mechanisms of spatial adaptation during the formation of Anticalin-ligand complexes: (i) induced fit, in which conformational alteration follows ligand binding, and (ii) conformational selection, which is based on a pre-existing mixture of conformational states. Taken together, these molecular mechanisms demonstrate remarkable resemblance between the binding site of lipocalins (natural or engineered) and the well characterized complementarity-determining region of immunoglobulins (antibodies), which represent two structurally and functionally different types of mammalian plasma proteins.

1. Richter, A., Eggenstein, E. & Skerra, A. (2014) Anticalins: exploiting a non-Ig scaffold with hypervariable loops for the engineering of binding proteins. *FEBS Lett.* 588, 213-218.
2. Deuschle, F. C., Ilyukhina, E. & Skerra, A. (2021) Anticalin® proteins: from bench to bedside. *Expert Opin. Biol. Ther.* 21, 509-518.
3. Achatz, S., Jarasch, A. & Skerra, A. (2021) Structural plasticity in the loop region of engineered lipocalins with novel ligand specificities, so-called Anticalins. *J. Struct. Biol. X.* 6, 100054.
4. Jerschke, E., Eichinger, A. & Skerra, A. (2023) Drastic alterations in the loop structure around colchicine upon complex formation with an engineered lipocalin indicate a conformational selection mechanism. *Acta Crystallogr. F Struct. Biol. Commun.* 79, 231-239.

Thursday, November 16, S5 - Activity 5

To establish a Biostudies database of protein engineering results

L23

NOVEL STRATEGIES AND WEB-BASED TOOLS FOR PROTEIN ENGINEERING

Jiri Damborsky[1,2], David Bednar[1,2], Stanislav Mazurenko[1,2], Zbynek Prokop[1,2], Martin Marek[1,2], Simeon Borko[1,2], Jan Stourac[1,2], Milos Musil[1,2], Jan Mican[1,2], Sergio M. Marques[1,2], Joan Planas-Iglesias[1,2], Petr Kabourek[1,2], Petr Kouba[1,3], Anton Bushiev[3], Jiri Sedlar[3], Josef Sivic[3]

[1]Loschmidt Laboratories, Department of Experimental Biology and RECETOX, Faculty of Science, Masaryk University, Kamenice 5, 625 00 Brno, Czech Republic
[2]International Clinical Research Center, St. Anne's University Hospital Brno, Pekařská 53, 656 91 Brno, Czech Republic
[3]Czech Institute of Informatics, Robotics and Cybernetics, Czech Technical University in Prague, Jugoslávských partyzánů 1580/3, 160 00 Prague, Czech Republic
jiri@chemi.muni.cz

We develop novel strategies and web-based protein engineering tools under the ELIXIR Czech Republic umbrella. These are fully automated computational workflows which can operated using the intuitive graphical user interface [1]. Protein sequence or structure is typically the only input required for the calculation. The tools can be accessed freely via the Protein Engineering Portal (**Figure 1**). The tools are particularly suitable for experimentalists without prior structural biology or bioinformatics knowledge. The National Supercomputing Centre IT4Innovations provides the infrastructure for high-performance computing. This talk will introduce some of our web tools and illustrate their use for engineering proteins for biotechnological and biomedical applications [2].

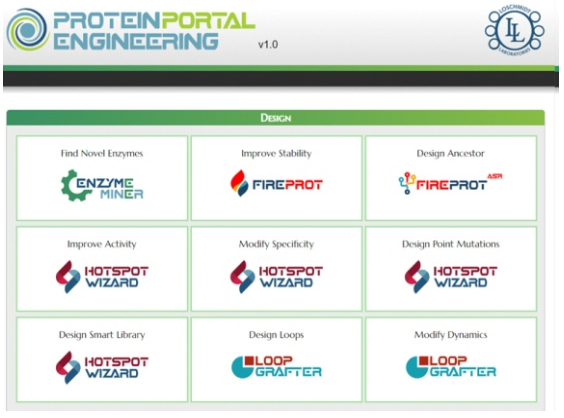

Figure 1. The graphical user interface of the Protein Engineering Portal provides unified access to the software tools and databases developed by the Loschmidt Laboratories and partners: https://loschmidt.chemi.muni.cz/portal/.

1. Marques, S. M., Planas-Iglesias, J., Damborsky, J., 2021: Web-based Tools for Computational Enzyme Design. *Current Opinion in Structural Biology* **69**: 19-34.
2. Vasina, M., Velecky, J., Planas-Iglesias, J., Marques, S. M., Skarupova, J., Damborsky, J., Bednar, D., Mazurenko, S., Prokop, Z., 2022: Tools for Computational Design and High-Throughput Screening of Therapeutic Enzymes. Advanced *Drug Delivery Reviews* **183**: 114143.

L24

NOVEL IMMUNOTHERAPEUTIC DRUGS THROUGH COMPUTATIONAL PROTEIN DESIGN

Clara Tabea Schoeder

Leipzig University, Faculty of Medicine, Germany
clara.schoeder@medizin.uni-leipzig.de

Recent advances in computational protein design have demonstrated its usability for tasks such as designing self-assembling nanoparticles, ß-barrel membrane proteins or the rapid discovery of picomolar binders for the SARS-CoV-2 receptor binding domain as antiviral drugs. In my laboratory, we are further leveraging these tools to design new immunotherapeutic drugs, including stabilized viral glycoproteins for rationally acting vaccines, antibodies and antibody fragments for cellular therapies, and adeno-associated virus capsids for targeted gene therapy. While computational protein design has been based on biophysical and knowledge-derived energy terms in the past, new machine learning methods are emerging with new capabilities. Here, we are presenting a study, where protein design in the Rosetta software framework was combined with the prediction of post-translational modifications using artificial neural networks. We integrated these models in the Rosetta framework, allowing the use of these predictions during design. With this, it is possible to enrich for intended post-translational modifications while altering the sequence, e.g. for the design of *N*-linked glycosylation, but also to decrease the occurrence of unintended modifications sites, such as deamidation of asparagine. This new method will be applied during epitope-focused immunogen design for influenza virus vaccines and for the stabilization of antibody therapeutics.

L25

USING MOLECULAR DYNAMICS (MD) CALCULATIONS FOR THE CHARACTERIZATION OF STRUCTURAL TRANSITIONS

O. Lampela[1], T-R. Kiema [1], T. Ohnuma[2,3], T. Fukamizo[3], R.K. Wierenga[4], A.H. Juffer[1,4]

[1]Biocenter Oulu, University of Oulu, Aapistie 5, 90220 Oulu, Finland
[2]Department of Advanced Bioscience, Kindai University, 3327-204 Nakamachi, Nara, 631-8505 Japan
[3]Agricultural Technology and Innovation Research Institute (ATIRI), Kindai University, 3327-204, Nakamachi, Nara, 631-8505, Japan
[4]Faculty of Biochemistry and Molecular Medicine, University of Oulu, Aapistie 7A, 90220 Oulu, Finland
Outi.lampela@oulu.fi

The increasing number of structures available for individual proteins in different conditions enable defining structural changes between protein states like unbound and liganded and active and non-active conformations. These structures do not reveal all the atomic movements (dynamics) needed for the transition between these states. Molecular dynamics (MD) simulations can be used to study and confirm anticipated structural changes. This requires an analysis of the available structures to define the states as well as the recognition of the path (transition) between these states from a simulated ensemble of conformations. The present work has applied correlation-based analysis tools to analyze MD simulation results to find possible states and transitions to understand the underlying mechanism of protein structure changes. We present examples of the information available from MD calculations, when combined with structural data. We also present some of the challenges to find the transition of interest in the simulation data.

For two proteins, vcCBP (*Vibrio cholerae* Carbohydrate Binding Protein [1] and MFE1(multifunctional enzyme, type-1) [2] several 500 ns all atom MD simulations were performed to study the dynamics of the proteins when starting from different structural states. vcCBP is a periplasmic solute binding protein specific for chitooligosaccharide, polymer of *N*-acetylglucosamine; $(GlcNAc)_n$. MFE1 is a monomeric enzyme with two active sites. In both studies structure-based predetermined distances between carefully selected c -atoms were analysed as function of the simulation time as well as by cross-correlation of these distances with each other. The atom distance analysis and dynamical cross-correlation (DCC) data from MD simulations have been used to investigate the intra-domain and inter-domain movements.

In the case of vcCBP, the distance analysis can differentiate between a bound and an unbound state, but it is in the DCC data of the c -atom positions which highlights key features of the binding mode of the two different ligands, $(GlcNAc)_2$ and $(GlcNAc)_3$, as shown in Fig. 1. The cross-correlation analysis reveals different behaviour of the domains and division of a single domain into two domains, which is consistent with the the thermodynamic

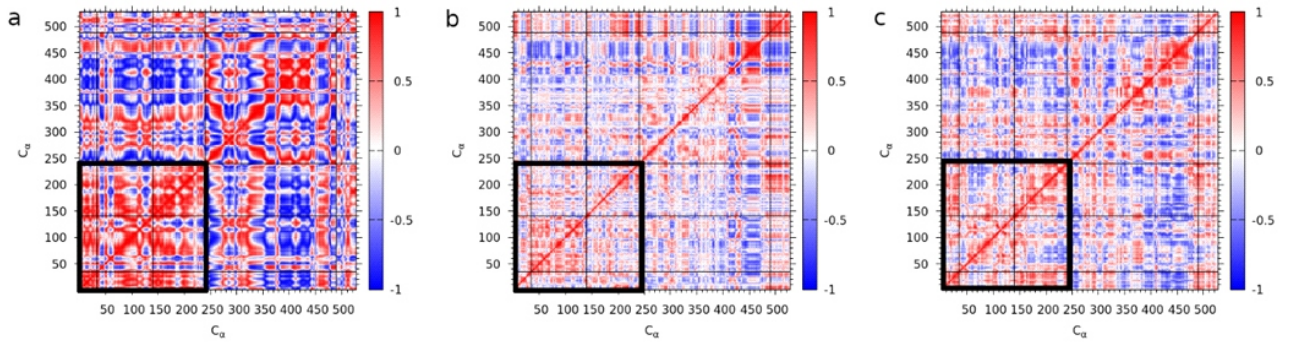

Figure 1. Dynamical cross-correlation (DCC) plots [4] from MD-simulations of unliganded (a) and $(GlcNAc)_2$ –bound (b) and $(GlcNAc)_3$ –bound vcCBP (c). The strong red and blue colours indicate positive correlation and negative correlation of atomic movements, respectively. The region comprising a single domain in unliganded vcCBP, having different dynamical properties in $(GlcNAc)_2$ –bound (b) and $(GlcNAc)_3$ –bound vcCBP (c) is highlighted by a black line.

data measured for these ligands. The results predict the existence of dynamical domains which is confirmed by a combination of experimental structural data [3].

For the MFE1 study two different conformational states are found in the asymmetric unit of the crystallised protein. The characteristic distances for the A and B states have been defined in previous publications [2]. The analysis of the 500ns simulations runs of the unliganded structure, using these characteristic distances, shows that these structural states do not converge to each other during this time scale attainable by the MD simulation. The simulations are extended and analysis of 500 ns simulations runs of liganded MFE1 complexes is still in progress.

The collection of the distance characteristics from two or more protein conformational states using extensive simulation data will make it possible to automate the search of the residues that are important for the function of these proteins.

1. W. Suginta, N. Sritho, A. Ranok, D.M. Bulmer, Y. Kitaoku, B. van den Berg, T. Fukamizo, *J Biol Chem.*, **293**, (2018), 5150.
2. S.Sridhar, W. Schmitz, J.K. Hiltunen, R. Venkatesan, U. Bergmann, T-R. Kiema, R.K. Wierenga, *Acta Crystallographica*, **D76**, (2020), 1256.
3. T. Ohnuma, J. Tsujii, C. Kataoka, T. Yoshimoto, D. Takeshita, O. Lampela, A. H. Juffer, W. Suginta, T. Fukamizo, *Manuscrip*t, (2023).
4. G. Battocchio, R. González, A.G. Rao, I. Schapiro, M. A. Mroginski, *The Journal of Physical Chemistry B*, **124**, (2020) 1740.

L26

FROM ALPHAFOLD TO PYMOL: ENABLING SEAMLESS ACCESS TO STRUCTURAL BIOINFORMATICS TOOLS

S. Rosignoli and A. Paiardini

Sapienza University of Rome, Piazzale Aldo Moro, 00185, Rome
serena.rosignoli@uniroma1.it

AlphaFold [1], the revolutionary artificial intelligence system developed by Google DeepMind, has ushered structural biology into a new era with its remarkable precision in predicting protein structures. However, like any pioneering technology, AlphaFold has encountered its fair share of challenges that have driven further innovation. These limitations are entirely justifiable and predictable, given the inherently statistical nature of the solution proposed for the ab-initio prediction problem. Nevertheless, the models generated by AlphaFold can be misleading if not rigorously scrutinised. To address this, numerous research groups have delved deep into both the algorithm and its predictions, seeking to uncover the various potential applications. Consequently, a multitude of tools [2-6] has emerged, aiming to enhance AlphaFold's predictive capabilities.

In light of these constraints, our objective is to seamlessly integrate AlphaFold into the PyMOL-PyMod environment [7,8], a well-established graphical interface that supports structural bioinformatics analyses. The development of a database indexed with UniProt codes, complete with APIs, has significantly broadened AlphaFold's accessibility [9], making predictions available even in the absence of cutting-edge computational resources. This integration will empower us to adapt database searches, supported by PyMod, to AlphaFold models, providing direct access to both sequences and structures for further manipulation within the PyMOL-PyMod environment. Some of the analyses supported by PyMod are sequence similarity searches, building multiple sequence/structure alignments, constructing phylogenetic trees, conducting evolutionary

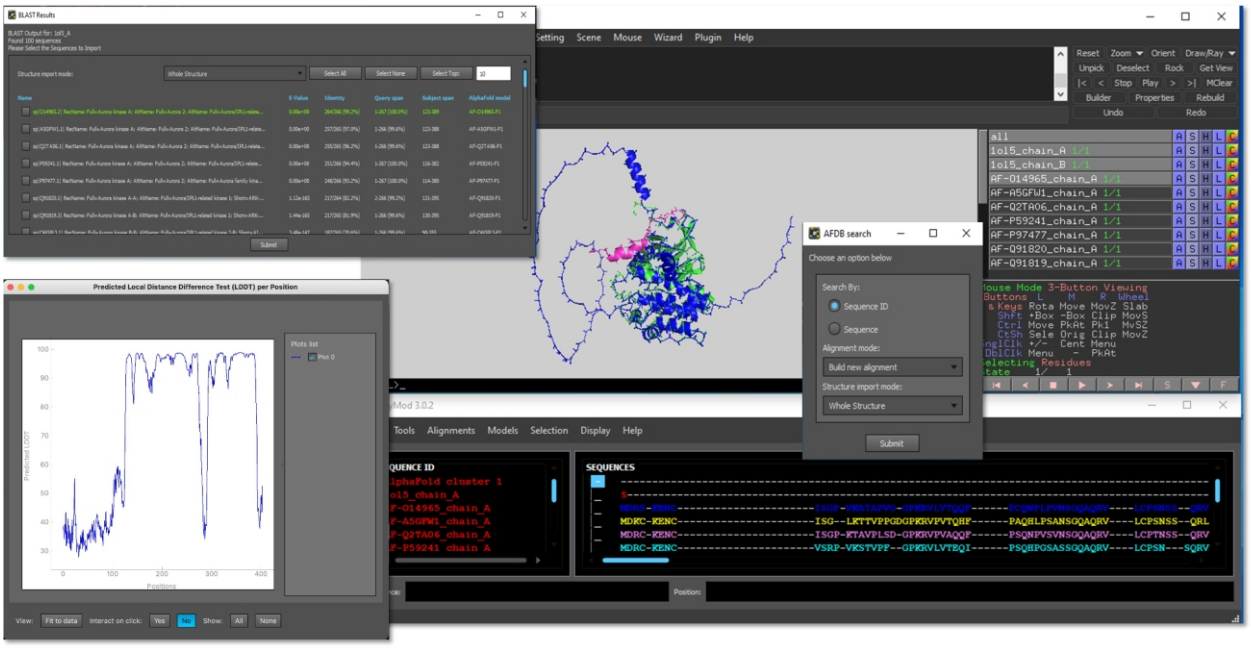

Figure 1. An Overview of PyMod's incorporation of AlphaFold-inspired features.

conservation analyses, parsing domains, and modeling single/multiplechains and loops.

As previously noted, AlphaFold is constrained to proposing protein models devoid of any cofactors or ligands. This lack is further exacerbated by the fact that AlphaFold algorithm only slightly explores the entire conformational space of the protein and completely excludes the potential influence of interactors in the folding mechanism. In this scenario, structural superimpositions and sequence searches facilitated by PyMod, will elegantly enable the identification of plausible ligands for these models whilst tailored PDB-API queries will facilitate the identification of alternative protein conformations. Importantly, these automatically gathered data will not only offer a clear visualization within PyMOL and PyMod but will also be subject to a comprehensive analysis of their binding probabilities and the potential influence of these ligands on the model's folding.

Despite the substantial advancement represented by *ab-initio* models in structural predictions, homology-based methods remain of enduring interest. However, their approach varies, as they can be effectively employed in conjunction with AlphaFold predictions. This cooperative strategy shows potential in empowering structural biologists to create elaborate models of protein complexes, especially when AlphaFold's predictive abilities may have limitations in determining the precise arrangement of domains or multi-subunit proteins.

Our proposal represents a significant step forward in streamlining research workflows and democratising access to cutting-edge structural biology tools. This initiative embraces the synergy between AI-driven predictions and user-friendly open-source platforms, creating a unified and easily accessible environment. Pymod has a well-established history in the context of education, thus bringing these educational benefits to AlphaFold, which in turn enhances its value for instructional purposes as well.

1. J. Jumper et al., *Nature*. **596** (2021), 583-589.
2. J. M. Thornton et al., *Protein Science*. **31** (2022), 283-289.
3. Lomize, Andrei L et al., *Protein Science*. **31** (2022), e4318.
4. M.L. Hekkelman et al., *Nature*. **20** (2023), 205–213.
5. H. Bagdonas et al., *Nature* **28**, (2021), 869–870.
6. P. Bryant et al., *Nature* **13**, (2022), 6028.
7. Schrödinger, L., & DeLano, W. (2020). PyMOL. Retrieved from http://www.pymol.org/pymol.
8. G. Janson et al., *Bioinformatics*. **37**, (2021), 1471-1472.
9. M. Varadi et al., *Nucleic acids research*. **50** (2022), D439–D44.

L27

DESIGN OF NOVEL PEPTIDES TARGETED TO HUMAN PRIMARY AMINE OXIDASE

Marion, Denise Alix, Gabriela Guédez, Tiina, Annamaria Salminen

Åbo Akademi university, Finland
marion.alix@abo.fi

Human primary amine oxidase (hAOC3), a transmembrane protein on the endothelium binds the Sialic Acid-binding Immunoglobulin-like lectin-9 (Siglec-9) on the leukocytes surface for the extravasation to the inflammatory site [1] During inflammatory conditions, hAOC3 is translocated to the endothelial cell surface and, therefore, a labelled Siglec-9 peptide targeting specifically [1]. hAOC3 can be used to detect tumors and acute or chronic inflammatory response in many diseases by positron emission tomography (PET) e.g. rheumatoid arthritis [2]. The aim of this study is to design improved Siglec-9 peptide derived from peptide for PET imaging and further develop then for anti-cancer and anti-inflammatory agents.

We have started our study by designing a novel Siglec-9 peptide that corresponds to the original phage peptide but is significantly smaller than the Siglec-9 peptide used for PET-imaging [1]. The experimental binding studies show that the designed peptide interacts with hAOC3 like the original phage peptide (positive control) whereas the mouse Siglec-E peptide (negative control) did not show any binding. In addition to these peptides, we have predicted the binding of even smaller Siglec-9 peptides containing WRG motif like the phage peptide, Siglec-10 peptide: containing QRG motif, and other related peptides by docking experiments. The docking results help us to understand the specific binding of human Siglec-9 peptide to hAOC3. Interestingly, the binding sites for the phage peptide and the human Siglec-9 peptides coincide with the known inhibitor binding site in hAOC3 around specificity site (Fig. 1).

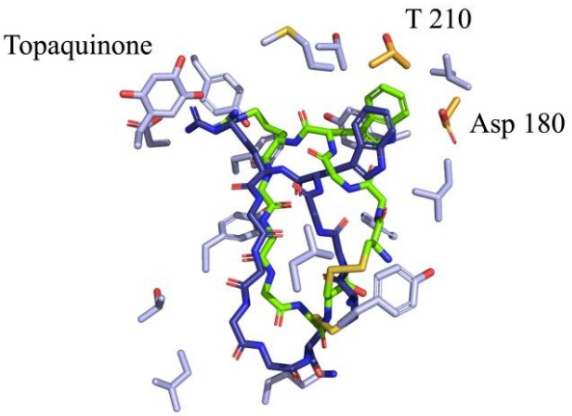

Figure 1. Phage (green) and Human (dark blue) Siglec-9 peptides docking into hAOC3 protein. Active site side chain in sticks lightblue and specificity site in sticks orange

1. K. Aalto, A. Autio, E. A. Kiss, K. Elima, Y. Nymalm, T.Z. Veres, F. Marttila-Ichihara, H. Elovaara, T. Saanijoki, P. R. Crocker, M. Maksimow, E. Bligt, T.A. Salminen, M. Salmi, A. Rovainen and S. Jalkanen, *Blood*, **118**, (2011), pp. 3725-33.
2. H. Virtanen, A. Autio, R. Siitinen, H. Liljenba¨ck, T. Saanijoki, P. Lankinen, J. Ma¨kila¨, M. Ka¨kela¨, J. Teuho, N. Savisto, K. Jaakkola, S. Jalkanen, and A. Roivainen, *Arthritis Research & Therapy*, **17**, (2015), 308.

L28

ASSESSING THE PERFORMANCE OF PROTEIN REGRESSION MODELS

Wouter Boomsma

University of Copenhagen, Denmark
wb@di.ku.dk

Machine Learning is increasingly used to predict protein fitness as a guide for selecting promising candidates in protein engineering. It is therefore natural to ask how good these predictions are, and how much we are able to learn from one round of experiments to the next. In this talk, I will discuss how different assessment criteria for regressor performance can lead to quite different conclusions, depending on the choice of metric, and the underlying definition of generalization. I will also highlight issues of sample bias in typical regression scenarios and how this can lead to misleading conclusions about regressor performance.

L29

PREDICTION OF BACTERIAL INTERACTOMES BASED ON GENOME-WIDE COEVOLUTIONARY NETWORKS: AN UPDATED IMPLEMENTATION OF THE CONTEXTMIRROR APPROACH

Fernandez-Martin M., Pontes C., Ruiz-Serra V., Valencia A.

Barcelona Supercomputing Center (Life Sciences), Plaça d'Eusebi Güell, 1-3, 08034 Barcelona
miguel.fernandez@bsc.es

The biological function of proteins is preserved through coevolution and can be quantified by computing the similarity between the phylogenetic trees of pairs of protein families [1]. When the phylogenetic similarity is high, it indicates that proteins are likely to interact. However, this similarity is influenced by many factors, including background evolution. Current coevolution-based methods treat protein pairs independently, despite proteins interacting with multiple others.

The ContextMirror methodology evaluates coevolution by integrating the influence of every interactor on a given protein pair (coevolutionary network), providing more accurate protein-protein interaction predictions [2]. In our study, we evaluate the ContextMirror pipeline, already shown to improve the prediction of protein-protein interactions, by predicting protein-protein interactions for the full proteome of *Escherichia coli* (4298 proteins). Preliminary predictions reveal the potential of this approach to improve our understanding of protein coevolution. The true positive rate of the top-500 predictions (60% accuracy) is approximate to other methods and compared to the STRING database [3], they map only to high-confident pairs (confident score > 0.8). The physical compatibilty between these pairs was confirmed by quantifying the structural PPI interface as the pDockQ score from structures modelled with AlphaFold-Multimer [4] (Figure 1).

In the current stage of our analysis, ContextMirror is being used to predict PPI networks for different bacterial proteomes and to identify differences in their predicted interactomes with potential applications in drug design and protein engineering.

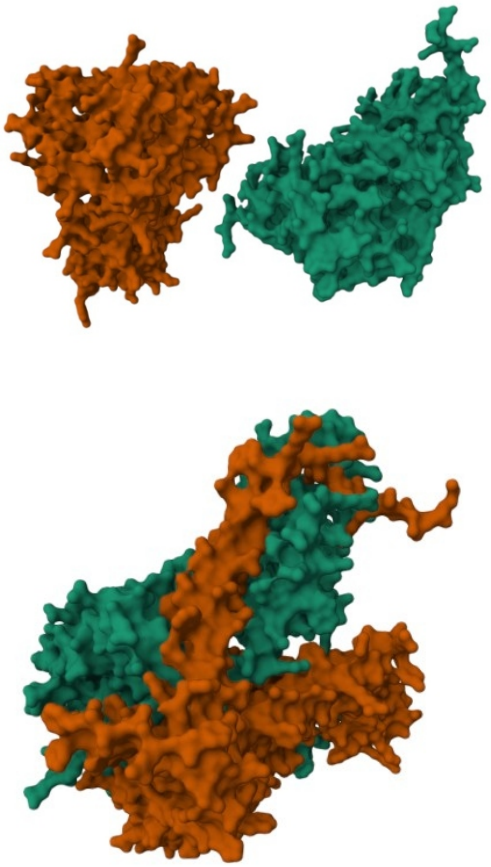

Figure 1. Structural compatibility validation for one low-scoring pair (left) and one high scoring pair (right). No suitable inferface was observed (pDockQ = 0.03) for one of the least confident predictions (RsmA-SrlR – confidence score = -1.3), while a valid interaction (pDockQ = 0.74) was found for the most confident prediction (UgpA-UgpE – confidence score = 1.93).

1. Pazos, F., & Valencia, A. (2001). Similarity of phylogenetic trees as indicator of protein–protein interaction. Protein Engineering, Design And Selection, 14(9), 609-614.
2. Juan, D., Pazos, F., & Valencia, A. (2008). High-confidence prediction of global interactomes based on genomewide.coevolutionary networks. Proceedings Of The National Academy Of Sciences, 105(3), 934-939.
3. Szklarczyk, D. et al. (2023). The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. Nucleic Acids Res, 51(D1), D638-D646.
4. Wensi Zhu and others, Evaluation of AlphaFold-Multimer prediction on multi-chain protein complexes, Bioinformatics, Volume 39, Issue 7, July 2023, btad424.