a *de novo* protein scaffold, using computational methods to design a binding site with covalently anchored flavin. We demonstrate that this approach represents a successful strategy for incorporating flavins into several *de novo* heme-containing proteins. Using this strategy, we obtain two proteins which both successfully bind heme and covalently anchored riboflavin in the desired stoichiometry, whilst maintaining highly helical fold and high thermal stability. We demonstrate these proteins' ability to perform light-activated heme reduction in the order of minutes, as well as light-induced electron transfer to a *de novo* designed heme-containing membrane protein. Furthermore, we investigate how natural flavin transferases can be exploited to covalently attach natural flavins to *de novo* protein scaffolds, with our preliminary results indicating that this is a viable strategy. The work described here creates a foundation for incorporating covalently bound flavins into designed scaffolds, enabling creation of minimal light-activated redox proteins. Future efforts will involve utilising flavin transferases to attach flavins *in vivo*, as well as using the described strategies to create a photoactivated *de novo* catalyst.

1. ***Hutchins, G. H***. *et al. Precision design of single and multi-heme de novo proteins. bioRxiv 2020.09.24.311514 (2020).*

Characterization of rare and novel AlphaFold structural space

Janani Durairaj [1,2], Mehmet Akdel [3], Miguel Correa Marrero [4], Pedro Beltrão [4], Torsten Schwede* [1,2]

[1]Biozentrum, University of Basel, Basel, 4056, Switzerland.
[2]SIB Swiss Institute of Bioinformatics, Biozentrum, University of Basel, Basel, 4056, Switzerland
[3]Wageningen University and Research, Wageningen, 6708 PB, Netherlands
[4]Institute of Molecular Systems Biology, Department of Biology, ETH Zurich, Zurich, Switzerland

The recent and ongoing releases of millions of highly accurate computationally predicted structures as part of the AlphaFold Protein Structure Database has greatly expanded structural coverage of proteomes across species and function. We perform a global comparison of structural elements between over a million proteins from the AlphaFold database and over 100,000 experimentally characterised structures from the PDB by combining rotation-invariant “shape-mers”, analogous to sequence k- mers, and topic modelling, an unsupervised approach to discover abstract “topics” in structural space. We focus on high confidence predictions of structural elements, i.e. combinations of shape-mers, which are absent or very rare in experimentally characterised structures. We pinpoint areas of protein functional and evolutionary space now expanded by AlphaFold as well as conformations which could represent novel folds. The shape-mer topic modelling approach allows for fast comparison of millions of structures while also defining connected structural elements spanning across a protein which, analogous to sequence motifs, can then be linked to protein function.
Code: https://github.com/TurtleTools/alphafold-structural-space

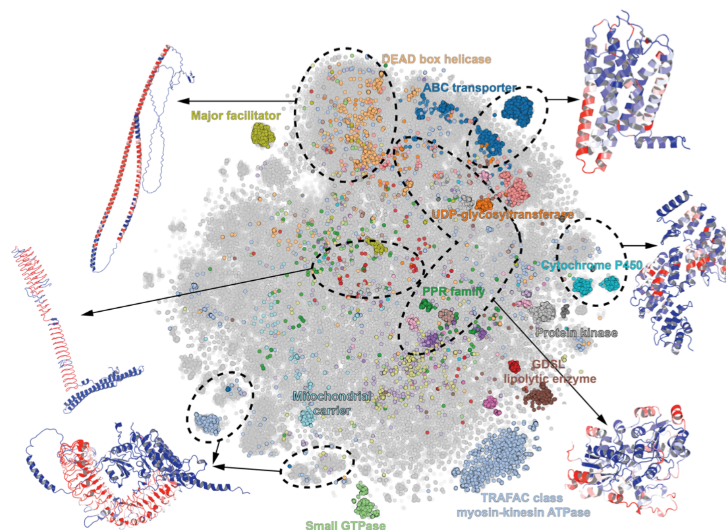


Figure 1: *Structural space of AlphaFold structures for 21 species. Proteins with structural elements rare in the PDB are highlighted with dashed circles and representative structures are shown for each with the rare elements coloured in red. Some common superfamilies are coloured and labelled.*

The binding of dermatan sulfate to alpha-L-iduronidase: An *in silico* analysis of the impact of pathogenic variants.

Darinka P. Durán-Gutiérrez[1*], Cesar A. Reyes-López[1], José Luis Castrejón Flores[2]

*[1] Sección de Estudios de Posgrado e Investigación, ENMyH, Instituto Politécnico Nacional. Guillermo Massieu Helguera, No. 239, Fracc. "La Escalera", Ticomán, C.P. 07320, Ciudad de México, México. [2] Sección de Estudios de Posgrado e Investigación, UPIBI, Instituto Politécnico Nacional *Corresponding author (ddurang1804@alumno.ipn.mx)*

Human alfa-L-iduronidase (IDUA) is a lysosomal glycoside hydrolase, involved in the sequential degradation of the glycosaminoglycans (GAGs) heparan sulphate (HS) and dermatan sulphate (DS). These types of GAGs are found in different tissues such as cornea, sclera, blood vessel walls, heart valves and umbilical cordon, in the case of DS, and within the cell surface and the extracellular matrix for the HS. Deficiency of IDUA enzyme activity causes undegraded DS and HS to accumulate in multiple tissues leading to organ dysfunction known as mucopolysaccharidosis type I (MPS I). The spectrum of symptoms in MPS I patients varies from very severe for the Hurler phenotype, intermediate for the Hurler-Scheie phenotype, and mild in the Scheie phenotype. The levels of enzymatic activity of the IDUA obtained with the synthetic substrate 4-MUI have been associated with a mild, intermediate, or severe phenotype for some pathogenic variants, however, there is a significant number of pathogenic variants that do not have this correlation. Molecular dynamics and molecular dockings were performed with representative structures of IDUA WT and the variants P533R, S633L and S633W, to evaluate their impact in the binding of a natural substrate, a dermatan sulphate tetrasaccharide and the synthetic substrate 4-MUI. The results obtained suggest that the GAG tetrasaccharide binds to IDUA through a groove at the left of the catalytic pocket, between glycosilations in Asn372 and Asn415. Interestingly, a significant frequency of pathogenic variants can be found in this region, most of which are associated with a severe phenotype. On the other hand, the molecular dockings with the variants allowed the formation of a viable complex with the tetrasaccharide with only one structure of the variant associated with the mild phenotype, while in the case of 4-MUI, all the variants allowed obtaining viable complexes with a binding ΔG lose to the those obtained with the IDUA WT structure. These data could support the proposal that the synthetic substrate is not capable of adequately sensing the catalytic activity of the pathogenic variants, so that they can be correlated with a specific phenotype as part of an early diagnosis of the pathology. It is possible that a larger synthetic substrate could sense changes in the GAG chain binding groove, which would allow enzymatic activity to correlate with an MPS I phenotype.

Designing stable metalloproteins using deep learning

Simon L. Dürr, Ursula Röthlisberger

See p. 25 (selected speakers)

De novo Proteins - Expression and Structural Analysis

Lars Albert Eicholt[1], Andreas Lange[1], Erich Bornberg-Bauer[1,2]

[1] *Institute Institute for Evolution and Biodiversity, University of Muenster, Germany.*
[2]*Department of Protein Evolution, Max Planck Institute for Developmental Biology, Tübingen, Germany*

6th January 2022

It is widely accepted that new genes emerge from duplication and divergence from prior existing genes. Improved methods of genome sequencing and comparison have led to the discovery of genes that show no homologs outside of a given species or lineage. Since convergent gene loss in multiple lineages or phases of accelerated evolution that led to loss of sequence similarity would be highly uncommon, it is assumed that "orphan genes" have in fact arisen *de novo* from non-coding DNA.

Although there is strong evidence from the genomic point of view there is a lack of knowl- edge regarding structure and function of *de novo* proteins.
One recent example of a structurally and functionally described *de novo* protein is Goddard in *Drosophila*, which plays an important role during spermatogenesis (Lange et al., 2021). A continuing problem remains the expression and purification of such proteins. Currently, we focus on how chaperones (GroEL/GroES) influence the expressability and solubilty of *de novo* proteins from *Drosophila melanogaster* and *Homo sapiens* in correlation to their disorder level.

We aim to express *de novo* proteins from different species with the help of chaperones in *E. coli* and analyse them functionally for enzymatic activity, foldability, disorder and further elucidate their structure *in silico* and experimentally via Cryo-EM and NMR.
Additionally, we use Ancestral reconstruction to investigate the emergence of *de novo* pro- teins and how their sequence and structure alternated during the course of evolution. Exploration of naturally arisen *de novo* proteins would give insights into early protein evolu- tion and broaden our knowledge of the sequence space explored by nature on earth.

Damietta: A Tensorised Protein Design Engine, and Its Application in Therapeutic Protein Design

Kateryna Maksymenko[1,2], Birte Hernandez Alvarez[1], Timo Ullrich[1,2], Patrick Mueller[2], Julia Skokowa[3], Andrei Lupas[1], Mohammad ElGamacy[1,2,†]

[1]Max Planck Institute for Developmental Biology, 72076 Tübingen (Germany) [2]Friedrich Miescher Laboratory of the Max Planck Society, 72076 Tübingen (Germany) [3]University Hospital Tübingen, Division of Translational Oncology, 72076 Tübingen (Germany)

Despite the groundbreaking successes of computational design in recent years, simultaneous improvements of design throughput and accuracy are continually needed to achieve better experimental success rates, and tackle more difficult design problems. We describe Damietta[1]; a novel protein design framework that maximises computational efficiency by tensorising energy calculations, and improves accuracy by relying on a self-consistent scoring function. This scoring function is not trained or contaminated by any learnt parameters, but relies purely on physics-based force field. We deploy these design concepts to tackle three design problems with different levels of difficulty, yielding agonists and antagonists of growth factor signaling pathways with therapeutic potential. First, we use Damietta to design epidermal growth factor (EGF) inhibitors based on an EGF receptor template structure. Testing only two designs, they were capable of binding EGF and inhibiting its signaling in cells. Second, we also use Damietta to create stabilised variants of the granulocyte-colony stimulating factor (G- CSF), which could proliferate G-CSF-dependent cell line.Third, we design a bispecific, single-domain cytokine, capable of engaging two different cytokine receptors (here, we start by a IL3-Rα/G-CSFR combination). Such a “novokine” possesses a novel fold, and can serve as a non-natural cytokine with novel function. These applications exemplify the design of proteins with therapeutic potential, and demonstrate Damietta to be applicable for a range of protein design and engineering problems.

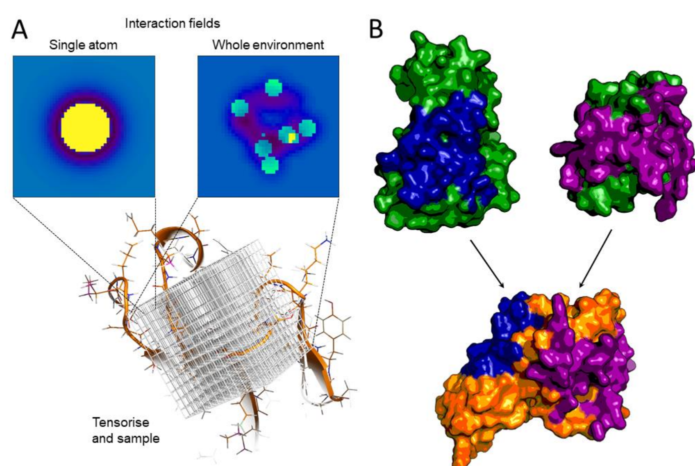


Figure 1. The concept and applications of the Damietta design framework. (A) *A tensorised representation of the non-bonded interaction fields can accelerate protein design calculations through i) single instruction, multiple data processing paradigm, ii) precomputing of rotamer fields, and iii) enabling parallel implementation. (B) An example application presented here is the design of a novokine; a novel cytokine that bind and co-localise a non-native combination of cytokine receptor subunits.*

A pre-release of the Damietta software is available at: https://www.eb.tuebingen.mpg.de/damietta/

Prediction and Engineering of Post-translational Modifications in Rosetta using Machine-learning during Structure-based Protein Design

Moritz Ertelt[1,2], Jens Meiler[1,3], Clara T. Schoeder[1]

1 Institute for Drug Discovery, University Leipzig Medical School, Germany; 2 Center for Scalable Data Analytics and Artificial Intelligence ScaDS.AI, Dresden/Leipzig, Germany; 3 Department of Chemistry, Vanderbilt University, Nashville, TN, USA

Post-translational modifications of proteins play a vital role in their function and stability. These modifications influence protein signaling, degradation, folding, aggregation, protein-protein interactions, enzyme activity, binding affinity, and many more. To date over 400 types of post-translational modifications have been described, highlighting the added complexity that lays beyond the limits of the amino acid code. Such modifications not only pose a challenge to the successful design of proteins, but also represent a mayor opportunity to diversify the protein engineering toolbox. Engineering of post-translational modifications, however, remains challenging as underlying mechanisms are not yet fully understood. To this end, we first trained machine-learning models to accurately predict four of the most abundant post-translational modifications, including protein glycosylation, phosphorylation, sulfation, and deamidation. In a second step these models were implemented inside the computational protein modeling suite Rosetta which allowed us the flexible combination with existing protocols to model the modified sites and understand their impact on protein stability as well as function. Lastly, we developed a new design protocol that either maximizes or minimizes the predicted probability of a particular site being modified. We find that this combination of accurate prediction and structure-based design can enable the modification of existing, as well as the introduction of novel, post-translational modifications. The potential applications of our work include, but are not limited to, stabilizing proteins through glycosylation, strengthening protein-protein interactions through phosphorylation or sulfation, as well as protecting proteins from deamidation liabilities. These applications are especially important for the design of new protein therapeutics where post-translational modifications can drastically change the therapeutic properties of a protein. In conclusion, our work contributes by adding a novel tool to Rosetta's protein engineering toolbox, that allows for the rational design of post-translational modifications.

A deep unsupervised language model for protein design

Noelia Ferruz[1,2*], Steffen Schmidt[3], Birte Höcker[1]

[1]*Department of Biochemistry, University of Bayreuth, 95447 Bayreuth, Germany.*
[2]*Current address: Institute of Computer Science and Applications, University of Girona*
[3]*Computational Biochemistry, University of Bayreuth, 95447 Bayreuth, Germany.*
noelia.ferruz@udg.edu

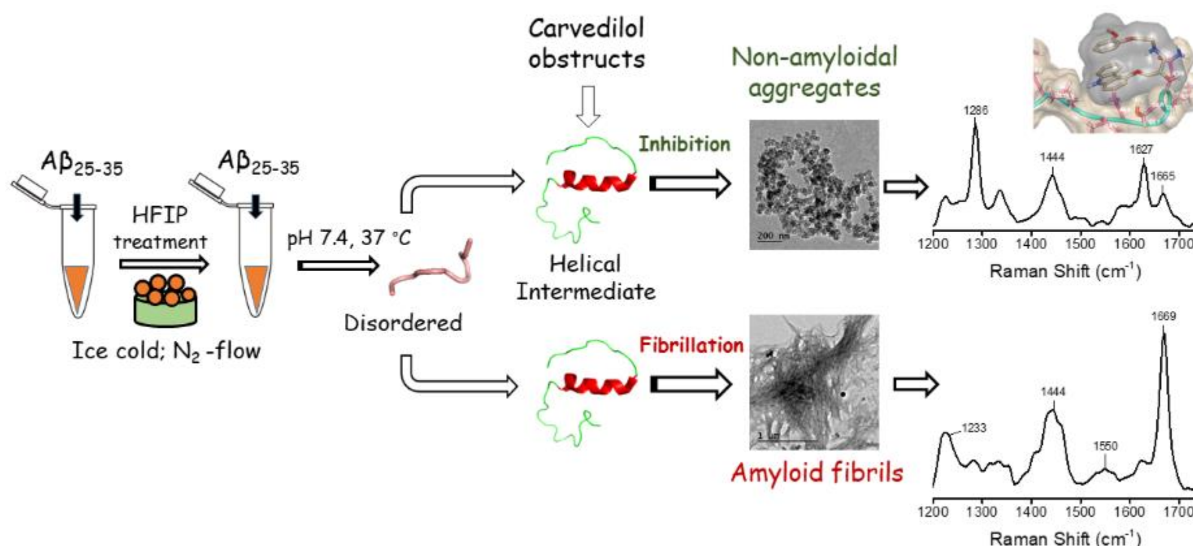
Protein design has the potential to tackle many environmental and biomedical problems. Recent progress in the field of natural language processing (NLP) has enabled the implementation of ever-growing language models capable of understanding and generating text with human-like capabilities. Given the many similarities between human languages and protein sequences, the use of NLP models offers itself for predictive tasks in protein research. Motivated by the evident success of generative Transformer-based language models such as the GPT-x series, we trained ProtGPT2, a language model trained on protein space that generates *de novo* protein sequences that follow the principles of natural ones. In particular, the generated proteins present amino acid propensities which resemble natural proteins, whereas disorder and secondary structure prediction show that 88% of ProtGPT2- generated proteins are globular, in line with natural sequences. Sensitive sequence searches in protein databases demonstrated that ProtGPT2 sequences are distantly related to natural ones, and similarity networks further evidenced that ProtGPT2 is sampling unexplored regions of the protein space. AlphaFold prediction of sequences revealed well-folded structures with high pLDDT scores. Therefore, ProtGPT2 has the potential to generate *de novo* proteins in a high-throughput fashion in a matter of seconds. The model is easy-to-use and available to the community.

Early-stage helical intermediate formation of A$\beta_{25\text{-}35}$ perturbed by Carvedilol: Insight into the aggregation pathway using the biophysical tools

Sudeshna Ghosh* and Sandeep Verma*

*Department of Chemistry, Indian Institute of Technology Kanpur, UP 208016, India

Amyloid-beta (Aβ) peptides undergo aggregation resulting in highly structured amyloid fibrils. Recent research has revealed that the early-stage structural changes of Aβ peptides are of great significance. It is challenging to acquire due to the nature of the amyloids and experimental constraints. Interestingly, inhibitors can modulate the aggregation pathway and the structural events. Our present work describes the aggregation cascade (early-late; 0-100 h) of the Aβ25-35 peptide in the absence and presence of carvedilol, a nonselective β-adrenergic receptor blocker. Thioflavin T (ThT) fluorescence, dynamic light scattering (DLS), circular dichroism (CD), Raman spectroscopic techniques, and imaging experiments are used to monitor the aggregation process of A$\beta_{25\text{-}35}$ peptide. We notice that the A$\beta_{25\text{-}35}$ peptide undergoes an early-stage (3-6 h) helical intermediate formation during fibrillation (using CD and Raman measurements). Carvedilol prevents the helical intermediate formation of A$\beta_{25\text{-}35}$ peptide resulting in inhibition.



The secondary structural changes are distinct during aggregation in the absence and presence of carvedilol. CD spectra and deconvolution of the Raman bands indicate the β-sheet formation (24-100 h) in the absence of carvedilol. Spectroscopic analysis shows a disordered structure for the peptide in the presence of carvedilol (24-100 h). Microscopic images ensure the formation of polymorphic fibrils and non-amyloidal aggregates for the peptide alone and with carvedilol, respectively. The plausible mode of interaction with carvedilol involves the C-terminal residues of the peptide, as suggested by the

molecular docking study. Our research provides intricate details of structural changes of the Aβ_{25-35} peptide over a considerable period (0-100 h), which can benefit strategic therapeutic design.

Published as: Ghosh S and Verma S, Carvedilol inhibits Aβ25-35 fibrillation by intervening the early stage helical intermediate formation: A biophysical investigation, International Journal of Biological Macromolecules, (2021),188, 263-271.

(1) *C.L.L. Pham, A.H. Kwan, M. Sunde, Functional amyloid: widespread in nature, diverse in purpose, Essays Biochem., 56 (2014), 207-219.*

(2) *(2) M. Bucciantini, E. Giannoni, F. Chiti, F. Baroni, L. Formigli, J. Zurdo, N. Taddei, G. Ramponi, C.M. Dobson, M. Stefani, Inherent toxicity of aggregates implies a common mechanism for protein misfolding diseases, Nature, 416 (2002), 507-511.*

(3) *F. Chiti, C.M. Dobson, Protein misfolding, functional amyloid, and human disease, Annu. Rev. Biochem., 75 (2006), 333-366.*

(4) *Y. Song, P. Li, L. Liu, C. Bortolini, M. Dong, Nanostructural differentiation and toxicity of amyloid-ß25-35 aggregates ensue from distinct secondary conformation, Sci. Rep., 8 (2018), 765.*

(5) *P. Kabiraj, J.E. Marin, A. Varela-Ramirez, M. Narayan, An 11-mer amyloid Beta peptide fragment provokes chemical mutations and Parkinsonian biomarker aggregation in dopaminergic cells: a novel road map for "Transfected" Parkinsons, ACS Chem. Neurosci., 7 (2016), 1519-1530.*

(6) *A. Accardo, V. Shalabaeva, M. Cotte, M. Burghammer, R. Krahne, C. Riekel, S. Dante, Amyloid ß peptide conformational changes in the presence of a lipid membrane system, Langmuir, 30 (2014), 3191-3198.*

Development of a versatile PET degradation assay platform and its application to a halophilic PETase

Sebastian Weigert[1], Florian J Gisdon[1], Pablo Perez-Garcia[2], Jennifer Chow[2], Andreas Gagsteiger[1], Matthias Ullmann[3], Wolfgang Streit[2], Birte Höcker[1]

[1]*Department of Biochemistry, University of Bayreuth, 95447 Bayreuth, Germany*
[2]*Department of Microbiology and Biotechnology, University of Hamburg, 22609 Hamburg, Germany*
[3]*Computational Biochemistry, University of Bayreuth, 95447 Bayreuth, Germany*

Dealing with plastic waste and the associated ubiquitous occurrence of microplastics is a serious challenge of our time (1). Recent discoveries, especially in the field of polyethylene terephthalate (PET) degrading enzymes potentially lead to innovative solutions for natural degradation and recycling applications. To discover, characterize, and develop such PET-degrading enzymes, we have developed a high-throughput assay (2), which is characterized by the application of a PET coating on lab consumables. This allows for convenient handling of a difficult substrate and, at the same time, provides unparalleled freedom in the composition of the coating. The system was showcased in the study of PET6 from Vibrio gazogenes, a proposed PET-degrading enzyme (3) from a halophilic organism common in the marine environment worldwide. The experiments revealed remarkable salt tolerance of PET6, which even showed the highest turnover above 1M sodium chloride. By introducing two minor mutations into PET6, we were able to increase its activity to about 1.5-fold, demonstrating the evolutionary potential of the enzyme. The enhancement caused by the mutations was elucidated by MD simulations revealing more refined enzyme-substrate interactions. Given the enormous amounts of plastic waste entering the ocean and that Vibrio gazogenes are widely distributed in marine biofilms and estuaries, the halophilic activity profile suggests PET6 as a suitable candidate for natural degradation of PET.

---

(1) Dris R, Agarwal S, Laforsch C: Plastics: From a Success Story to an Environmental Problem and a Global Challenge. Global Challenges ***2020****, 4 2000026. doi:10.1002/gch2.202000026*
(2) Weigert S, Gagsteiger A, Menzel T, Höcker B: A versatile assay platform for enzymatic poly(ethylene-terephthalate) degradation. Protein Engineering, Design & Selection ***2021****, 34 , 1-9. doi:10.1093/protein/gzab022*
(3) Danso D, Schmeisser C, Chow J, Zimmermann W, Wei R, Leggewie C, Li X, Hazen T, Streit WR: New Insights into the Function and Global Distribution of Polyethylene Terephthalate (PET)-Degrading Bacteria and Enzymes in Marine and Terrestrial Metagenomes. Appl Environ Microbiol ***2018****, 84 (8), e02773-17. https://doi.org/10.1128/AEM.02773-17.*

Design of programmable and addressable DNA-protein hybrid nanostructures

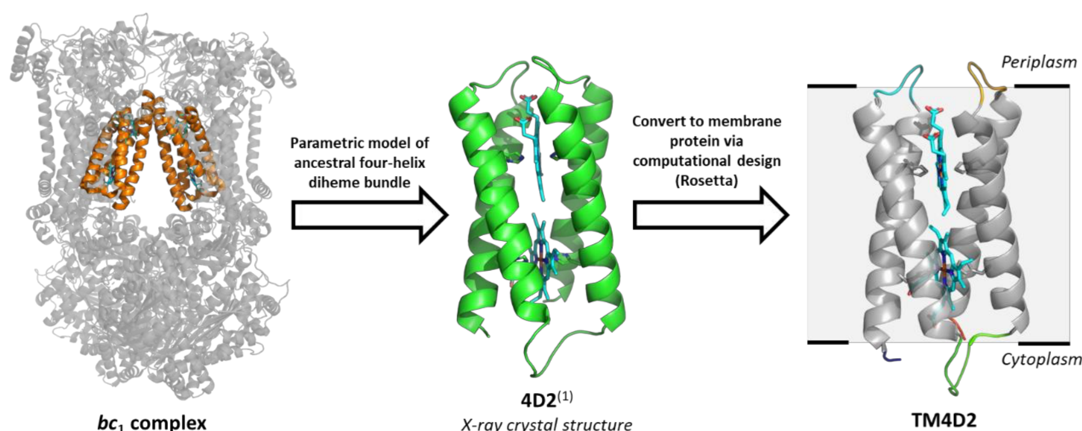
Robbert de Haas

Currently, DNA nanotechnology is the only way of creating truly programmable and addressable nanostructures, but these structures are fundamentally limited in size and are not optimally suited for integration with the cellular machinery. Inspired by TALEs (transcription activator-like effectors), we have designed highly stable repeat proteins that bind sequence non-specifically to DNA by wrapping around its major groove. Sequences for TALE-mimics were designed using Rosetta computational protein design. The DNA-templated protein assembly is highly cooperative, such that the protein-coated DNA template is highly protected against nucleases. By combining these polymerizing, sequence non-specific TALE-mimics with sequence specific TALEs, we will ultimately be able to create large addressable hybrid nanostructures that are highly compatible with the cellular machinery, with a wide range applications in material science and medicine, such as designer vaccines.

Computational design of a minimal bioenergetic integral membrane protein

Ben Hardy[1], J. L. Ross Anderson[1], Paul Curnow[1]

[1]School of Biochemistry, University of Bristol, BS8 1TD Corresponding email: 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. Many respiratory and photosynthetic complexes share an ancestral four-helix bundle architecture at their core that positions two b-type hemes for transmembrane electron transport. This fundamental scaffold is an ideal basis for creating man-made bioenergetic membrane proteins. We report here the design, recombinant production and characterisation of such an artificial protein through the computational transformation of an existing water-soluble design[1] into a sequence capable of membrane integration. This design process retains the key interactions that govern bundle assembly and heme binding but otherwise the resulting construct, which we call Transmembrane-4D2 (TM4D2), has no direct sequence homology with any natural metalloprotein. TM4D2 can be recombinantly expressed in E. coli, is efficiently inserted into the plasma membrane, and appears to recruit heme in vivo. This artificial protein can be successfully purified from cellular membranes using mild non-ionic detergents. Absorbance spectroscopy confirms that TM4D2 binds to two hemes via bis-histidine coordination, and redox potentiometry reveals two distinct reduction potentials split by about 112 mV, characteristic of electronic coupling by two hemes in close proximity. E. coli Flavodoxin reductase can rapidly transfer electrons from NADPH to TM4D2 in micelles, demonstrating that the hemes of TM4D2 are accessible to small diffusive redox proteins. Future work will use the TM4D2 module to construct larger bioenergetic membrane proteins by fusion to other de novo redox domains and will diversify cofactor binding sites.

---

*(1) Hutchins, G. H.; Noble, C. E. M.; Blackburn, H.; Hardy, B.; et al. Precision Design of Single and Multi-Heme de Novo Proteins. bioRxiv **2020**, 2020.09.24.311514. https://doi.org/10.1101/2020.09.24.311514.*

Recombination of TIM-Barrel and flavodoxin-like fold fragments can lead to two different folds

Farid Lukas Hassine, Abhishek Anan Jalan, Julian Beck, Birte Höcker

Department of Biochemistry, University of Bayreuth, Bayreuth, Germany

The current functional and structural diversity of the protein world is the consequence of protein evolution. Proteins evolve via mutation, deletion, recombination and duplication of protein fragments. We have focused on the recombination of subdomain-sized fragments targeting the following questions: When do the resulting proteins properly fold? When are they misfolded and dismantled? What determines the success of these events and when might new protein folds emerge? Here we present two protein chimera that were designed using fragments from the TIM-Barrel and the flavodoxin-like fold with the goal to construct a new $(\alpha\beta)_4$ -half barrel called Chelix-1 and its duplicate Chelix-2 that we expected to form a $(\alpha\beta)_8$-TIM-barrel (see Figure). While initial characterizations with circular dichroism spectroscopy suggest well-folded proteins with high thermal stability (>95°C), analysis by 2D NMR resembles rather a molten globule state based on poor dispersion of amide cross-peaks. Thus, we modelled the putative structure of Chelix-1 and Chelix-2 using RosettaFold and AlphaFold to investigate this behaviour. The predicted structures suggest a flavodoxin-like fold for Chelix-1 and a new fold with an extended $(\alpha\beta)_4$-$(\alpha\beta)_4$ like topology for Chelix-2. Analysis of the oligomerization state of Chelix-2 with SEC- MALS shows a larger hydrodynamic radius than expected for a TIM-barrel. The data suggests the protein to either be packed loosely or to form a new extended structure as predicted. Further optimizations of the interfaces are considered to improve packing of the protein both towards a TIM-barrel as well as the newly predicted structure.

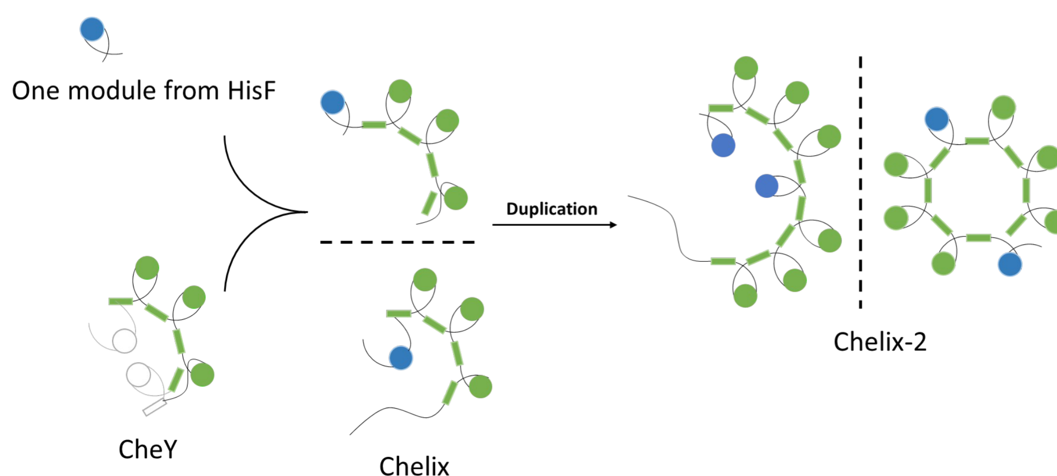


Figure 1: Chelix-1 and Chelix-2 chimeras build from fragments of the proteins HisF and CheY. The position of the helix module from HisF determines the topology of the protein, with two possible outcomes.

Archaeal Triosephosphate Isomerase Sequences and Structures: An evolutionary approach

Andrea Hernandez Garcia[1], Ricardo Muñiz Trejo[2], Sergio Romero Romero[3], Daniel Alejandro Fernández Velasco[1]

[1]School of Medicine UNAM | Department of Biochemistry. [2]The University of Chicago | Department of Ecology and Evolution. [3]University of Bayreuth

The (β/α)8 or TIM barrel fold is a widespread domain that catalyzes 6 of the 7 catalytic functions. It is the most common fold in metabolic pathways and is present in the three domains of life. Its ability to explore many functions keeping its fold (robustness) and its wide genetic background turn it an outstanding model to study evolvability. The TIM barrel name comes from the Triose phosphate isomerase enzyme (TIM) that is an excellent model to study thermostability and evolvability because it is conserved in the three domains of life and many protein structures from thermophilic, mesophilic, and psychrophilic organisms are available. The TIMs from Bacteria and Eukarya are dimers with ~250 aa length, whereas the TIMs from Archaea are ~10% smaller with ~225 aa length, and can be found as dimers, or tetramers. Tetramerization has been proposed as a temperature adaptation. We have characterized physicochemically the TIMs of *Methanocaldococcus jannaschii* (MjaTIM) and *Methanotorris formicicus* (MfoTIM), two thermophilic archaea that show 81% identity. Both TIMs are tetramers but have a considerable difference in melting temperature (~20°C). This is clearly not a cause of the oligomeric state, so the differences must be structural. Analysis of the MjaTIM and MfoTIM structures show that the tetrameric interface is shaped by hydrophobic surfaces between helices 4, 5 and 6. In tetramers helix 5 presents a conserved alanine sequence (AVAAL) that tightly packs to form the tetramer. We aligned this zone with the crystallographic structure of *Thermoplasma acidophilum* TIM (TacTIM) (PDB:5CSR) which is a dimer and found a deletion of the only two in helix 6 residues that participate in the interface, whereas the alanine sequence was replaced by (KFREY) where the aromatic and charged residues cause steric clashes, hampering the tetramer assembly. We searched if these sequence features are present in mesophilic and psychrophilic TIM sequences, as expected for the temperature adaptative hypothesis. Surprisingly, these changes are not associated with mesophilic and psychrophilic organisms but with acidophilic organisms. This indicates that dimerization is in this case a phylogenetic divergence not an adaptation.

Natural history and phyletic distribution of accessory subunits e and g that participate in the formation of dimeric/oligomeric F_1F_o ATP synthase.

José Alfredo Hernández-Zúñiga[1], Oscar Flores-Herrera[1], Enrique García- Hernández[2], Hector Riveros-Rosas[1].

[1] *Departamento de Bioquímica, Facultad de Medicina, Universidad Nacional Autónoma de México. Ciudad Universitaria, 04510, Ciudad de México, CDMX, México.* [2]*Instituto de Química, Universidad Nacional Autónoma de México, Ciudad Universitaria, 04510 Ciudad de México, CDMX, México.*

F_1F_O ATP synthase complex is largely conserved across bacteria, mitochondria and chloroplast. In mitochondria, this complex can be found as a dimer/oligomer formed through accessory subunits. The dimers of mitochondrial ATP synthase are unique because they organize into long rows that induce membrane invaginations named cristae. However, a wide diversity at the morphology of mitochondrial cristae can be found among Eukaryotes.

On other hand, several dimeric/oligomeric F_1F_O ATP synthase arrangements have been reported in eukaryotes: ovine, bovine, *Saccharomyces cerevisiae*, *Chlamydomonas reinhardtii*, *Tetrahymena thermophila*, *Trypanosoma Brucei*, and *Euglena gracilis*. In some model organisms different subunits seems to be involved in dimer formation.

To get insights about how these subunits were recruited into the ATP synthase complex, we performed a phylogenetic analysis of reported accessory subunits e and g involved in dimer/oligomer ATP synthase formation.

Amino acid sequences were retrieved either from Pfam (http://pfam.xfam.org/), or using BlastP searches against the complete genomes available at NCBI's RefSeq genome database (ftp://ftp.ncbi.nlm.nih.gov/genomes/refseq). Progressive multiple amino acid sequence alignments were performed with ClustalX (http://www.clustal.org/clustal2/) or MUSCLE (https://www.drive5.com/muscle/). Phylogenetic analyses were conducted using the MEGA_X (http://www.megasoftware.net).

Accessory subunit e and g were found only in eukaryotes; therefore, these proteins originated after the last eukaryotic common ancestor. Both subunits were found in animals, fungi, and plants. In both cases conserved the motive GXXXG. This suggest a broad distribution of dimeric/oligomeric ATP synthase complexes, but probably other accessory subunits involved in dimer/oligomer formation evolved independently several times through the evolution of eukaryotes.

Supported by UNAM-PAPIIT grants IN218819 and IN219022. JAHZ is a student of Master's Degree in the Biological Sciences Program, UNAM (Supported by CONACYT 1033911).

Scaling adaptive seeding strategies for protein folding to 2000+ GPUs

Eugen Hruska, Emory University

High performance computing enables to simulate protein folding for larger proteins. To efficiently utilize the computational resources, good scaling of both the software and the algorithms is necessary. The algorithmic scaling of complex protein folding workflows like adaptive seeding strategies is not fully understood. The presented results demonstrate that both software and algorithms can scale to 2000+ concurrent GPUs for adaptive seeding strategies. This scaling performance is better than for plain parallel molecular dynamics simulations. The accuracy of the resulting protein folding, and protein kinetics is validated with benchmark simulation.

---

[1] Hruska, E., Balasubramanian, V., Lee, H., Jha, S. & Clementi, C. Extensible and Scalable Adaptive Sampling on Supercomputers. J. Chem. Theory Comput. 16, 7915–7925 (2020).
[2] Hruska, E., Abella, J. R., Nüske, F., Kavraki, L. E. & Clementi, C. Quantitative comparison of adaptive sampling methods for protein dynamics. J. Chem. Phys. 149, 244119 (2018).

Global analysis of large multi-mutant libraries to discover stabilizing amino acid substitutions.

Kristoffer E. Johansson, Kresten Lindorff-Larsen and Jakob R. Winther

The Linderstrøm-Lang Centre for Protein Science, Department of Biology, University of Copenhagen, Denmark

The identification of amino acid substitutions that both enhance the stability and function of a protein is a key challenge in protein engineering. Advances in biotechnology have enabled assaying thousands of protein variants in a single high-throughput experiment, known as multiplexed assays of variant effects (MAVE) or deep mutational scanning (DMS). Such experiments are a promising source of protein-specific information to guide studies of protein engineering and design.

We present a Global Multi-Mutant Analysis (GMMA) that exploits the presence of multiply- substituted variants to identify individual amino acid substitutions that are beneficial for both the function and stability across a large library of protein variants. We have applied GMMA to two experiments: Firstly, a published study of >54,000 variants of green fluorescent protein (GFP), each carrying on average 4 amino acid substitutions and with measured fluorescence output. Secondly, a purpose-made library of >14,000 variants of a redesigned thioredoxin fold, each carrying on average 9 amino acid substitutions and with measured folding competence.

In both cases, the GMMA method achieves a good fit to the data while being analytically transparent. And in both cases, the top-ranking substitutions are shown to directly achieve a state- of-art performing proteins without further experiments: For GFP, the six top-ranking substitutions are demonstrated to progressively enhance GFP to a level similar to super-folder GFP and our analysis recovers nearly all the substitutions previously reported to be beneficial for GFP. For the redesigned thioredoxin, the nine top-ranking substitutions are demonstrated to progressively enhance the structural stability from 50 kJ/mol to ~70kJ/mol or a melting temperature to >150°C.

Expanding the repertoire of *de novo* protein assemblies: secretion optimized and polymorphic assemblies

Alena Khmelinskaia[1,2,3], Jing Yang (John) Wang[1,2,4], Neil P. King[1,2]

See p. 27 (selected speaker)

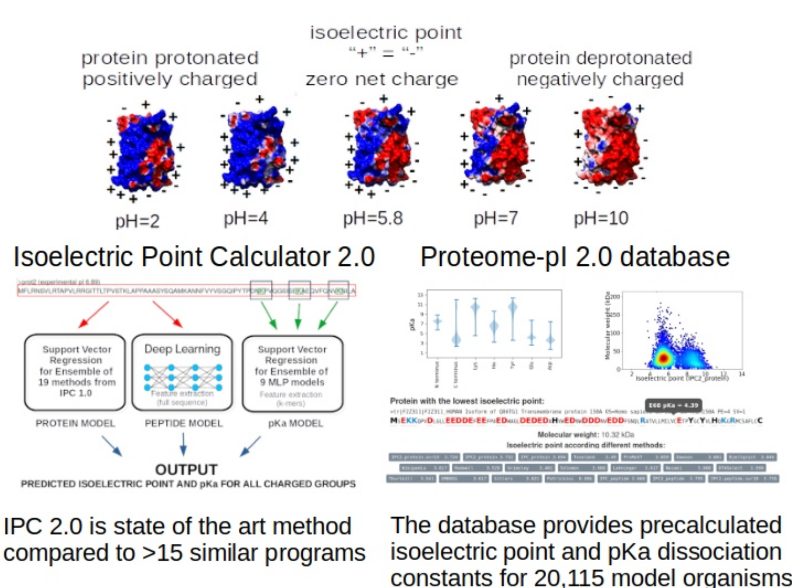
IPC 2.0 & Proteome-*pI* 2.0 – high-throughput prediction of isoelectric point and pKa dissociation constants

Lukasz P. Kozlowski

*Institute of Informatics, Faculty of Mathematics, Informatics and Mechanics, University of Warsaw, Poland, *lukaskoz@mimuw.edu.pl*

In proteins and peptides, the isoelectric point (*pI*), the pH at which a particular molecule is electrically neutral due to the equilibrium of positive and negative charges, mostly depends on dissociation constant (pKa) of charged groups of seven amino acids and NH+ and COO- groups at polypeptide termini. The information about pI and pKa's is extensively used in 2D gel electrophore sis (2D-PAGE), capillary isoelectric focusing (cIEF), X-ray crystallography, and mass spectrometry (MS). Therefore, there is a strong need for in silico prediction of pI and pKa's values.

Here, I present Isoelectric Point Calculator 2.0 (IPC 2.0, http://ipc2-isoelectric-point.org), a web server for the prediction of isoelectric points and pKa values using a mixture of deep learning (DL) and support vector regression (SVR) models that outperforms previous algorithms being at the same time a few folds faster. Next, the IPC 2.0 has been used for large-scale prediction of isoelectric point in 20,115 proteomes of model organisms (Proteome-pI 2.0 database, http://isoelectricpointdb2.org). Moreover, the proteomes have been in silico digested with the five most frequently used proteases (Trypsin Chymotrypsin, Trypsin+LysC, LysN, ArgC).

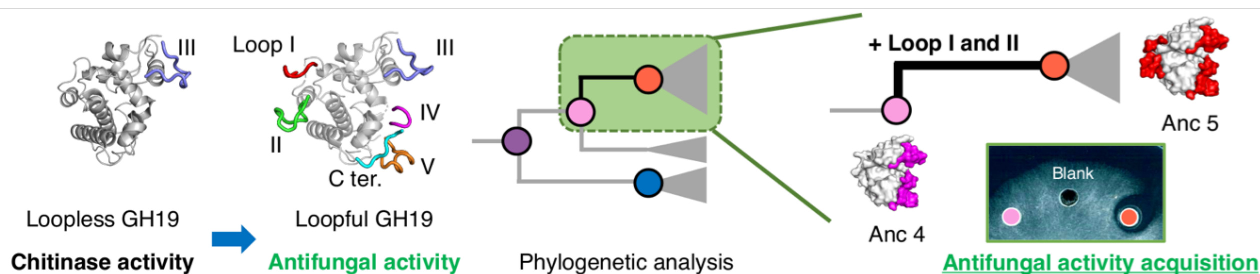


1. *Kozlowski LP (2021) IPC 2.0 – prediction of isoelectric point and pKa dissociation constants. Nucleic Acids Res., doi: 10.1093/nar/gkab295*
2. *Kozlowski LP (2021) Proteome-pI 2.0: Proteome Isoelectric Point Database Update. Nucleic Acids Res., doi: 10.1093/nar/gkab944*

InDels in the loop region shapes new protein function in the evolution of GH19 chitinase

Dan Kozome, Paola Laurino

Protein Engineering and Evolution Unit, Okinawa Institute of Science and Technology, 1919- 1 Tancha, Onna, Okinawa, Japan 904-0495.



Graphical Abstract

Substitutions, insertions and deletions (InDels) play a key role in the emergence of enzyme functions. However, our understanding of the role of InDels is limited compared to the one of substitution. Most protein engineering methods focus on exploring the sequence space of the fixed length while sequence space including InDels has been overlooked due to the ambiguity in evolutionary information and the difficulties in creating libraries. Herein we choose glycoside hydrolase family 19 (GH19) chitinase as a candidate to investigate how InDels in the loop regions contribute to the emergence of new functions in proteins. The function of GH19 varies from chitinase activity to antifungal activity, depending on the presence/absence of loop regions. Phylogenetic analysis revealed the ancestral nodes where the InDels in loop regions occurred. Characterizations of eight protein variants based on two reconstructed ancestral proteins with different loop combinations revealed the critical loop for new function acquisition. This study suggests that non-catalytic loop additions/removals are one of the strategies that Nature applies, to acquire new protein functions.

Potentials of PotF – Characteristics, Engineering & Application

Pascal Kröger, Sooruban Shanmugaratnam, Birte Höcker

University of Bayreuth

Periplasmic binding proteins (PBPs) are a highly adaptable superfamily of proteins that cover a vast spectrum of ligands in eubacteria and archaea. They display a "Venus-flytrap" architecture which consists of two β/α-lobes connected by a hinge region with the ligand binding site at their interface [1]. PBPs are predominantly open in solution and – similar to a flytrap – undergo a large conformational change (closure) upon ligand recognition.

PotF is the PBP that recruits putrescine (PUT) to its ABC transporter system in *E. coli* [2]. PotF has proven to promiscuously bind several other biogenic amines besides its wild type ligand PUT in a nM to µM range [3]. Two of these biogenic amines raise particular interest: spermidine (SPD) and agmatine (AGM)

SPD uptake in *E. coli* is facilitated by PotD, a homolog of PotF that displays 7 differences in its binding pocket. By generating different mutant variants, we were able to switch the polyamine specificity of PotF regarding PUT and SPD and pinpoint effects to certain residues [1,4]. SPD uptake is linked to pathogenesis in multidrug resistant gram-negative bacteria and has been targeted by treatment strategies [5]. Having highly adaptable PBPs like PotF and its homo- and orthologs at their disposal, pathogens might be able to hijack the PotF system to enable SPD uptake and bypass treatment strategies over time.

AGM on the other hand shows an influence on multiple neurotransmitter systems and is assumed to be a "magical shotgun", a non-selective drug with multiple targets, that can lead to more effective treatments [6,7]. Still, AGMs specific mechanisms are largely unknown and unexplored. Thus, we developed a biosensor by inserting a circularly permuted GFP into PotF to track its spatiotemporal pattern in a non-invasive manner. The new sensor allows fluorescence tracking of binding events upon closure.

1: Scheib et al. (2014) *Journal of structural bioloy 185*, 168-192.
2: Pistocchi et al.(1993) *The Journal of biological chemistry 268*, 146-152. 3: Kröger et al. (2021a) *Structure 29*, 433-443
4: Kröger et al. (2021b) *The Journal of biological chemistry 297*, 101419 5: Zhang et al. (2019) *Journal of molecular biology 431,* 4882-4896
6: Roth et al. (2004) *Nature reviews. Drug discovery 3,* 353-359
7: Piletz et al. (2013) *Drug discovery today 18,* 880-893

Artificial photo-regulation of tryptophan feedback inhibition in anthranilate synthase

Ashok Kumar Bhagat, [a] Sandra Schlee,[a] Kristina Straub,[a] Astrid Bruckmann,[b] Julian Nazet [a] Reinhard Sterner[a]*

[a] Institute of Biophysics and Physical Biochemistry, Regensburg Center for Biochemistry, University of Regensburg, 93040 Regensburg, Germany; [b] Institute of Biochemistry, Genetics, and Microbiology, University of Regensburg, 93040 Regensburg, Germany

Light-driven biocatalysis is an emerging concept in the field of synthetic biology, which allows for the spatio-temporal regulation of enzymatic activity in the concept of basic research or biotechnological applications. Among the different tools available for the use of light to gain control over the activity of enzymes, genetic code expansion is the most advanced and versatile one. Genetic code expansion uses the reprogramming of an artificially introduced stop codon to site-specifically incorporate photosensitive unnatural amino acids (UAAs) into positions of the enzyme that are crucial for its activity or regulation.

Here, on the example of anthranilate synthase (AS), we show that feedback inhibition, which is an important mechanism to regulate key metabolic enzymes, can be controlled by light via the introduction of a photosensitive UAA by means of genetic code expansion. AS catalyzes the conversion of chorismate and ammonia into pyruvate and anthranilate, which is the first out of seven steps that are required for the biosynthesis of the essential amino acid tryptophan in bacteria, fungi, and plants. It has been shown that AS activity is regulated by feedback inhibition that is triggered by the binding of the end-product tryptophan to an allosteric site of the enzyme. By incorporating the bulky photosensitive UAA *o*-nitrobenzyl-*O*-tyrosine (ONBY) close to the allosteric site of AS, tryptophan binding was sterically prevented. Upon UV irradiation, ONBY decaged to the less space-consuming tyrosine residue, leading to the restoring of tryptophan binding and activity inhibition. Our study shows that the introduction of UAAs by genetic code expansion is a powerful method for modulating important biological regulatory mechanisms. From an applied point of view, our approach could be a valuable strategy for the development of photoactivable antibiotics against mammalian pathogens.

On the role of initial seeding of high throughput molecular dynamics for effective and accurate simulation of ligand transport processes in enzymes with buried active sites

Dheeraj Kumar Sarkar,[1,2] Jan Brezovsky[1,2]

[1]Laboratory of Biomolecular Interactions and Transport, Institute of Molecular Biology and Biotechnology, Faculty of Biology, Adam Mickiewicz University, Poznan, Poland, [2]International Institute of Molecular and Cell Biology, Warsaw, Poland

Many proteins have the active site buried in their hydrophobic core. These sites are then connected to the surface or bulk solvent by molecular transport paths known as tunnels or channels. Understanding the (un)binding processes of small molecules by utilization of transport pathways in facilitating ligand transport to the active site is of growing interest in the field of structure-based drug design and protein engineering. *Haloalkane dehalogenase* LinB belongs among typical representatives of enzymes featuring buried active sites. Recently, a mutant of this enzyme with markedly altered transport was constructed by closing a primary tunnel and designing a novel tunnel, creating a valuable model system to study the ligand (un)binding.

In the current study, we considered several approaches to construct ligand-protein complexes as effective input structures, also called seeds, for adaptive high-throughput molecular dynamic simulations guided by Markov state models. We designed four schemes for selecting the initial seed positions of ligands to start these simulations, ranging from randomly placed substrates in the bulk solvent to gradually more knowledge-based positioning into the locations relevant for the transport process. We then evaluated the robustness of various schemes and rigorously compared the rates connected with the usage of individual tunnels for ligand transport. The study enabled us to understand the non-trivial paths taken by transported molecules including exploration of high-affinity sites outside the tunnels. Overall, one of the schemes resulted in more kinetically meaningful Markov state models by effectively sampling the relevant transport paths cost-effectively, facilitating more efficient and accurate analyses to study ligand transport mechanisms in enzymes.

———————————————

This research was supported by POWR.03.02.00-00-I006/17 project; and National Science Centre, Poland (grant no. 2017/26/E/NZ1/00548). Calculations were performed at Poznan Supercomputing and Networking Center.

Designing Proteins to Protect Dyes in Light-Emitting Diodes

Horst Lechner, Gustav Oberdorfer

Institute of Biochemistry, Graz University of Technology, Austria

We want to employ artificial fluorescent proteins as down-converting filters for the emitting source of (blue) light-emitting-diodes (LEDs) to create white Bio-LEDs. The proteins are used as hosts for small organic compounds with fluorescent properties, similar to the ones used in organic LEDs (OLEDs). This should protect these compounds from photo-deactivation, which represents a central problem with these kinds of fluorescent molecules and often renders them less attractive for the use in white LEDs.

Our major aim in this multi-disciplinary project is to computationally identify and redesign proteins that can accommodate and specifically bind the compounds, while retaining high thermal stability. Additionally, the fluorescent compounds might be anchored into the binding pocket of the protein via covalent attachment using a non-canonical amino acid or a cysteine residue. To accomplish this, a suitable anchoring point in the pocket must be identified during design.

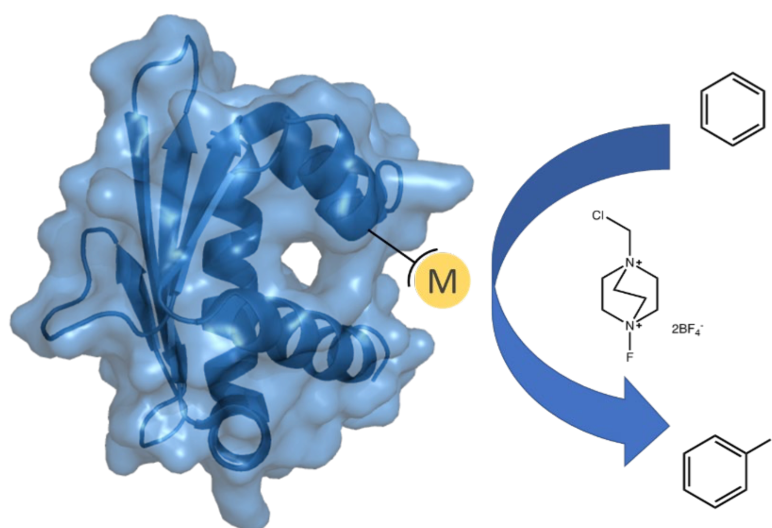
Finally, the derived proteins will be embedded in a polymer matrix to increase their stability against pH changes, temperature and the high photon-flux occurring when the blue LED-emitter is covered with this biological filter.

We currently have a few dozen designs for several of our envisaged dyes expressed and are evaluating their binding, spectroscopic and biophysical properties.

Development of New Artificial Metalloenzymes for Selective Fluorination

Isabeau Lüddecke[a], Amanda Jarvis*[a]

[a] *School of Chemistry, College of Science and Engineering, University of Edinburgh, Joseph Black Building, 40 David Brewster Road, EH9 3FJ, Edinburgh, UK.*
Email: amanda.jarvis@ed.ac.uk



To date, only one enzyme is known in nature that can catalyse the formation of a C-F bond. In 2002, the fluorinase of Streptomyces cattleya was discovered.[1] Since then no fluorinating enzymes other than fluorinase homologues have been identified, although it is suspected that there are other fluorinating enzymes.[2] Therefore, it is of great interest to develop new methods for enzymatic fluorination and to create an artificial fluorinase.

Previous work by the Jarvis group, demonstrating the catalytic activity of a newly developed artificial rhodium enzyme using a human sterol carrier protein, shows the ability of metalloenzymes to expand the toolbox of enzyme catalyzed reactions.[3] A similar approach has been explored for the development of an artificial metalloenzyme for fluorination using the same protein and bioconjugating it with different nitrogen- based ligands to bind copper and palladium salts, with the aim of performing regio- and enantioselective electrophilic fluorination in an aqueous environment.[4,5] In this poster, we will present our results to date on this project and future directions addressing the challenges related to fluorine chemistry in water.

1. Schaffrath,C.,etal.FEBSLett.547,111–114(2003).
2. Chen,Y.,etal.Org.Biomol.Chem.19,10081–10084(2021). 3. Deuss,P.J.&Jarvis,A.G.Supramol.Catal.547–560(2022).
4. Shibata,N.,etal.Synlett38,1153–1157(2007).
5. Yamamoto,K.etal.Nature554,511–514(2018)

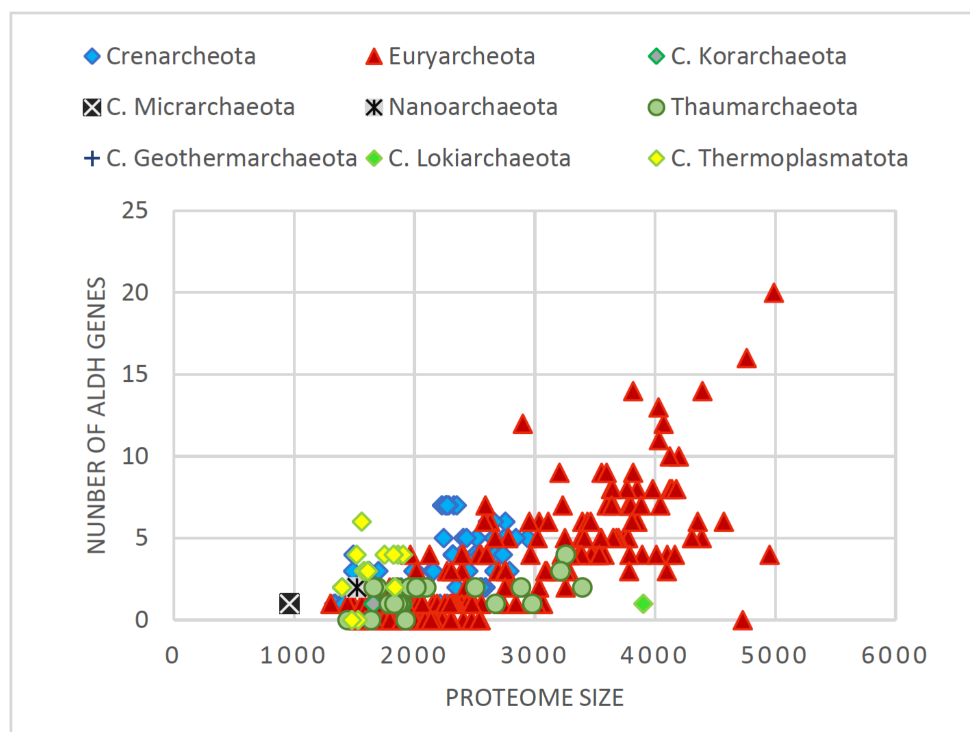
Natural history and functional diversity of aldehyde dehydrogenase superfamily in Archaea.

Danny Gabriel Madrigal-Ceballos[1], Gabriel Moreno-Hagelsieb[2], Héctor Riveros-Rosas[1]

[1]Departamento de Bioquímica, Facultad de Medicina, Universidad Nacional Autónoma de México, Ciudad de México. CDMX. MÉXICO. [2]Department of Biology, Wilfrid Laurier University, Waterloo, ON, CANADA. danny093@hotmail.com ; hriveros@unam.mx

Aldehyde dehydrogenases (ALDHs) comprise one of the most ancient protein superfamilies, it contains a large number of isoenzymes that perform a large amount of tasks that go from oxidative functions to non catalytic functions. Their members have been extensively studied in animals and plants, and their participation in diverse metabolic pathways using a broad variety of aldehyde substrates has been documented.

The ALDH superfamily has a wide phyletic distribution, being present in the three domains of life: Archaea, Bacteria and Eucarya. However, very few studies have been published characterizing ALDHs from archaea. Therefore, in this work a phylogenetic analysis of archaeal ALDHs was performed to get insights into the diversity and evolution of this protein family.



Using Blastp and Hmmer, we retrieved 1088 ALDH sequences from 365 complete genomes from archaea available at NCBI's RefSeq database. Of these genomes, 304 possess at least one *aldh* gene (average: 3.6 *aldhs*/genome).

We found that ALDHs in archaea grouped into 34 protein families, with the majority being present in Eury- archeota phyla (see Figure). Indeed, *Haloterrigena turkme- nica* DSM5511 possess 20 *aldh* genes. More than half of these ALDH sequences belong to only six ALDH families: ALDH11, ALDH18, ALDH21, and three unnamed ALDH families, identified in the CDD protein database as cd07100, cd07097 and cd07131. The broad phyletic distribution of ALDH18 suggest that this protein family could be the most ancient within ALDH superfamily.

Supported by UNAM-PAPIIT grants IN218819 and IN219022.

Enzyme catalysis prior to aromatic residues: reverse engineering of a dephosphocoenzyme A kinase

M. Makarov[1], A. K. Dunker[2], K. Hlouchová[1]

[1]*Department of Cell Biology, Faculty of Science, Charles University, Viničná 7, 128 00, Prague 2, Czech Republic.* [2]*Centre for Computational Biology and Bioinformatics, Indiana University School of Medicine, Indianapolis, Indiana, USA.*
**makarovm@natur.cuni.cz, klara.hlouchova@natur.cuni.cz*

It is well-known that the large diversity of protein structures and functions derives from the diverse physicochemical properties of the 20 canonical amino acids that comprise modern proteins. According to the generally accepted hypothesis [1], evolution of protein structures and functions was continuously associated with enrichment of the genetic code, with aromatic amino acids being considered the latest addition to the genetic code to increase structural sta- bility of proteins and enhance specification of their catalytic functions.

The main objective of this study was to test whether enzyme catalysis can occur without the aromatic residues (aromatics) by studying the structure and function of dephospho-CoA kinase (DPCK) following aromatic residue depletion. Two aromatics-lacking variants of a pu- tative DPCK from *Aquifex aeolicus* were designed by substituting (i) Phe, Tyr, Trp with Leu (DPCK-LH) and (ii) all aromatics (including His) by various non-aromatic amino acids based on the best predicted preservation of thermodynamic stability (DPCK-M).

Structural characterization of DPCK variants using far-UV CD spectroscopy, 1D NMR spectroscopy and limited proteolysis indicates that substituting the aromatics does not mark- edly alter their secondary structure but leads to a drastic loss of a firm globular arrangement implying their molten globule conformation [2]. Enzyme assays demonstrate that both mutant variants still possess ATPase activity, although with 150–300 times lower efficiency in com- parison with the wild-type phosphotransferase activity. The transfer of the phosphate group from ATP to the dephospho-CoA substrate becomes heavily uncoupled, and only the His-con- taining variant (DPCK-LH) is still able to perform the phosphotransferase reaction. Further investigation on ATP binding by DPCK-LH using steady-state fluorescence spectroscopy, dy- namic light scattering and 2D HN NMR spectroscopy revealed the formation of a tightly packed protein interior upon ligand binding.

These data support the hypothesis that proteins in the early stages of life could support catalytic activities, albeit with low efficiencies as a result of lack of firm hydrophobic core. For- mation of firm hydrophobic cores, which enable the assembly of stably structured active sites, is suggested to provide a selective advantage for addition of the aromatic residues. Before that, proteins would be more dependent on other structural scaffolds, such as small organic cofactors.[1]

1. *Wong Tze-Fei J. Coevolution theory of the genetic code at age thirty. Bioessays, 2005; 27(4):416–25.*
2. *Makarov M, Meng J, Tretyachenko V, Srb P, Březinová A, Giacobelli VG, et al. En- zyme catalysis prior to aromatic residues: Reverse engineering of a dephospho-CoA kinase. Protein Science, 2021;30(5):1022–34.*

---

[1] *This work was supported by the Czech Science Foundation (GAČR) grant number 17-10438Y and by the project BIOCEV (CZ.1.05/1.1.00/02.0109) from the European Regional Development Fund.*

Protein design of growth factor inhibitors

Kateryna Maksymenko[1], Patrick Müller[2], Julia Skokowa[3], Andrei Lupas[1], Mohammad ElGamacy[1,3]

See p. 27 (selected speaker)

Exploration of rare tunnels using Gaussian accelerated molecular dynamics simulations

Nishita Mandal1,2, Bartłomiej Surpeta1,2 and Jan Brezovsky1,2

[1]Laboratory of Biomolecular Interactions and Transport, Department of Gene Expression, Institute of Molecular Biology and Biotechnology, Faculty of Biology, Adam Mickiewicz University, 61-614 Poznan, Poland, and [2]International Institute of Molecular and Cell Biology in Warsaw, 02-109 Warsaw, Poland; Email: nisman@amu.edu.pl

1. Introduction

Enzymes catalyze various chemical reactions to maintain biological homeostasis through tightly regulated biochemical pathways. Understanding complexities involved in enzyme function requires thorough knowledge of catalytic mechanisms and structural features that render these reactions possible. The location of the active site largely determines the functions of the enzymes and many enzymes from six enzyme classes have buried active sites. The enzyme core is connected to the bulk solvent through molecular transport pathways called tunnels as shown in Figure 1. Moreover, the engineering of tunnel residues can produce mutants with enhanced properties and tunnels also represent very interesting drug targets. Since most of these tunnels are transient in their nature, the preferred method to study them is molecular dynamics (MD) simulations. However, given the rare opening of tunnels, their identification and analysis often require extensive and time-demanding sampling. Gaussian accelerated MD simulations (GaMD) [1], employing dual boost potentials to smooth biomolecular potential energy surfaces and reduce energy barriers, have been shown to overcome sampling limitations of classical MD (cMD) in many relevant biomolecular transitions. Therefore, we investigated to what extent GaMD models are suitable for the analyses of the rare opening of gates in transient tunnels.

2. Approach

We performed 5x 5µs cMD and GaMD simulations for three model systems, wild-type haloalkane dehalogenase (DEHAL) enzyme along with two mutants using AMBER18 package, following the recommended standards in the field. Tunnels were analyzed with the CAVER3 package [2] and TransportTools library [3] was applied to contrast the tunnel information gathered by CAVER3 from the cMD and GaMD simulation sets.

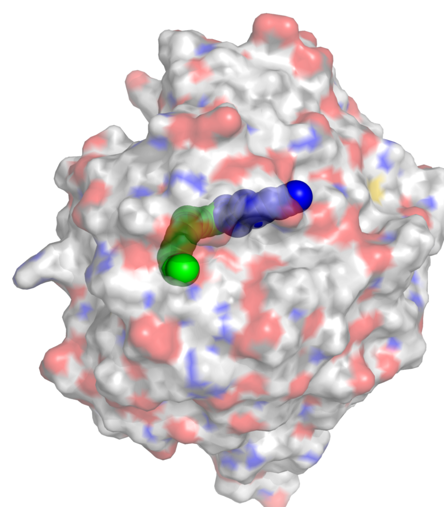


Figure 1. *Example of tunnels (the set of blue and green spheres) observed in proteins (color surface)*

3. Results

We have observed that GaMD simulations managed to identify all experimentally confirmed tunnels while maintaining similar geometries in agreement with cMD. Additionally, we also observed that there is an enhanced tendency for GaMD simulations to systematically explore tunnels, which were

observed only exceedingly rarely in the cMD simulations, but still on the biologically relevant time scales.

4. Discussion

Overall, GaMD approach to investigate tunnels in proteins that overcomes sampling limitations without compromising geometrical properties of these ligand transport pathways, and hence open up new possibilities for the identification of rare tunnels in large protein systems.

5. Acknowledgments

This research is supported by the National Science Centre, Poland (2017/26/E/NZ1/00548), and the calculations were performed at the Supercomputing and Networking Center.

---

1. *Miao, Y., Feher V., and McCammon JA (2015) Gaussian Accelerated Molecular Dynamics: Unconstrained Enhanced Sampling and Free Energy Calculation. J Chemical Theory and Computation, 11(8): 3584-3595.*
2. *Pavelka A, Sebestova E, Kozlíková B, Brezovsky J, Sochor J, Damborsk .J, (2016) CAVER: Algorithms for Analyzing Dynamics of Tunnels in Macromolecules. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 13, 505-517.*
3. *Brezovsky J, Thirunavukarasu AS, Surpeta B, Sequeiros-Borja CE, **Mandal N**, Sarkar DK, Dongmo Foumthuim CJ, Agrawal N, 2022: TransportTools: a library for high-throughput analyses of internal voids in biomolecules and ligand transport through them. Bioinformatics (Advance article DOI: 10.1093/bioinformatics/btab872).*

Trypanosoma cruzi mitochondrial peroxiredoxin (TcMPX) shares spatial distribution of active site with human peroxiredoxin 3

Ignacio Martinez, Lucio Rivera-Santiago, Rubén Arroyo-Olarte and Bertha Espinoza

Immunology Department, Instituto de Investigaciones Biomédicas, Universidad Nacional Autónoma de México. e-mail address: besgu@iibiomedicas.unam.mx

Trypanosoma cruzi is the parasite that causes Chagas disease, which according to WHO data affects 6-8 million people in the world. This protozoan has several enzymes to combat reactive oxygen species (ROS), both own and host. One of these proteins, mitochondrial peroxiredoxin (TcMPX), has been proposed as a parasite virulence factor and identified as a potential therapeutic target. However, no drugs have been proposed to inhibit this enzyme so far. This is partly due to the absence of basic models of its structure and active site arrangement. Therefore, in the present work, the TcMPX of T. cruzi Mexican strain Ninoa was sequenced (GenBank QKE53460.1), aligned with known peroxiredoxins, and modeling by homology using Swiss-Model platform. The model was validated by package SAVES v6.0, and visualized by PyMol Ver. 2.4.0. TcMPX was found to have more than 60% similarity to human peroxiredoxin 3 (PRX3). Furthermore, when comparing the model of TcMPX (pink) with the human PRX3 (PDB 5JCG, green color), it was found that the spatial arrangement of both proteins was very similar (Figure 1).

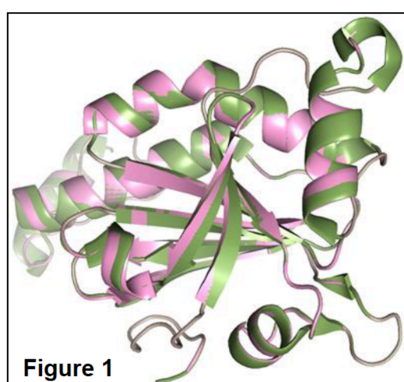


Figure 1

Likewise, the position of the peroxidatic (CP, red spheres) and resolutive (CR, yellow spheres) cysteines of the active site are similar between both proteins (Figure 2). Also, the amino acids of the so-called catalytic triad (cyan) and those possibly involved in resistance to hyper oxidation (blue) have a similar spatial distribution between both proteins. These results suggest that TcMPX could be susceptible to being inhibited by molecules such as Thiostrepton, which have been shown to inhibit the activity of human PRX3. If this is verified, the use of these drug for the treatment of Chagas disease could be proposed.

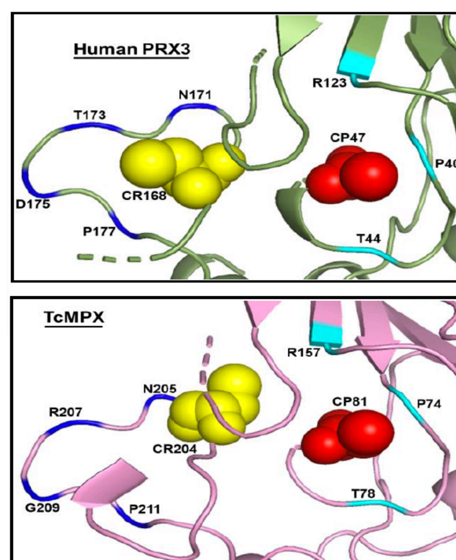


Figure 2

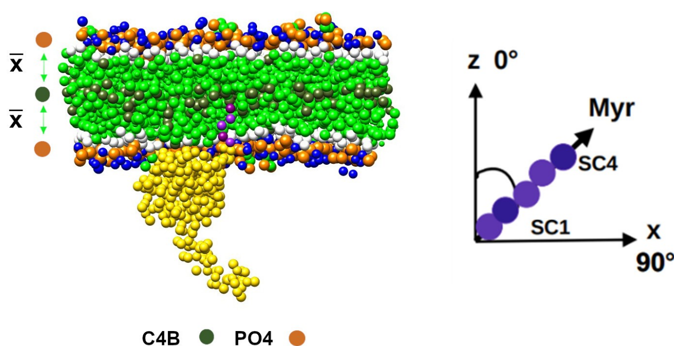
The authors acknowledge the support of DGAPA-UNAM (IN206620) and NUATEI- UNAM Institutional Program

HIV-1 Gag matrix domain interaction with plasma membrane from coarse-grain molecular dynamics simulations

P. Mendoza-Espinosa*,[1], J. Quintana-H[1], J.C. Armas-Perez[2] , O. Guzmán[3], G. A. Chapela[3]

[1]*Instituto de Quiímica, UNAM, Cd. de México, México.*
[2]*División de Ciencias e Ingenierías, Universidad de Guanajuato, Loma del Bosque 103, colonia Lomas del Campestre, CP 37150, León, Guanajuato, México.*
[3]*Departamento de Física, Universidad Autónoma Metropolitana, Iztapalapa, 09340, Cd. de México, México. E-mail:* *mendozapaola1@gmail.com

We performed coarse-grain simulations with explicit solvent to analyze the role of the myristoyl moiety in the matrix (MA) domain of the HIV-1 capsid protein Gag when interacting with three different models of lipid bilayer. These models were specifically chosen because they mimic the composition in the mammalian inner plasma membrane, lipid-raft domains, and a cholesterol-free bilayer. The inner plasma membrane is the target for the Gag protein when forming the HIV-1 capsid, so the other two models allow for investigation of bilayer composition effects on the interaction between the MA domain and the lipid bilayer. The simulation trajectories of these coarse-grained model allowed sampling over several hundreds of microseconds, in contrast to recent atomistic-detail simulation trajectories that can cover a few microseconds. First, we measured the probability distribution associated with position and orientation of the myristoyl and MA protein domain interacting with the plasma membrane models. In addition, we quantified the degree to which the myristoyl moiety aligns to the normal of the membrane surface. In order to realize both, we quantified the orientational order in these two measurements and determine the potential of mean force and torques associated with the observed probability distributions. Due to the longer sampling time, we achieved much improved statistical uncertainty of these potentials, allowing us to determine that their corresponding binding-energy values are on the order of 8 kT , or equivalently to ~ 5 kcal mol^{-1} at physiological temperature. We show that the interaction of side chains of residues in MA and the inner-membrane in our simulations is similar to previous experimental reports; some of these residues are key in the interaction of RNA molecules with MA, therefore these interactions are important milestones for the coarse-grained models presented here. Our method can be applied to other acylated viral proteins interacting with the plasma membrane, not only for HIV-1, but other retroviruses and enveloped viruses.



A) Leaflet membrane thickness and myr position at lipid membrane. The average coodinate z, normal to the membrane surface, of the center of mass of PO4 groups (orange) and C4B of the acyl chains (dark green) was used to find the edge and center of the membrane for each leaflet of the bilayers.

B) For the myr position, extension and orientation, we located the centers of mass of SC1 and SC4, corresponding to the first and last carbons in the myr chain, and defined the azimuthal angle Θ between the membrane normal and the vector pointing from SC1 to SC4.

Design of a T cell receptor targeting the NY-ESO-1$_{157-165}$ epitope using Rosetta

Möller, J.[1]; Schoeder, C. T.[1]

[1]*Institute for Drug Discovery, Medical Faculty, University of Leipzig, Brüderstraße 34, 04103 Leipzig, Germany*

Multiple myeloma (MM) is a hematologic malignancy. It originates from post germinal-center B cells with monoclonal proliferation, antibody production and end-organ damage [1]. Emerging treatment options utilize targeted personalized immunotherapy, with CAR T cell therapy showing promising results in MM [2]. In contrast to antibodies targeting soluble and surface proteins, T cells can identify intracellularly processed antigens presented on major histocompatibility complexes (MHCs). MHC class I molecules are expressed by all nucleated human cells and present intracellular proteins in the form of small peptides of 8-10 amino acids in length. The peptide MHC complex (pMHC) builds the T cell epitope that is recognized by T cell receptors (TCRs). If a cell is infected or malignant and produces altered or foreign proteins, the "non-self" T cell epitope is detected by T cells resulting in an immune reaction [3]. Typically, TCRs, although effectively targeting pMHCs, bind with low affinities [4]. To increase affinity of a TCR to the T cell epitope SLLMWITQC, a peptide derived from the NY-ESO-1 protein and known to be presented on MM cells, structure-based protein design was conducted using the Rosetta software suite. Input structures were obtained from the PDB (PDB-IDs 2BNR, 2F53, 2F54, 2P5E, 2P5W, 2PYE). All structures included the MHC allele HLA-A*0201 bound to SLLMWITQC and one out of five TCRs. Designable residues of the TCR were selected as being within 5 Å of the pMHC. The binding energy was determined for the whole interface as well as between the TCR and the peptide or MHC, respectively. Sequences of the best scoring models were analysed. The binding energy between TCR and pMHC was increased through protein design. Detailed analysis revealed this was mostly due to increased binding affinity of the TCR to the MHC, whereas the affinity of the TCR to the peptide could only be increased in very few models, contributing little to the overall observed effect. This has previously been shown in regards to TCR affinity maturation and poses the problem of off-target cross reactivity. A possible solution is the design of TCR-like antibodies, harnessing the structural homology between antibodies and TCRs and the generally high affinity and specificity of antibodies [4].

1. *Jurczyszyn, A, Suska A: Encyclopedia of Biomedical Gerontology. Elsevier 2019,461-478.*
2. *Wang, L et al.: Int J Cancer 2015, 136(8):1751-1768*
3. *Neefjes, J et al.: Nat rev Immunol 2011, 11(12):823-836*
4. *Steward-Jones, G et al.: PNAS 2009, 106(14):5784-5788*

Finding rationale for diversity; associative and independent Cas4s

Devanjan Mondal[1], Manisha Goel[1*]

[1]*Department of Biophysics, University of Delhi-South Campus, New Delhi, India*

CRISPR-associated (Cas) protein systems are crucial parts of the CRISPR-Cas based adaptive immune systems in prokaryotes. These proteins function to mediate adaptation, expression, and interference processes, with the constitutive proteins differing and therefore resulting in different types of CRISPR-Cas systems. Although, two such proteins, Cas1 and Cas2 are found conserved across most systems, and are reported to function in the adaptation module. Cas4 though not ubiquitously present, is also noted to be functional in the adaptation step. However, the Cas4 proteins among other such proteins of the CRISPR-Cas system are reported to show extensive diversity in their existence as Solo-Cas4s and Cas-Cas4s occurring together in a cell, with no existing correlations to define such occurrences or the co-existence of such systems towards the CRISPR-Cas process. Such trends in diversity are more prevalent in Archaea as compared to bacterial systems. Our work is an attempt to make sense of Solo Cas4s and Cas-Cas4s in archaeal systems, and towards finding reasons for such diversification. Through the use of several in-silico tools, we have performed cataloging, phylogenetic analyses, motif analyses, and structure-based analyses, therefore underlining comparative studies and definitions to the nature of Solo-Cas4s and Cas-Cas4s.

MAPPING HOW GTPASES CATALYZE GTP HYDROLYSIS – RESOLVING A DECADES OLD CONTROVERSY

Cátia Moreira[1*], Ana Rita Calixto[1], Anna Pabis[2], Carsten Kötting[3], Klaus Gerwert[3], Till Rudack[3], Shina Carolyn Lynn Kamerlin[1]

[1] Department of Chemistry-BMC, Uppsala University, Box 576, S-751 23 Uppsala, Sweden.
[2] Department of Cell and Molecular Biology , Uppsala University , BMC Box 596, S-751 24, Uppsala , Sweden. [3]Department of Biophysics, Ruhr University Bochum , 44801 Bochum , Germany. e-mail: catia.moreira@kemi.uu.se

GTP hydrolysis is a biologically crucial reaction, participating in the regulation of almost all cellular processes[1]. Because of this, the enzymes that catalyze this reaction, GTPases, are among the most important drug targets, as well as being in the center of decades of substantial research effort[2]. However, the fundamental mechanism of both non-enzymatic and enzyme-catalyzed GTP hydrolysis remains highly controversial. In the case of the enzyme-catalyzed reaction, key questions include trying to understand how GTPases hydrolyze GTP without an obvious general base present in the active site, what the preferred reaction pathway and nature of the corresponding transition states is, and how the interaction of GTPases with their corresponding GTPase activating proteins (GAP) enhances the catalytic rate of GTP hydrolysis. These continued uncertainties hamper both our fundamental understanding of enzymatic GTP hydrolysis, as well as ongoing drug discovery efforts.

To answer these questions, we recently performed empirical valence bond simulations of GTPase-catalyzed GTP hydrolysis, comparing different reaction pathways in three distinct GTPases, Ras, Rab, and the $G_{\alpha i}$ subunit of a heterotrimeric G-protein, both in the presence and in the absence of the corresponding GAP[3]. We propose that in all systems, GTP hydrolysis preferentially proceeds through a conserved 'solvent-assisted' pathway, involving the rate-limiting nucleophilic attack of a water molecule, and leading to a short-lived intermediate that then tautomerizes to form H_2PO^{4-} and GDP as the final products. Our results also provide insight into the forces that drive the enhancement of the intrinsically slow catalytic rates of GTPases upon binding to their specific GAP. Our work thus provides fundamental biochemical insights into the enzymatic regulation of GTP hydrolysis[3] that not only resolves a decades-old mechanistic controversy, but also has high relevance for drug discovery efforts.

*1) S.C.L. Kamerlin, P. K. Sharma, R. B. Prasad, A. Warshel, Q. Rev. Biophys. **2013**, 46, 1-132*
*2) J.K. Lassila, J.G. Zalatan, D. Herschlag. Annu. Rev. Biochem. **2011**, 80, 669-702*
*3) A. R. Calixto, C. Moreira, A. Pabis, C. Kötting, K. Gerwert, T. Rudack, S. C. L. Kamerlin. J. Am. Chem. Soc. **2019**, 10, 141 (27)*

RGS1 – G(i) alpha complex formation – correlating DEER data to structure dynamics using computational modeling and MD simulations

Victoria Most[1], Hope Woods[2], Kevin Eden[3], Davide Sala[1], Mathias Elgeti[1,3], Jens Meiler[1,2]

[1]Leipzig University Medical School, Institute for Drug Discovery, Leipzig, Germany;
[2]Center for Structural Biology, Vanderbilt University, Nashville, TN, USA.;
[3]Department of Chemistry and Biochemistry, University of California, Los Angeles, CA , USA.

G-protein coupled receptors (GPCRs) are the largest class of human membrane proteins. Their ability to transmit a signal through the cell membrane makes them important drug targets. Regulators of G protein signaling (RGS) are proteins that can bind to activated G alpha subunits and accelerate GTP hydrolysis, resulting in a negative modulation of GPCR signaling[1]. They pose an interesting target for small molecule inhibitors to enhance the GPCR signaling response. In this study we focused on the RGS1 protein for which an X-Ray structure is available as a monomer and in complex with the G_i alpha subunit[2]. Using double electron-electron resonance (DEER) distance measurements we investigated the dynamics of the RGS1-G_i alpha complex formation. In contrast to the information from available structures, RGS1 seems to occupy multiple distinct conformations in its monomeric and bound form. We used a combined approach of molecular dynamics simulations and novel molecular modeling techniques to determine conformations that are in correspondence with the measured DEER distances.

---

1 Berman, D. M., Kozasa, T. & Gilman, A. G. The GTPase-activating Protein RGS4 Stabilizes the Transition State for Nucleotide Hydrolysis. J. Biol. Chem. ***271****, 27209–27212 (1996).*
2 Soundararajan, M. et al. Structural diversity in the RGS domain and its interaction with heterotrimeric G protein α-subunits. Proc. Natl. Acad. Sci. U. S. A. ***105****, 6457-6462 (2008).*

Consolidation and evaluation of the PocketOptimizer software for computer-aided design of ligand-binding proteins

Jakob Noske[1], Josef Kynast[1], Birte Höcker[1]

[1]*Biochemistry III, University of Bayreuth*

The engineering of customized proteins that can perform specific tasks in living systems is of great interest. Computational methods are used to narrow down the immense combinatorial sequence space to find the best solution, and thus can provide starting points for experimental procedures. However, success rates of computational techniques strongly depend on the accurate modelling of protein conformation and its energetic evaluation. The design of sensitive and specific ligand-binding proteins for biotechnological or biomedical applications is particularly challenging, as not only intra- but also intermolecular interactions have to be taken into account. To tackle this problem, we developed the software PocketOptimizer. PocketOptimizer allows to predict mutations in the binding pockets of proteins that increase the affinity to a given ligand. Here we present the latest version of PocketOptimizer, version 2.2, which was improved in various aspects. We implemented a cleaner user interface, a modular architecture with even more supported options such as force fields and scoring functions, a backbone-dependent rotamer library, as well as different improvements in underlying algorithms. Version 2.2 of PocketOptimizer was evaluated extensively against the original benchmark set of design cases as well as a new and improved benchmark set. Initial results indicate a similar performance of version 2.2 as before while providing a significant advantage in processing time. Additionally, newly implemented modules such as the new rotamer library result in improved prediction accuracy within shorter time. PocketOptimizer 2.2 provides a robust and versatile environment for the design of small molecule binding pockets in proteins that can be widely applied and expanded further due to its modularity.

Experimental evolution of transporter specificity under a multitude of selective pressures

Sebastian Obermaier, Bert Poolman

Department of Biochemistry, Groningen Biomolecular Sciences and Biotechnology Institute, University of Groningen, Netherlands

Natural evolution of proteins is marked by diversification. Members of a protein family experience diverse selective pressures, typically yielding proteins with vastly different substrate profiles. Yet, laboratory-based experimental evolution is so far not able to provide a model system that captures such substrate diversification. In this project, we are setting up a platform for in-vivo evolution of the substrate specificity of amino acid transporters.

In-vivo experimental evolution

A *Saccharomyces cerevisiae* strain that is deficient in amino acid transport is the base of the platform. Transport of individual natural amino acids can be restored by complementation with a specific transporter, e.g., a proline transporter. This allows the strain to use proline as the sole nitrogen source. Targeted *in vivo* mutagenesis of the proline transporter then leads to sequence variation. Variants with new specificities can be selected by changing the nitrogen source in the growth medium to other amino acids. The exceptional opportunity here is that there are 20 amino acids that can sustain growth of yeast, if they are transported. This creates the possibility to challenge the transporter with a multitude of substrate pressures.

How will the transporters adapt to these substrate pressures? Can a proline transporter evolve to transport tryptophan? Can a promiscuous transporter become specific for one amino acid? Will similar substrates yield similar transporter variants? The answers will be essential for understanding diversification of protein families in nature, a ubiquitous phenomenon that is difficult to capture with traditional experimental evolution approaches.

Structural and functional characterization of DNA/RNA methyltransferase specificity

○Yoshiki Ochiai, Paola Laurino

(Protein Engineering and Evolution Unit, OIST)

M.EcoGII is the only known bacterial methyltransferase that catalyze non-site- specific N6-adenine methylation on both DNA and RNA. However, the evolutionary origin of this unusual promiscuity is still unknown. Here, we performed ancestral sequence reconstruction to investigate the evolutionary origin of sequence specificity in M.EcoGII. Five ancestors of M.EcoGII were selected from a maximum-likelihood phylogeny for ancestral sequence reconstruction. These ancestral enzymes were expressed in E. coli, purified and characterized. In vitro methylation assays showed that all ancestors had a higher activity for RNA than M.EcoGII, with Anc284 showing a significant increase in selectivity for RNA. Multiple sequence alignment indicated the amino acid residues in charge of substrate specificity. Structural models of the ancestral proteins and multiple sequence alignment predicted that amino acid residues in charge of functional switch were located on the catalytic domain. The results in this study will contribute to the understanding of the molecular mechanism of selective activity between RNA or DNA methyltransferases. In addition, the RNA selective M.EcoGII ancestor would be applicable to the epigenetic tools for changing RNA methylation states.

Exploring the protein space of poly(ethylene) terephthalate depolymerases and the emergence of PET degrading activity

Liliana Oliveira[1] and Bruce Lichtenstein[1]

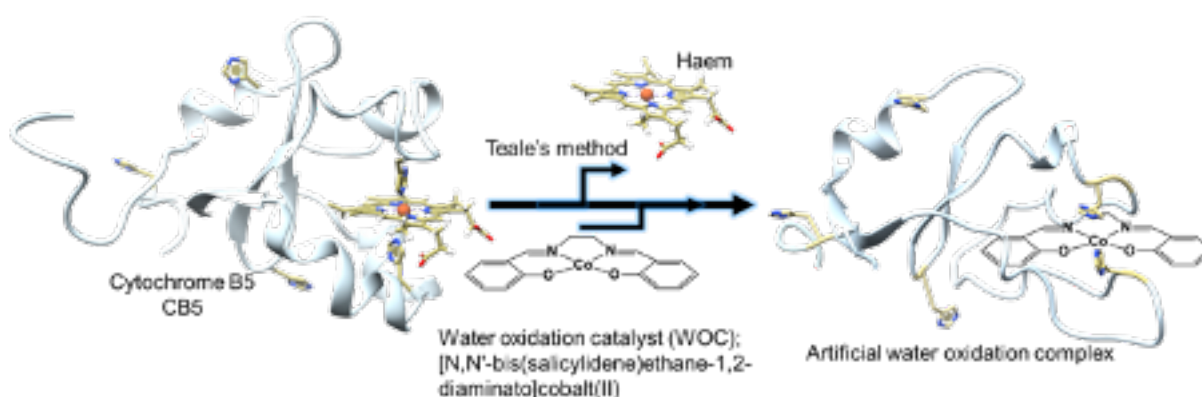
[1]Centre for Enzyme Innovation, School of Biological Sciences, University of Portsmouth, UK

Plastics have become indispensable in today's society owing to their durable and varied physical properties, yet, the same properties that make these materials so desirable also make them recalcitrant pollutants allowing their accumulation in the environment. Natural enzymes however have the ability of depolymerising certain plastics, such as poly(ethylene) terephthalate (PET), and are a potential biological alternative to the current, and often open-loop, solutions for plastic waste management. In this study, bioinformatics approaches are used to explore the protein sequence space occupied by known PETases, revealing that PET-degrading activity is not an evolved function, but rather a result of catalytic promiscuity. Ancestral sequence reconstruction is used to calculate putative ancestors of extant PET depolymerases, which will be resurrected and tested for PET-degrading activity in the laboratory. The resurrected enzymes are expected to have properties, such as a higher thermodynamic stability, more desirable for the enzymatic recycling of post-consumer plastics.

Photo-activated water oxidation catalysis in an artificial metallo-protein

Laura Opdam[a], Ehider Polanco Rivas[b], Boyd de Regt[a], Cas Bakker[a], Nicole Lambertina[a], Sylvestre Bonnet[b], Anjali Pandit[a]

[a]SSNMR/BPOC and [b]MCBIM departments, Leiden Institute of Chemistry, Leiden University, Einsteinweg 55, 2333 CC Leiden, The Netherlands



Photoactivated water oxidation is a clean source of electrons in the production of H_2 or valuable chemicals, avoiding use of fossil fuels. Water oxidation can be catalyzed by synthetic metal complexes. The turn over frequency of these complexes has improved significantly over time as a result of changes to the first coordination sphere, the turn over number however, has stayed behind. In nature, metalloproteins function as a second coordination sphere to their bound metal complex. Our goal is to design protein-based water-oxidation catalyst complexes, potentially offering a stabilizing environment in which water oxidation catalysis is tuned by the second coordination sphere. We selected 3 proteins: HasAp, cytochrome B5, and myoglobin to axially coordinate synthetic catalysts in their binding pocket. A library of ruthenium- and cobalt- based transition catalysts were chosen to bind to these proteins. An efficient way of screening for promising combinations was therefore developed; semi-native PAGE allows for the detection of complex formation by separation based on a difference in structural stability. Electron spray ionization mass spectrometry was used to determine the binding stoichiometry of the formed complexes. NMR, using histidine side chain focused HSQC combined with site directed mutagenesis, was finally used to identify the binding locations of a promising catalyst. We are currently testing the activity and stability of our protein-catalyst complexes and are looking forward to optimization via mutagenesis.

Conformational selection at play in the assembly of a Cre intasome

Marco Ramírez and Nina Pastor

*Centro de Investigación en Dinámica Celular, IICBA, UAEM, México *nina@uaem.mx*

Cre is a tyrosine site-selective recombinase that recognizes 34 bp loxP sites in dsDNA. Each loxP site is bound by two Cre, and these assemble with another loxP-Cre complex to form an intasome. Cre is a 343 amino acid protein composed of two domains: an N-terminal CBD domain comprising helices A to E, and a C-terminal CAT domain formed by helices F to N and a three-stranded beta sheet; the catalytic residues are located in the CAT domain and the CBD domain contributes to the binding energy to loxP through the interaction of helices B and D with the major groove [1].

NMR analyses of Cre have shown that there are few interactions between the two domains in the absence of dsDNA [2], justifying the study of each domain separately. The CBD domain has a disordered N-terminal tail of ~20 amino acids, which can be deleted without loss of function. We simulated this domain from amino acid 21, the beginning of helix A, to the end of helix E, in four independent 5 µs molecular dynamics runs, starting from the same structure (PDB 1Q3U), in explicit solvent using CHARMM36m [3] in Gromacs [4]. In the absence of quaternary contacts, helix A samples three different positions on the surface of the core of the domain, formed by helices B, C, and D. Helix E tends to unravel at the C-terminus, but otherwise remains in the same position as in the complete intasome. Comparison of the angle formed between helices A and B in the simulations and in the available structures in the PDB reveals that one of the angles corresponds to that found in intasomes (~30 degrees), while another one is found in recently released structures of Cre monomers and dimers bound to loxP (~330 degrees) [5]. The other conformation located at ~75 degrees has not been found in any structure yet.

The CAT domain exists in active and inactive conformations. It is stabilized in the intasome by swapping the C-terminal helix N with the neighboring CAT domain, cradled by a cavity in its surface; this is a *trans* interaction [1]. Recent studies of the isolated CAT domain [2] proposed the existence of a self-inhibitory *cis* complex, where the helix N binds at the surface of the active site, opposite to the *trans* binding site. We simulated three 5 µs runs of this domain starting from PDB 1Q3U in six conditions: active and inactive conformations, and with helix N in *cis* (based on the PRE data in [2]), in *trans* as an independent peptide, and without helix N. We found a hinge between two subdomains, one formed by helices F to J and the other by helices K to M. The docking site for helix N in trans lies at this hinge, and the active site lies at the opposite face of the hinge. In the absence of helix N, the two subdomains rotate and separate from each other, sampling a large variety of conformations, especially in the simulations starting from the active state. Helix N, whether in *cis* or in *trans*, severely restricts the motion of the domain to the vicinity of the starting conformation, representative of all available intasomes.

---

[1] Meinke, G, et al. (2016). Chem. Rev. 116:12785–12820; [2] Unnikrishnan, A, et al. (2020) PNAS 40:24849-24858; [3] Huang, J, et al. (2016) Nat. Methods 14:71–73; [4] Hess, B, et al. (2008) J. Chem. Theory Comput. 4:435-447; [5] Stachowski K, et al. (2022) Nucleic Acids Res. 50:1753-1769

We acknowledge supercomputing time at Laboratorio Nacional de Supercómputo del Sureste de México (grant 202101023N) and at LANCAD (grants in 2020 through 2022).

Exploring the dark areas of the natural protein universe for unknown protein families

Joana Pereira[1,2], Torsten Schwede[1,2]

[1]*Biozentrum, University of Basel, Basel, Switzerland.*
[2]*SIB Swiss Institute of Bioinformatics, Biozentrum, University of Basel, Basel, Switzerland.*

The collection of all natural protein sequences is known as the "protein universe", a multidimensional-dimensional space where individual proteins correspond to points whose coordinates are defined by their evolutionary relationship to others. In this universe, protein families form galaxies and protein superfamilies form clusters of galaxies, surrounded by empty, dark areas seemingly unexplored by Nature. However, the increasing feasibility of large-scale genomic projects is promoting an exponential increase in the number of protein sequences deposited in protein repositories every year. These are automatically organized and linked to others well-studied on the basis of homology (e.g., UniRef clusters and automated annotation of protein domain families and sequence features) but the number of so-called "hypothetical proteins" and "proteins of unknown function" is increasing proportionally. This can be due to low sensitivity of the methods behind it, but also due to the presence of sequences belonging to novel not hitherto described biological systems populating dark areas of the protein universe that could be explored in biotechnology, biomedics or pharmaceutical research.

In 2009, Levitt estimated that more than 20% of all non-redundant proteins known to date could not be mapped to an already known family. As a first step, we revised this hypothesis and analysed all active non-redundant sequences in UniProt and UniParc, classifying them into dark based on the current annotation level of their closer homologs as by UniProt and InterPro. This allowed us to revise and study the proportion and taxonomic distribution of sequences and UniRef clusters that are poorly annotated for domains and sequence features (i.e., predicted coiled coil and disordered regions), but also analyse individual proteomes for their darkness content. We found that almost 25% of the sequences, covering about 40% of all UniRef50 clusters, have a maximum of 5% annotation coverage, corresponding to proteins without clear homology to any other group of known sequences, i.e. they are seemingly dark. These sequences are widespread in the tree of life but the most common are from marine sediment metagenomes. Looking specifically at the proteomes available through the AlphaFold2 database, pathogens as *Staphylococcus aureus* and extremophiles as *Methanocaldocus janischii* are those with the highest frequency of proteins with no clear homology to well-known families or specific sequence features, which AlphaFold2 now helps to enlighten.

The next step is to evaluate how many of these correspond to novel families and not just divergent forms out of the scope of the current annotation methods. Preliminary analysis using deep learning protein language models suggest that such dark proteins complement the landscape occupied by those with at least one well annotated homolog (i.e., bright proteins). While most sequences seem to form bridges between well studied areas, some sequences were found to form unique, well-delimited clusters, suggesting they may correspond to independent novel families unrelated to any other previously described.

Physicochemical characterization of ancestral triosephosphate isomerases obtained by different reconstruction methods.

Pérez-Niño, Jorge Alejandro[1]; Guerra Borrego, Yasel[1,2], Rodríguez Romero, Adela[3]; Fernández-Velasco, D. Alejandro[1].

See p. 32 (selected speaker)

Allosteric network analysis in galectin-7 uncovers key residues controlling positive cooperativity between two opposite glycan binding sites

N. T. Hang Pham[1], Alex Paré[1], Myriam Létourneau[1], Marlène Fortier[1], Gabriel Bégin[3], David Chatenet[1], Yves St-Pierre[1], Patrick Lagüe[2,3], Charles Calmettes[1,3] & Nicolas Doucet[1,3]*

[1]Centre Armand-Frappier Santé Biotechnologie, Institut National de la Recherche Scientifique (INRS), Université du Québec, Laval, QC, Canada. [2]Département de Biochimie, de Microbiologie et de Bio-informatique and Institut de Biologie Intégrative et des Systèmes (IBIS), Université Laval, Québec, Canada. [3]PROTEO, the Québec Network for Research on Protein Function, Engineering, and Applications, Université Laval, Québec, QC, Canada.
**Corresponding author: nicolas.doucet@inrs.ca*

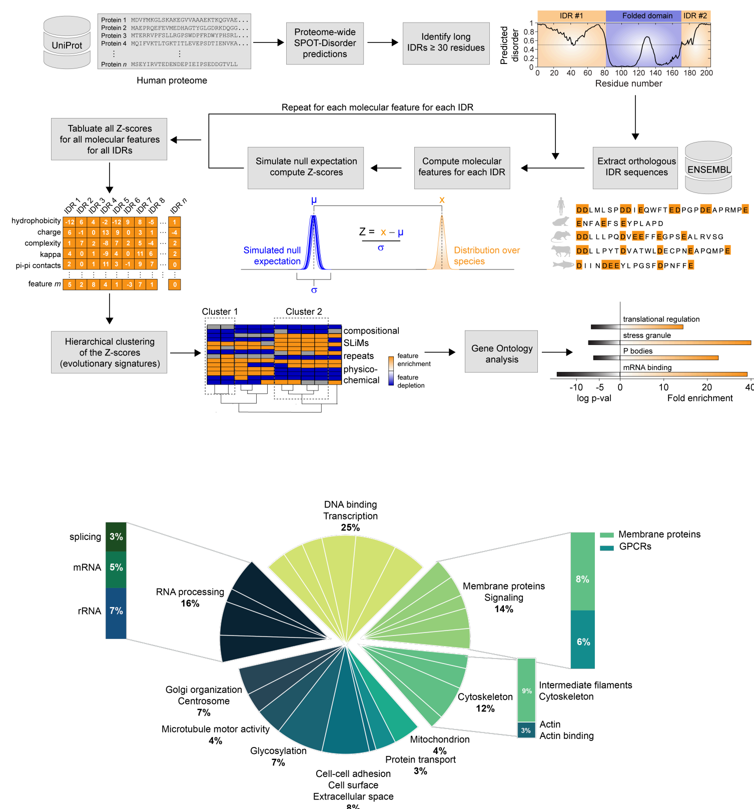
Triple-negative breast cancer (TNBC) cells abnormally express galectin-7 (GAL-7), leading to metastasis. GAL-7 is a prototypical galectin characterized by a β-galactoside binding site (GBS) and a homodimeric molecular architecture. It binds to glycoreceptors on the surface of activated T cells via its GBS and triggers T cell apoptosis, favoring tumor cell immune evasion. Consequently, GAL-7 represents a promising target for TNBC treatment. To this day, the development of galectin modulators has almost exclusively focused on glycan-based inhibitors aimed at perturbing glycoreceptor interactions. However, due to high GBS similarity among several galectin homologs in the cell, this remains a high-risk strategy because of unwanted off-target effects on other beneficial anti-tumoral galectins. Furthermore, GBS inhibitors are ineffective at targeting glycan-independent functions of GAL-7. New approaches are thus required to develop effective and highly specific GAL-7 inhibitors. Our previous study showed that perturbing GAL-7 dimer interactions alters its immunosuppressive activity. We further used dynamical network analysis to model the first allosteric communication network across the dimer interface and between opposite protomers. However, it is unclear which residue networks are important to control positive cooperativity between opposite GBSs. In this study, we uncover how crucial residue contacts at the dimer interface can be altered to control the biological function of GAL-7 by allosterically modulating communication between protomers. We show that introducing a covalent disulfide bridge at position G16 strengthens protomer interactions to improve positive cooperativity between GBSs by favoring interprotomer communication between residue pair Arg20-Asp103 at the dimer interface. Accordingly, we describe a new approach to investigate the global communication flow between two GBSs, revealing that altered electrostatic interactions between residue pairs Arg20-Asp103 and Arg22-Asp103 reduces communication between the two distant GBSs. This is consistent with decreased pro-apoptotic activity in GAL-7 variants R20A, R22A, and D103A, further suggesting that Arg20, Arg22, and Asp103 are potential residues controlling positive cooperativity between the two GBSs. Our study illustrates how the shortest pathway analysis can illuminate global flow communication to reveal key residues controlling allosteric communication between long- range functional sites within a protein.

A Functional Map of the Human Intrinsically Disordered Proteome

Iva Pritišanac[1,2], Taraneh Zarin[1,4], Khaled Elemam[1], T. Reid Alderson[3], Aqsa Alam[1], Alex Lu[1,5], Julie D. Forman-Kay[2,3]* & Alan M. Moses[1]*

1. Department of Cell and Systems Biology, University of Toronto, Toronto, Canada
2. Program in Molecular Medicine, The Hospital for Sick Children, Toronto, Canada
3. Department of Biochemistry, The University of Toronto, Canada
4. The Centre for Genomic Regulation, Barcelona, Spain
5. Microsoft Research Lab – New England, Boston, USA

Intrinsically disordered protein regions (IDRs) represent at least one third of the human proteome and defy the traditional structure-function paradigm. Because IDRs often show low positional sequence conservation, the functional classification of IDRs using standard bioinformatics can provide only limited insight. It has previously been shown that the bulk molecular features of IDRs can be conserved in the absence of positional sequence conservation.



Here, we show that evolutionarily conserved bulk molecular features of the intrinsically disordered human proteome (IDR-ome) enable classification and functional predictions. Clustering of the human IDR-ome reveals strong enrichment for frequently studied functions of IDRs in transcription (~25% of the

IDR-ome) and RNA processing (~16% of the IDR-ome), as well as a full range of other diverse functional annotations ranging from subcellular localization and proteinaceous membraneless organelles (PMLOs) to the constitution of the cytoskeleton and transmembrane signalling. We show that using feature conservation of IDRs alone allows prediction of protein localization to PMLOs. Clustering of the IDRs encoded by disease-causing genes and signatures of their conserved molecular features can explain the impact of some of the mutations that map to IDRs, which allows the formulation of testable hypotheses for experimental studies. Our results show that the conserved molecular features of IDRs are predictive of biological function, subcellular localization, and in some cases can rationalise disease-causing mutations in IDRs.

Analysis of the three subfamilies of ATG18-autophagy protein and particular focus on PvATG18b in *P. vulgaris* (Common bean)

Elsa H Quezada-Rodríguez[1], Homero Gómez-Velazco[2] , Manoj-Kumar Arthikala[1], Miguel Lara[3], Kalpana Nanjareddy[1]

1. Ciencias Agrogenómicas, Escuela Nacional de Estudios Superiores Unidad León, Universidad Nacional Autónoma de México (UNAM), León C.P. 37684, Mexico
2. Instituto de Química, Universidad Nacional Autónoma de México (UNAM), Cuidad Universitaria, Cuidad de Mexico C.P. 04510, Mexico.
3. Departamento de Biología Molecular de Plantas, Instituto de Biotecnología, Universidad Nacional Autónoma de México (UNAM), Cuernavaca C.P. 62271, Mexico

Autophagy is a fundamental degradation process in eukaryotes. Macroautophagy is a type of autophagy that involves the autophagy genes (ATG). The ATGs allow the formation of double membrane vesicles called autophagosomes that engulf damaged organelles, toxins, and damaged proteins to be recycled by vacuoles to maintain cellular homeostasis. The manipulation of processes shows promise to improve crop productivity under diverse stress (Thanthrige et al., 2021). We are interested in legumes for their nutritional content and their contribution to sustainable agriculture. The legumes can develop symbiosis, a relationship with *Rhizobium* and *Mycorrhiza*. In the way to understand this relationship in the common bean, our data show that autophagy genes play a role in symbiosis. Therefore, we reported 17 different families in the autophagy genes. We found that ATG8 and ATG18 are the most abundant families. Here, we contribute to identifying the different subfamilies in the ATG18 family using 280 proteins through 27 photosynthetic organisms. We performed evolution analysis to generate a phylogenetic tree that is supported by principal analysis components and conserved motifs. We found and described 3 subfamilies in the ATG18 family. Then we focused on the ATG18 family of common beans, exploring the microsynteny and the exploration of the primary and secondary structures. Based on these results, we chose PvATG18b that appeared in all the photosynthesis organisms selected in this study to generate a 3D model and evaluate the stability of the secondary structure by molecular dynamics simulation. Finally, these studies have led us to have accurate classification of 3 subfamilies in one of the most abundant families in autophagy genes. Also, we enhance the homologs of ATG18 in 27 photosynthetic organisms and show the structure of the PvATG18b protein and the possible interaction with membrane according to the observed by molecular docking studies.

The authors wish to acknowledge the DGAPA for partially funding this research (DGAPA/PAPIIT-UNAM grant. IN213221 to M.-K.A. and IN216321 to KN.) AND cONACYT cf-316528 to M.-K.A.

Directed evolution of laccase from Thermus thermophilus to increase activity and redox potential

Beatriz Miranda-Zaragoza[1], Paul Gaytán-Colín[2], Guillermo Huerta-Miranda[3], Margarita Miranda-Hernández[3], Leticia Ramírez-Silva[4], Claudia Rodríguez-Almazán[1*].

[1]Departamento de Micro y Nanotecnologías, Instituto de Ciencias Aplicadas y Tecnología, Universidad Nacional Autónoma de México, Av Universidad 3000, Col. UNAM, Ciudad de México, CP 04510, México. [2]Unidad de Síntesis y Secuenciación de ADN, Instituto de Biotecnología, Universidad Nacional Autónoma de México, Avenida Universidad 2001, Cuernavaca, Morelos 62210, México. [3]Instituto de Energías Renovables, Universidad Nacional Autónoma de México, Priv. Xochicalco, 62580 Temixco, Morelos, Mexico.[4]Departamento de Bioquímica, Facultad de Medicina, Universidad Nacional Autónoma de México, Av Universidad 3000, Col. UNAM, Ciudad de México, CP 04510, Mexico. claudiar@comunidad.unam.mx*

Laccases (EC 1.10.3.2: benzenediol oxygen oxidoreductases) belong to the multicopper oxidase family (MCO). Enzymes of this family catalyze the oxidation of different substrates with the simultaneous reduction of oxygen to water. The most used industrial laccases are those of fungal origin. However, bacteria laccases have more thermal stability, and more significant activity in extreme pH and temperature conditions than fungal laccase have a high impact on industrial processes. In the present work, the Thermus termophilus HB27 laccase (Tth-Lac) was engineering by directed evolution. Culture conditions (temperature, aeration, and copper inclusion) for the overexpression of Tth-laccase were determined. Copper and reduction of aeration produced an increase in the catalytic activity of Tth-Lac. The library obtained by random mutagenesis had a size of 1.2 x10^5 colonies. For the selection of mutants, decolorization of malachite green was assayed in the solid medium at pH 5 and pH 7 at 30 °C, 72 h.

The results show that the mutant did not cause changes in secondary structure. The mutant and wild type had the same behavior at different pH, while the specific activity at 90°C in the mutant demonstrated a significant increase. The kinetic constants were determined and did not show a significant difference between both enzymes. Tth-Lac and mutant's voltammetric response was measured at 25 °C and 61.5 °C. Only the mutant had an increase in current, this was observed at 61.5 °C, and the potential of copper T1 was displaced 100 mV. The mutant from*T. thermophilus* HB27 has excellent potential for use in the industry in extreme temperature conditions processes.

Founding: grant IT201522 (PAPIIT-UNAM).

Kinetic stability of the human γ-crystallins and the effect of metal ions involved in cataract disease.

Oscar Rodríguez-Meza[1], Giovanni Palomino-Vizcaino[2], Liliana Quintanar[2], Miguel Costas[1*]

[1]*Laboratorio de Biofisicoquímica, Departamento de Fisicoquímica, Facultad de Química, Universidad Nacional Autónoma de México, Mexico City 04510, México.*
[2]*Departamento de Química, Centro de Investigación y Estudios Avanzados (Cinvestav), Mexico City 07360, México.*
* *costasmi@unam.mx*

The lens is a highly specialized tissue responsible for transparency and focusing light on the retina. The correct performance is achieved due to a high concentration and suitable distribution of proteins: α-, β-, and γ-crystallins. They are into cells lacking organelles and null protein turnover, thus these crystallins have to remain folded for the entire life of the organism. The stability of the crystallins can be affected by several factors, such as the presence of metal ions, which results in misfolded proteins. These alterations can lead to protein aggregation into high-molecular-weight complexes and finally to the development of cataracts. Herein, we evaluated the effect of the Cu^{2+}, Zn^{2+}, and Hg^{2+} ions on the high kinetic stability of the most abundant human γ-crystallins: HγC, HγD, and HγS, by differential scanning calorimetry (DSC). It is shown that Cu^{2+} and Hg^{2+} ions affect the kinetic stability of the human γ-crystallins in different magnitudes, while Zn^{2+} ions did not. The interaction between each γ-crystallin with the metal ions was studied using isothermal titration calorimetry (ITC) experiments. Cu^{2+} and Hg^{2+} ions exhibited a clear exothermic interaction with the three γ-crystallins, but with higher energy for the mercury binding. However, the system HγD/Zn^{2+} showed an endothermic interaction, which coincides with the zinc-induced aggregation observed only in HγD. This work gives us a better understanding of the mechanisms of metal-induced aggregation of these proteins and their relevance in the development of cataract disease.

Exploring the stability landscape of *de novo* TIM barrels by protein design and engineering

Sergio Romero-Romero[1], Sina Kordes[1], D. Alejandro Fernández-Velasco[2], Birte Höcker[1]

[1] *Department of Biochemistry, University of Bayreuth, Germany.*
[2] *Faculty of Medicine, Universidad Nacional Autónoma de México, Mexico. Poster communication: sergio.romero-romero@uni-bayreuth.de*

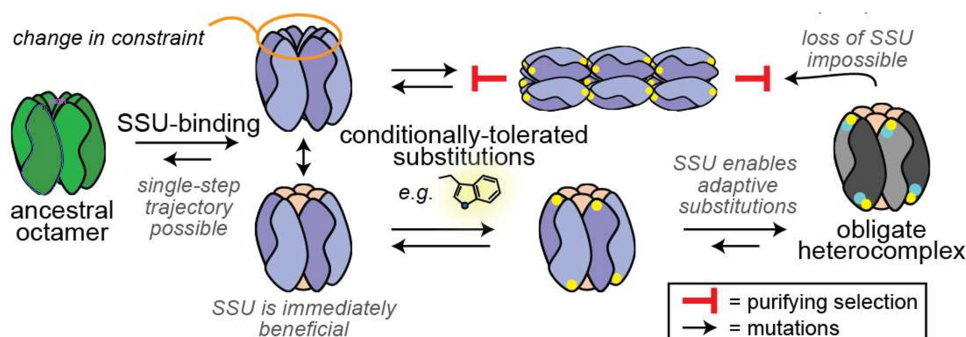
The ability to create and engineer stable proteins with custom-made functions is a fundamental goal in biochemistry, with practical relevance for our environment and society. Protein stability can be fine-tuned by modifying diverse structural features such as hydrogen-bond networks, salt bridges, hydrophobic cores, disulfide bonds, loop extension, or protein-protein interfaces, among others. Since the $(\beta/\alpha)_8$-barrel or TIM-barrel fold is one of the most abundant topologies in nature and a common functional scaffold, in this work, we designed and engineered a collection of stable de novo TIM barrels (DeNovoTIMs) using a computational fixed-backbone and a modular approach based on improved hydrophobic packing (Romero-Romero et al., 2021) and the introduction of salt-bridges (Kordes et al., 2022). DeNovoTIMs were subjected to thorough biochemical and folding analyses using computational, biophysical, structural, and thermodynamic methods to explore their stability and conformational landscape. DeNovoTIMs navigate a region of the stability landscape previously uncharted by natural proteins, with variations spanning 60 degrees in melting temperature and 22 kcal per mol in conformational stability throughout the designs. Significant non-additive or epistatic effects were observed when stabilizing mutations from different barrel regions were combined. Salt-bridge variants from some DeNovoTIMs exhibit important differences in comparison with the parental proteins, both in conformational stability and structural properties. The design of stable proteins increases the applicability of de novo proteins and provides crucial information on the molecular determinants of the sequence-structure-stability relationships, with this study being an essential step towards fine-tuned modulation of protein stability by protein design.

Romero-Romero, S., Costas, M., Silva Manzano, D. A., Kordes, S., Rojas-Ortega, E., Tapia, C., Guerra, Y., Shanmugaratnam, S., Rodríguez-Romero, A., Baker, D., Höcker, B., & Fernández- Velasco, D. A. (2021). The Stability Landscape of de novo TIM Barrels Explored by a Modular Design Approach. Journal of Molecular Biology, 433(18), 167153.

Kordes, S., Romero-Romero, S., Lutz, L., & Höcker, B. (2022). A newly introduced salt bridge cluster improves structural and biophysical properties of de novo TIM barrels. Protein Science, 31(2), 513– 527.

How a small protein evolved to enable the transition to aerobic life - Rubisco and its small subunit

Luca Schulz, Zhijun Guo, Jan Zarzycki, Wieland Steinchen, Jan M. Schuller, Thomas Heimerl, Simone Prinz, Oliver Mueller-Cajar, Tobias J. Erb, Georg K. A. Hochberg



Most autotrophic microorganisms assimilate CO_2 using ribulose-1,5-bisphosphate carboxylase/ oxygenase (Rubisco). Rubiscos are phylogenetically widespread and different forms of the enzyme exhibit varying catalytic parameters and complex stoichiometries. Aerobic autotrophs employ a form of Rubisco that assembles into a 16meric assembly of 8 large and 8 small subunits (SSU). Phylogenetic inferences show that this form evolved from ancestors that did not yet interact with a SSU. Due to the enrichment of SSU-bearing RubisCOs in aerobic habitats, the SSU is considered an adaptation to oxygen and is thought to help suppress the oxygenation side reaction of Rubisco. However, no biochemical evidence of the SSUs function exists. This is because the SSU is essential for both the solubility and activity of SSU-bearing Rubiscos, which hinders studies surrounding its influence on catalysis.

We identified novel Rubisco forms that are closely related to the evolutionary interval over which the SSU first evolved, use ancestral sequence reconstruction to resurrect ancestral Rubiscos that existed before and after the invention of the SSU and biochemically characterize these ancestral proteins.

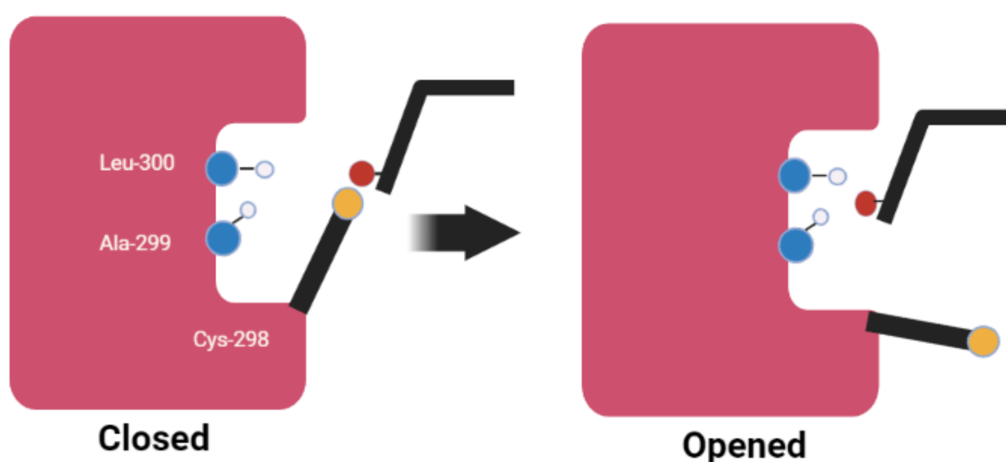
We show that modern day form I Rubiscoss evolved from an octamer of large subunits that started to interact with- and depend on the SSU more than 2.5 billion years ago. Immediately upon recruitment, the SSU improved Rubiscos catalysis and opened evolutionary paths that led to increased specificity. Thereafter, a single surface-exposed and functionally neutral substitution was able to cause Rubiscos dependence on the SSU for solubility. This substitution causes self-assembly of the octameric large subunit complex into fibers via an interface that is shielded and sterically inaccessible in the SSU-bound form. Our findings suggest that natural selection can fix protein-protein interactions for their beneficial effects and that such interactions can subsequently become essential via non-adaptive evolution. Additionally, we reveal the function of RubisCOs SSU.

Molecular dynamics simulation of the recognition of glutathione by aldose reductase

Gurprit Sekhon*‡, Balvinder Singh†, and Ranvir Singh*

Department cum National Centre for human genome studies and research, Panjab University, Chandigarh-160014, India. †CSIR Institute of Microbial Technology, Chandigarh-160039, India. Email: sekhongurprit@gmail.com

Aldose reductase (AR) is an NADPH-dependent oxidoreductase that has been extensively studied for its implications in secondary complications associated with diabetes. As a part of the cellular detoxification system, the physiological function of AR is reduction of toxic aldehydes and their glutathione conjugates to respective alcohols. We used molecular dynamics and simulation of a glutathione analog into the binding pocket of AR to elucidate the atomic details of the recognition of glutathione by AR. Hydrogen bond interaction between backbone nitrogen of Ala-299 and Leu300, and a specific oxygen atom of the glycine arm of glutathione forms the basis of this recognition mechanism. In this dynamic interaction, Cys-298 of AR functions as a gatekeeper to either restrict or permit the access of glutathione to the backbone nitrogen of Ala-299 and Leu-300. These atomic levels details are relevant for structure-based specific inhibition of AR.



1. Sekhon G, Singh B, Singh R. Role of Cys-298 in specific recognition of glutathione by aldose reductase. J Biomol Struct Dyn. 2021 Feb 25:1-9.

2. The Graphical figure was created with BioRender.com.

Uncovering the Cell-Penetrating Anticancer Activities of Peptide Prion-like Domains

Anjali Sharma[a], **Siddhant Sharma**[a,b*], Vineeta Kashyap[a]

[a] Department of Biochemistry, Deshbandhu College, University of Delhi, New Delhi, India 110019 [b] Blue Marble Space Institute of Science, Seattle, WA, 98154, USA
***Email:** siddhaantsharma.ss@gmail.com*

There has been intensive research into building bioinformatics and experimental workflows to identify prion-forming proteins in all domains of life, one of them being the prion-prediction tool Prion-Like Amino Acid Composition (PLAAC). PLAAC uses Hidden- Markov Models trained on experimentally verified prions and is capable of identifying the prion-like composition of amino acids in disordered fragments of proteins called prion domains (PrDs). Recent studies (Ding *et al.*, 2021) have pointed to the role of prion-like domains in cancer biology (Antony *et al.*, 2012), thus providing a potential therapeutic target. In this preliminary study, we aim to elucidate the cell-penetrating anticancer activities of the peptide PrDs identified through PLAAC. For the initial toy workflow, we scanned 1705 human mitochondrial proteins from MitoProteome Database through PLAAC, using the default settings leading to 17 putative PLAAC positive candidates having prion-like domains (PrD's). The amino acid sequences of these select PrD's were passed to AntiCP 2.0 and CpACpP servers for anti-cancer cell-penetrating peptide predictions based on amino acid composition, conserved features, and physicochemical properties. Additional benchmarking for cell-penetrating anticancer peptides (Cp-ACPs) identification was done through the CellPPD server. The putative anti-cancer, membrane penetrating peptides were checked for toxicity through ToxinPred 2.0 server to retain only the non-toxic peptides. The resultant list of peptides was passed through the SOLpro and Protein-Sol servers to calculate their solubilities. Finally, PSI-Pred and APPTEST servers will be used to predict the secondary and tertiary structures of the non-toxic, membrane penetrating anti-cancer peptide PrD's. This preliminary pipeline serves to validate if PrD's can show promising effects as anti-cancer peptide therapeutics, and this discovery workflow can be extended to multiple model organisms of therapeutic interest.

---

[1] Ding, M., Chen, Y., Lang, Y., & Cui, L. (2021). The Role of Cellular Prion Protein in Cancer Biology: A Potential Therapeutic Target. Frontiers in Oncology, 11
[2] Antony, H., Wiegmans, A.P., Wei, M.Q., Chernoff, Y.O., Khanna, K.K., & Munn, A.L. (2012). Potential roles for prions and protein-only inheritance in cancer. Cancer and Metastasis Reviews, 31(1-2), 1-19.

The DE-STRESS webserver and the expression of antibody designs

Michael J. Stam, Nadanai Laohakunakorn, Diego Oyarzún, Christopher W. Wood

University of Edinburgh

Natural proteins have a vast array of functionality, such as catalysts, materials, signaling molecules and more. They also have applications outside of their natural context as therapeutics, sensors and industrial feedstocks. Despite the huge number of natural proteins, nature has only sampled a small proportion of the possible proteins that could exist. *De novo* protein design aims to find new protein sequences from areas not sampled by nature, which can be used to solve challenges in medicine, agriculture, energy and beyond. Unfortunately, most proteins designed in this way fail when tested in the lab, usually due to low expression, misfolding, aggregation or lack of function. As a result, the design process is inefficient, unreliable and expensive, making it inaccessible to a lot of researchers.

However, it is possible that this high failure rate could be reduced by assessing the quality of protein designs before taking them into the lab. One potential way to approach this is to generate a set of high-quality metrics for protein designs, which could be used to make reproducible and data-driven decisions about which designs to characterise experimentally. Here, we present DE-STRESS (DEsigned STRucture Evaluation ServiceS), a web application for evaluating structural models of designed and engineered proteins [1]. DE-STRESS has been designed to be simple, intuitive to use, and responsive. It provides a wealth of information regarding designs, as well as tools to help contextualise the results and formally describe the properties that a design requires to be fit for purpose.

In addition to presenting the DE-STRESS webserver, we show how DE-STRESS can be used to assess the quality of antibody designs. We applied the DE-STRESS webserver to a data set of 193 antibody designs from the Fleishman Lab [2], to generate a set of high-quality metrics capturing different properties of the protein structures. The data set from the Fleishman Lab included a measure of expression which was used as the predictor variable. After this, a range of machine learning techniques were applied to this data set, in order to gain an understanding of which metrics were important for expression. Future work will involve using this understanding to design antibody sequences with high expression, with the aim of reducing the failure rate of the design process.

---

[1] Stam, M. J., & Wood, C. W. (2021). DE-STRESS: a user-friendly web application for the evaluation of protein designs. Protein Engineering, Design and Selection, 34. https://doi.org/10.1093/PROTEIN/GZAB029
[2] Baran, D., Pszolla, M. G., Lapidoth, G. D., Norn, C., Dym, O., Unger, T., Albeck, S., Tyka, M. D., & Fleishman, S. J. (2017). Principles for computational design of binding antibodies. Proceedings of the National Academy of Sciences of the United States of America, 114(41), 10900–10905. https://doi.org/10.1073/pnas.1707171114

Functionalization of de novo TIM-barrels with an ancient phosphate-binding motive

Cinthya R. Tapia-de-Aquino[1], D. Alejandro Fernandez-Velasco[1] and Birte Höcker[2]

[1]*Universidad Nacional Autónoma de México,* [2]*University of Bayreuth*

One of the most versatile and widely distributed protein folds in nature is the TIM- barrel (also known as $(\alpha/\beta)_8$-barrel). This fold is present in a wide variety of proteins that catalyze 5 of the 7 classes of chemical reactions in cells (Sterner and Höcker, 2005). Advances in developing *de novo* TIM-barrels as molecular scaffolds for specific functions were gained in the last few years. The first *de novo* TIM-barrel, sTIM11, (Huang *et al*. 2016) was further developed through hydrophobic packaging resulting in a collection of TIM-barrels (DeNovoTIMs) (Romero-Romero *et al*. 2021) as well as the introduction of a salt-bridge cluster (Kordes *et al*. 2021). These variants are valuable scaffolds for the generations of *de novo* enzymes and binders.

We consider the functionalization of DeNovoTIMs to bind phosphate as a relevant first target. Phosphate is ubiquitous in all molecules of life and, moreover, enzyme regulation and metabolic pathways are frequently mediated by phosphate transfer to other proteins. For these reasons, it was previously suggested that binding phosphate and nucleotide binding motifs emerged in early evolution (Hirsch *et al*. 2007). A number of ancient phosphate-binding units have been described by Longo *et al*. (2020). We explored the possibility to introduce different binding motifs into the existing DeNovoTIM scaffolds. Here we present our computational efforts in transplanting the motif from the HUP-domain into DeNovoTIMs as well as our initial experimental results.

Hirsch A.K.H., Fischer F.R., Diederich F., 2007. Phosphate Recognition in Structural Biology. Angew. Chem. Int. Ed. 46, 338–352.

Huang, P.-S., Feldmeier, K., Parmeggiani, F., Fernandez Velasco, D.A., Höcker, B., Baker, D., 2016. De novo design of a four-fold symmetric TIM-barrel protein with atomic-level accuracy. Nat. Chem. Biol. 12, 29–34.

Longo LM, Petrović D, Kamerlin SCL, Tawfik DS. Short and simple sequences favored the emergence of N-helix phospho-ligand binding sites in the first enzymes. Proc Natl Acad Sci. 2020 Mar 10;117(10):5310-5318.

Sterner R., Höcker B., 2005. Catalytic Versatility, Stability, and Evolution of the (βr)8-Barrel Enzyme Fold. Chem. Rev. 2005, 105, 11, 4038–4055.

Kordes S, Romero-Romero S, Lutz L, Höcker B. A newly introduced salt bridge cluster improves structural and biophysical properties of de novo TIM barrels. Protein Sci. 2022 Feb;31(2):513-527.

Romero-Romero S, Costas M, Silva Manzano DA, Kordes S, Rojas-Ortega E, Tapia C, Guerra Y, Shanmugaratnam S, Rodríguez-Romero A, Baker D, Höcker B, Fernández-Velasco DA. The Stability Landscape of de novo TIM Barrels Explored by a Modular Design Approach. J Mol Biol. 2021 Sep 3;433(18):167153.

Exploring the Effect of Water Models on Water Transport through Tunnels in Haloalkane Dehalogenase

Aravind Selvaram Thirunavukarasu[I,II] , Jan Brezovsky[I,II]

[I]*Laboratory of Biomolecular Interactions and Transport, Institute of Molecular Biology and Biotechnology, Faculty of Biology, Adam Mickiewicz University, Poznan, Poland*

[II]*International Institute of Molecular and Cell Biology in Warsaw, Warsaw, Poland*

Haloalkane dehalogenase is an environmentally important biocatalyst that catalyzes the hydrolysis of the carbon-halogen bond. These are hydrolytic enzymes, requiring the presence of water molecules in their active site to perform their action. The active site of these enzymes is deeply buried, meaning that the substrate and water molecules have to travel to this location from the bulk solvent for reaction to occur. The transport pathways used by the waters and substrates are called tunnels.

At present, molecular dynamics simulation is considered the best method to identify and characterize tunnels and capture the water transport across these tunnels. The preparation of the system and the water models used is a crucial part of how water transport through these transport pathways will be represented. First, we present our method to identify and evaluate water transport through these tunnels in massive simulation datasets. Next, we explore the effect of different popular water models on transport efficiency.

The presented research provides new insights into the water transport in haloalkane dehalogenase through different permanent and transient tunnels. The methodology could be adopted by other researchers to systematically identify transport pathways used by water or other ligands in different enzymes with buried active sites. Finally, by comparing the result of such analyses with different water models, we provide an overview of the models used in the description of the transport process obtained.

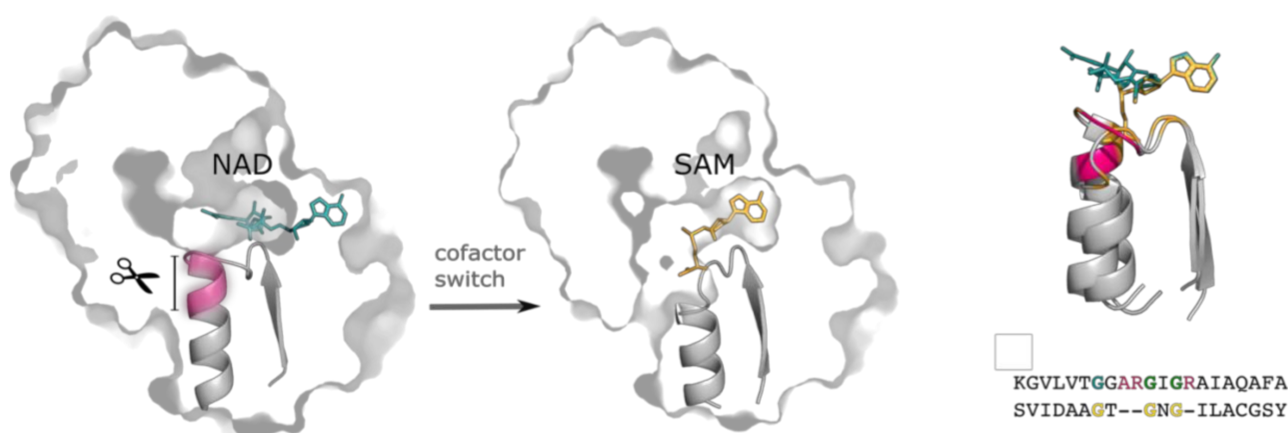
This research was supported by POWR.03.02.00-00-I006/17 project, National Science Centre, Poland (grant no. 2017/25/B/NZ1/01307), and grant of the Dean of faculty of biology, UAM (grant no. GDWB-05/2020). The computations were performed at the Poznan Supercomputing and Networking Center

An evolutionary path to engineer cofactor specificity from NAD to SAM employing InDels

Saacnicteh Toledo-Patino[1] & Paola Laurino[1]

[1] *Okinawa Institute of Science and Technology Graduate University*

Enzymes that employ cofactors to perform catalysis are ubiquitous across the tree of life and are necessary to assist about half of the enzymatic reactions existing in nature. Among cofactors, nucleotide derivatives are believed to be the catalytic fossils of a hypothetical RNA- based world. They may have become associated with early peptides, which evolved towards the enormous catalysis we observe today. Understanding how proteins developed to adopt different coenzymes, shades light into the emergence of the first proteins and their chemistry. This work studies the potential evolutionary steps that led to a change in cofactor specificity between Rossman proteins in virtue of their vast chemical variety. We observed that InDels may have modulated enzymatic functionality by allowing different coenzyme binding modes. As a proof of principle, we switched coenzyme specificity from NAD to SAM. These results draw an evolutionary bridge across reactions of different chemical nature (redox vs methylation) and provide an alternative way by which protein diversity is achieved.



A switch in cofactor specificity can be accomplished by InDel-based protein engineering between NAD- and SAM-binding Rossmann domains. An InDel (pink) in NAD-binding domains was detected through sequence-based profile-profile alignments, this InDel is missing in SAM-binding Rossmann domains and can be extracted from oxidoreductase scaffolds to accommodate the SAM cofactor. InDels may have mediated similar cofactor switches across other enzymes.

Computationally-guided activity enhancement of de novo designed granulopoietic proteins

Timo Ullrich[1], Mohammad ElGamacy[1,2], Julia Skokowa[2], *Andrei Lupas*[1]

[1]*Max-Planck-Institute of Biology, Department Protein Evolution, Tübingen*
[2]*Translational oncology, Internal Medicine II, University Hospital Tübingen*

De novo protein design has evolved into a powerful tool capable of providing a new generation of potential therapeutics. We have recently designed a novel protein (Boskar4) to act as a granulocyte-colony stimulating factor receptor (G-CSFR) agonist. Unlike the native receptor ligand (G-CSF), Boskar4 is small, highly stable, and can be produced with high yields in *Escherichia coli*. While Boskar4 showed nanomolar activity in cell-based assays and in differentiating neutrophils *ex vivo*, and in animal models, it was still less active than recombinant human G-CSF (rhG-CSF). This could be in part attributed to the lower affinity of Boskar4 to the G-CSFR compared to the native ligand. We therefore developed a pipeline, with Boskar4 as an initial candidate, that is able to quickly and efficiently affinity-maturate *de novo* designed proteins. In essence, a bacterial display is used in combination with computationally-designed libraries, utilizing a new design software Damietta. A set of 19 better binding monomeric Boskar4 variants were identified with up to 100-fold higher affinity. Nine of them were further investigated as tandem fusions in a cell-based assay, from which the most active variant nearly reached the activity of rhG-CSF itself. Additionally, high affinity variants of Boskar4 monomers can be used as competitive antagonists as well as the bases for new protein design applications like Novokines; novel-function cytokines.

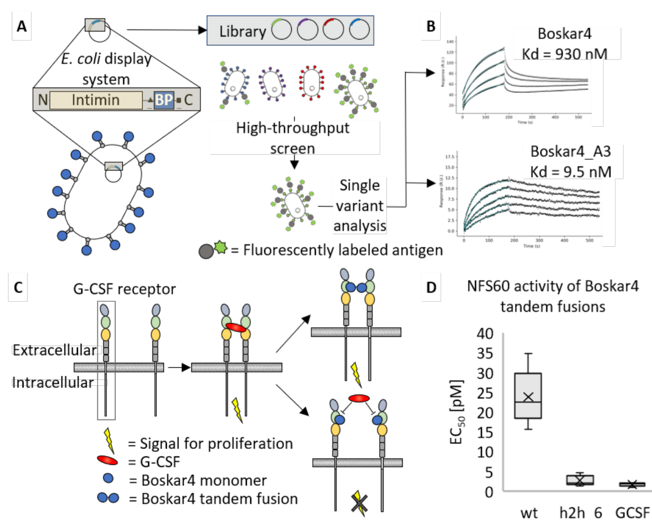


Figure 1: *A) Affinity maturation pipeline overview, B) surface plasmon resonance of enhanced Boskar4 variant, C) granulocyte-colony stimulating factor (G-CSF) receptor activation and respectively proposed mode of action of Boskar4 monomer or tandem fusion, D) EC50 of tandem fusions in NFS60 activity assays (n=5 independent experiments)*

Identification of Xanthomonas euvesicatoria proteins induced in plant tissue- based medium by peptide mass fingerprinting techniques

Valerio-Landa Sergio David*, Quiñones-Aguilar Evangelina Esmeralda, Hernández-Gutiérrez Rodolfo[1], Hernández-Montiel Luis Guillermo[2], Bach Horacio[3], Rincon-Enriquez Gabriel*

[1]Centro de Investigación y Asistencia en Tecnología y Diseño del Estado de Jalisco, A.C. Camino Arenero 1227, El Bajío del Arenal, Zapopan 45019, Jalisco, México. [2]Centro de Investigaciones Biológicas del Noroeste, Av. Instituto Politécnico Nacional 195, Col. Playa Palo de Santa Rita Sur, La Paz 23096, B.C.S., México. [3]Department of Medicine, Division of Infectious Diseases, University of British Columbia, 410-2660 Oak St., Vancouver, BC V6H3Z6, Canada.
** Correspondence (grincon@ciatej.mx / sergiodavidvale@gmail.com)*

Pathogenicity and hypersensitive response (hrp) genes encode type III secretion systems, effectors, and virulence toxins, and produce hypersensitive response reactions in non-host plants [1]. Hrp proteins from *Xanthomonas euvesicatoria* (Xae) can be induced *in vitro* in AWSL protein-Hrp induction medium containing plant host-related elements [2]. In this work we identify Xae proteins induced in AWSL medium and that use the tissue of a *Solanum lycopersicum* host plant as the only source of carbon and mineral nutrients. Proteins from Xae cultured in AWSL medium and NYGA rich medium (control condition) were separated by isoelectrofocus techniques in combination with 2D electrophoresis. Xae proteins up-regulated in AWSL medium were isolated and identified by peptide mass fingerprinting techniques for mass spectrometry. Our results showed 23 differentially abundant proteins in the interaction of Xae with elements related to its host plant in AWSL medium compared to control media. The biological relationship of overregulated proteins was associated with virulence mechanisms such as: secretion systems, hydrolytic enzymes, lipopolysaccharides and/or quorum sensing molecules. This is the first report identifying Xae proteins induced in AWSL medium based solely on plant tissue from a host plant. The results of this work are key in the identification of proteins related to pathogenesis and virulence.

[1] Wengelnik, K., Marie, C., Russel, M., & Bonas, U. (1996). Expression and localization of HrpA1, a protein of Xanthomonas campestris pv. vesicatoria essential for pathogenicity and induction of the hypersensitive reaction. Journal of Bacteriology, 178(4), 1061–1069.
[2] Valerio-Landa, S. D., Quiñones-Aguilar, E. E., Enriquez-Vara, J. N., Hernández-Gutiérrez, R., Hernández-Montiel, L. G., & Rincon-Enriquez, G. (2021). Method of induction, isolation, and selection of HRP proteins from phytopathogenic bacteria, a potential source of inducers of plant resistance. Mensaje Bioquímico, 45, 70–78.

Enhanced resistance of Mycobacterium tuberculosis β-lactamase against sulbactam

Ilona van Alen[1], Aleksandra Chikunova[1], Adil A Safeer[1], Misbha Ud Din Ahmad[2], Anastassis Perrakis[2], Marcellus Ubbink[1]

[1] Leiden Institute of Chemistry, Leiden University, Einsteinweg 55, 2333CC Leiden, The Netherlands.

[2] Division of Biochemistry, The Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands.

The current rise of antibiotic resistant forms of *Mycobacterium tuberculosis* is a global health threat that calls for new antibiotics. The pathogen *M. tuberculosis* produces a serine β-lactamase, BlaC, that can hydrolyse a wide spectrum of β-lactam antibiotics. This enzyme can be inhibited using β-lactamase inhibitors, though rapid evolution and the emergence of resistance against these inhibitors can be expected. We studied the evolutionary adaptability of BlaC against the inhibitor sulbactam. Laboratory evolution resulted in several amino acid substitutions in BlaC that conferred resistance against this inhibitor. The mutant proteins were characterized in vitro using a broad range of biochemical techniques to determine the effect of the mutation on protein structure, stability and kinetics. We showed that single nucleotide mutations in the gene can cause resistance against sulbactam by affecting the entrance to the active site cavity.

Biochemistry 60 (28), 2236-2245. doi: 10.1021/acs.biochem.1c00168

Bagel, a Novel Robust Eight-fold Symmetric Designer Protein Building Block

Laurens Vandebroek, Jeroen Vranken, Xiao Yin Lee, Arnout Voet

Laboratory for biomolecular modelling and design (KU Leuven), Celestijnenlaan 200G box 2403, 3000 Leuven, Belgium

Within nature, symmetry or pseudo-symmetry is ubiquitous. It can be found in fauna and flora, in organs, and even at the molecular level, such as the helical structure of DNA and protein oligomers. Although monomeric proteins generally do not express any form of symmetry at the tertiary level, it is at the quaternary level. The pseudo- symmetry exhibited at the tertiary level is usually observed within repeat protein families. Well-known examples include the trefoil fold for their three-fold symmetry, α/β-barrel with a four-/eight-fold symmetry, and the β-propellers which range from four-fold to ten-fold symmetry. The β-propellers consist of multiple families of propellers, greatly varying in sequence and symmetry, sometimes even within a family. Yet, what they all have in common is circularly arranged set of anti-parallel sheets, also called 'blades', which in turn consist of four strands.

We created a symmetric 6-bladed β-propeller by reverse engineering the natural evolution process, creating the Pizza protein, which was a useful building block from catalysis to assemblies and hybrid materials. However attempts to create an 8-bladed only partially succeeded. The first design, Tako, failed to self assemble and the subsequently designed Ika only partially self-assembled. The more recently designed Cake was highly interesting from a design perspective, but displays structural plasticity which makes it less interesting as a protein building block. Here we continued to engineer the Cake family of proteins to yield the Bagel protein, which is able to self-assemble in solution and doesn't suffer from the same structural plasticity as its predecessor. The protein permits truncation of the propeller into virtually any number of blades, which can be expressed and purified easily and in solution assembles into the complete Bagel β-propeller. In order to enhance the potential of Bagel as a building block, the protein is decorated with His-based metal binding sites. The basis of this design enables controlled complex formation at high pH in the presence of metal ions, while low pH allows for the binding of inorganic metal clusters. Here we present an overview of the features of Bagel, from a design perspective into the proof-of-concept assembly experiments showcasing the potency of Bagel as a novel protein building block.

Potential inhibitors of the catalytic site of OTUB1, to develop an anti-cancer drug using in-silico approaches.

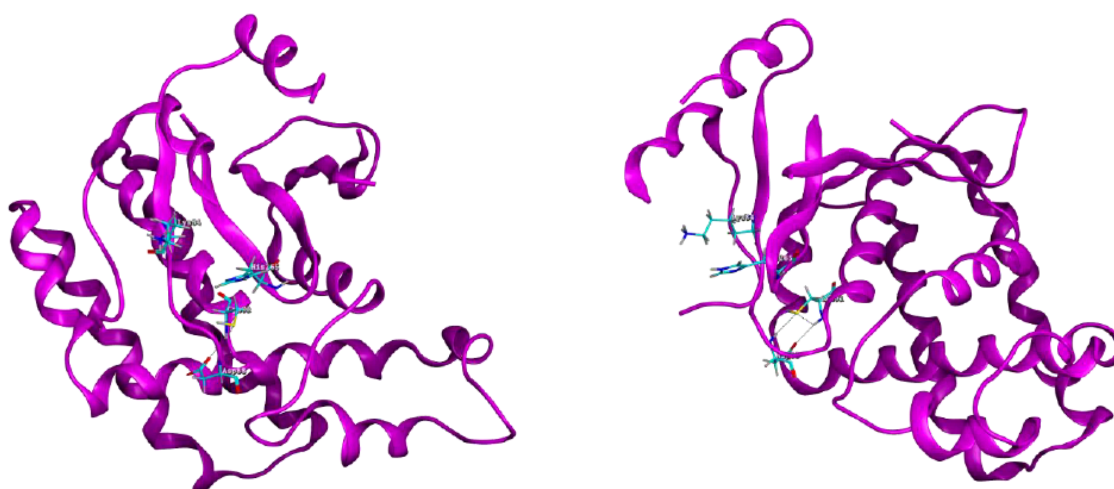
Jose Luis Vique-Sánchez[1] and Lizbeth Angelina García-Salazar[1]

[1]Facultad de Medicina, Universidad Autónoma de Baja California, Calle de la Claridad S/N, Col. Plutarco Elías Calles, C. P. 21376 Mexicali, B.C
E-mail: jvique@uabc.edu.mx

Cancer continues worldwide. It has been reported that OTUB1 plays a critical role in a variety of tumors and is strongly related to tumor proliferation, migration, and clinical prognosis by its functions on deubiquitinating. The drug advances continue against new therapeutic targets, in this study we use OTUB1 to develop a specific pharmacological treatment to regulate the deubiquitinating by OTUB1.

The aim of this research is to regulate functions of OTUB1, this study proposes ten compounds (OT1 - OT10) selected by molecular docking using a library of nearly 500,000 compounds, to develop a new anti-cancer drug to decrease the OTUB1 functions in the cancer processes. These compounds have a specific potential site of interaction in OTUB1, between Asp88, Cys91, and His265 amino acids, this site is necessary for deubiquitinating function of OTUB1.

It will be another way to attack the cancer, hindering the OTUB1 functions and this way to propose the develop of a new adjuvant treatment against cancer.



OTUB1 (Pink) shows the main amino acids in the potential binding site: Lys84, Asp88, Cys91, and His265 (Cyan).

SAKe: Computationally Designed Modular Protein Building Blocks for Macromolecular Assemblies

Staf M. L. Wouters[1], David E. Clarke[1,2], Hiroki Noguchi[1], Steven De Feyter[2], Arnout R. D. Voet[1]

[1]Laboratory of Biomolecular Modeling and Design, Department of Chemistry, KU Leuven, Celestijnenlaan 200G, 3001 Heverlee, Belgium
[2]Division of Molecular Imaging and Photonics, Department of Chemistry, KU Leuven, Celestijnenlaan 200F, 3001 Heverlee, Belgium
*staf.wouters@kuleuven.be

Advances in computational protein design have allowed for the development of new proteins with unique properties. Symmetric designer proteins have remarkable stability and can serve as versatile building blocks for the creation of macromolecular assemblies. Here we present the development and structural determination of SAKe: A new symmetric, stable protein building block with modifiable loops. Following the observation of pH induced 3D self- assembly, we engineered metal binding sites along the protein's internal rotational axis to fabricate 2D surface arrays. Using atomic force microscopy, we demonstrated Cu(II) dependent on-surface 2D self-assembly. Additionally, using dynamic light scattering and x-ray diffraction, we identified and characterized a SAKe mutant which shows in solution Zn(II) mediated nanocage formation. This work showcases a stable and highly modifiable SAKe protein scaffold, which holds promise as a building block for the creation of multi-functional macromolecular materials.

Modular gold binding antifouling protein

Chuanbao Zheng[1], Nicolò Alvisi[1], Robbert de Haas[1], Han Zuilhof[1], Renko de Vries[1]

[1]*Wageningen University*

Rendering surfaces antifouling is a crucial challenge whenever man-made solid materials need to be interfaced with biological fluids, living cells or tissues. As physical or chemical properties of metal surfaces can be tuned by via thin molecular layers adsorbed on- or chemically attached to them. Here, we designed protein that can adsorb on gold surfaces to tune antifouling properties of the gold surface. The structure of the protein can be modular as the length of the antifouling part can be tuned. By applying the surface binding antifouling protein, in principle, we can tune the antifouling properties of like silica, plastic or other surfaces. In principle, a wide range of biosensing application can be improved based on different surface binding proteins.

Towards the design and synthesis of artificial autonomous protein walkers

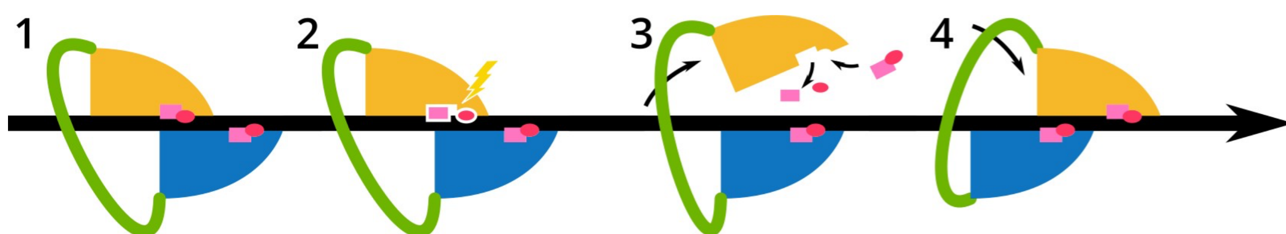
Anna Zink[1] , Olivier M. C. Laprévote[1] , Heiner Linke[2] , Paul M. G. Curmi[3] , Birte Höcker[1]

[1]Department of Biochemistry, University of Bayreuth, 94447 Bayreuth, Germany
[2]NanoLund and Solid State Physics, Lund University, Box 118, SE 221100 Lund, Sweden
[3]School of Physics, University of New South Wales, Sydney, NSW 2052, Australia

Sophisticated studies of biological protein motors – like kinesin, dynein and myosin – have led to much structural and biophysical information and the development of models for motor function. It remains however difficult to discern detailed mechanisms, for example about the relative role of different force generation processes, or how information is communicated across proteins to achieve the necessary coordination. A complementary approach to answering these questions is to build synthetic protein motors from the bottom up [1].

Here we will present a roadmap for designing and building autonomous, synthetic protein motors walking along a DNA track, based on existing, non-motor protein modules of known molecular function. The approach requires (i) the engineering of enzymatic activity controlled by the position of each foot toward each other, and whether they are bound to the track or not; and (ii) the engineering of asymmetry. We will discuss different ways of implementing these features and how enzymatic activity control can be envisioned as well as the associated challenges of these approaches.

Figure: Concept for an autonomous bipedal walker where allosteric communication is achieved by having the active site at an interface. Two asymmetric feet (blue and yellow) are tethered together where each foot has only one partial catalytic site. The full active site requires both feet to bind to adjacent sites in the DNA major groove in a forward direction. Step 1: 2nd foot (blue) binds adjacent to the 1st foot (yellow) completing the catalytic site, hence (2) generating products (pink and red). Step 3: after catalysis, the original foot (yellow) is released from the track and can (4) bind substrate and diffuse and bind to a forward binding site.



[1] Linke, H., Höcker, B., Furuta, K., Forde, N. R., & Curmi, P. M. G. (2020). Synthetic biology approaches to dissecting linear motor protein function: towards the design and synthesis of artificial autonomous protein walkers. Biophysical Reviews, 12(4), 1041–1054.