FINAL SCIENTIFIC REPORT – K 127961 – Tamas Hegedus, PI
Membrane and protein interactions of disordered regions in transmembrane proteins

Introduction

Interactions of intrinsically disordered regions (IDRs) in transmembrane (TM) proteins are important in many regulatory processes and can be formed not only with a protein partner but also with a membrane bilayer. We named IDRs involved in such membrane lipid-induced disorder-to-order transition as MemMoRFs, in an analogy to IDRs exhibiting disorder-to-order transition upon interaction with protein partners termed Molecular Recognition Features (MoRFs). Although these IDR-membrane interactions are critical in numerous physiological and pathological processes, in contrast to protein-protein interactions involving IDRs, IDR-membrane interactions are not studied extensively.

Our aims:

1) To collect disordered regions from transmembrane and membrane associated proteins;
2) To characterize MemMoRFs' interactions with both proteins and lipids using *in silico* methods;
3) To investigate the conformational ensemble of the CFTR R domain using experimental methods.

Specific aim #1 The comprehensive analysis of unstructured regions in transmembrane proteins.

Currently, both the experimental detection and computational characterization of MemMoRFs are challenging, and information about these regions are scattered in the literature. To facilitate related studies, we generated a comprehensive database of experimentally validated MemMoRFs based on manual curation of literature and structural NMR data. To characterize the dynamics of MemMoRFs, secondary structure propensity and flexibility calculated from NMR chemical shifts were incorporated into the database. These data were supplemented by inclusion of sentences from papers, functional data, and disease-related information. We had published the collection of a dataset including transmembrane and peripheral proteins, which possess MemMoRFs (https://memmorf.hegelab.org; Csizmadia *et al.* Nucleic Acid Research 2021), which serves a gold standard set to develop MemMoRF predictors.

In collaboration with Lukasz Kurgan (Virginia Commonwealth University, USA), a predictor for MemMoRFs were created as the combination of state-of-the-art tools. The selected tools predict residues that possess key characteristics of MemMoRFs, such as intrinsic disorder, disorder-to-order transition and lipid-binding. We identify and combine results from three tools that include flDPnn for the disorder prediction, DisoLipPred for the prediction of disordered lipid-binding regions, and MoRFCHiBiLight for the prediction of disorder-to-order transitioning protein binding regions. Our empirical analysis demonstrates that combining results produced by these three methods generates accurate predictions of MemMoRFs. We also show that use of a smoothing operator produces predictions that closely mimic the number and sizes of the native MemMoRF regions. Importantly, the resulting CoMemMoRFPred

method is available as an easy-to-use webserver at http://biomine.cs.vcu.edu/servers/CoMemMoRFPred.

We are also developing in our laboratory a MemMoRF predictor based on protein language models (pLM), which became recently available for development of predictors based on protein sequence. In the first step, pLM embeddings are generated for our gold standard set and are used for machine learning of MemMoRF properties. We tested linear regression models, fully connected neural network (NN) and convolutional NNs. The best performing model was based on fully connected NNs and its performance is much higher than CoMemMoRFPred.

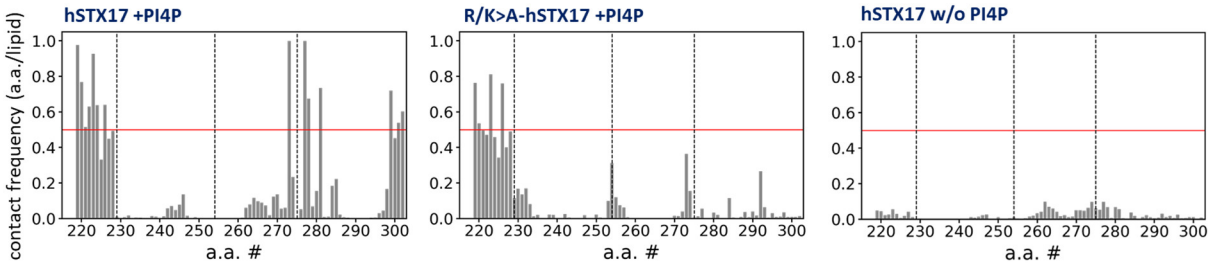
These tools will aid future studies of MemMoRFs in the context of exploring their abundance, cellular functions, and roles in pathologic phenomena.

The idea of MemMoRFs originated from the observation of a segment, which we observed in the cryo-EM map but was not resolved in the all-atom structure, in the regulatory insertion of ABCG2. Therefore, we also studied how the trafficking and function of this protein may be affected by its disordered parts (Mózner *et al.* Cells 2019, Nagy *et al.* CMLS 2020). During these studies we realized that the experimental data on the boundaries of membrane-embedded regions, which are generally sparse, is present in many cryo-EM maps of transmembrane proteins. Therefore, we also developed a pipeline to extract this information on membrane protein embedment in the bilayer (http://memblob.hegelab.org; Farkas *et al.* Bioinformatics 2019).

Specific aim #2 *In silico* investigation of IDR structural ensembles in solution and in complex with a bilayer or a protein.

We aimed to characterize the interaction of MemMoRFs with both lipids and proteins using molecular dynamics (MD) simulations. The progress of these studies was slow because of the difficulty of force field selection and unexpectedly long runtime required for disordered/elongated polypeptides. We tested a unique force field combination proposed my collaborators (Helmut Grubmüller, Max Planck Institute, Göttingen). However, we finally decided to used CHARMM36m, which was partially developed in his laboratory. This force field seems to describe lipid-protein interactions well, demonstrated via our new collaborative project with Gábor Juhász (ELTE, Budapest).

His group experimentally investigated the membrane interaction of an autophagosomal SNARE protein Syntaxin 17 (STX17). Its membrane binding occurs via the C-terminal disordered region and is dependent on negatively charged PI4P lipid molecules. We performed MD simulations with the wild type STX17 and a mutant, in which potential membrane-interacting positively charged residues were replaced with alanine. The simulations were performed in the presence and absence of PI4P lipids. Contact frequencies of residues with the lipid bilayer a plotted in the figure below. Our results highlight the interacting residues and their most probable conformations. Similar results were obtained with the sequentially different disordered region of *D. melanogaster* STX17.



For several studies, including those with ABCG proteins and STX17, we also used AlphaFold structural models. As a first step, we investigated the behavior of AlphaFold models in MD simulations. We found that the AF-based ABCG36 model performed better than our conventional homology model (Hegedus *et al.* CMLS 2022). In the same study, we demonstrated that AlphaFold works nicely also for membrane proteins. However, our results also indicated that it has somewhat low performance for regulatory complexes and complexes involving mobile segments (Berta *et al.* https://www.biorxiv.org/content/10.1101/2023.06.13.544745v1). In this latter study, we propose that SARS-Cov-2 Envelope protein exhibit similar structural properties, including a MemMoRF, to SERCA regulating regulins, such as phospholamban.

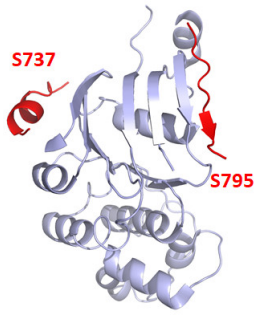
Making our structures and conformational ensembles of disordered proteins available for the public is challenging. We joined the efforts of various structural databases (e.g. Protein Data Bank in Europe, Protein Ensemble Database) in collaboration with the groups of Sameer Velankar (EBI, UK) and Christine Orengo (UCL, UK) to implement a system for distributing structural data on the web. We contributed to the development of the 3D beacon client (https://www.ebi.ac.uk/pdbe/pdbe-kb/3dbeacons) that sub-system aims to ease the inclusion of structures generated in small laboratories into large databases (Varadi *et al.* Gigascience 2023 https://pubmed.ncbi.nlm.nih.gov/36448847). We also lunched our site (https://3dbeacon.hegelab.org) that was aimed to be a prominent example of connecting laboratories into the 3D beacon system. As a test for our 3D beacon client we used important structures, generated ABC protein dimers, which are present neither in the PDB nor in the AlphaFold Structural Database (Tordai *et al.* IJMS 2022).

Specific aim #3 Experimental exploration of MemMoRFs in the CFTR R domain.

We found the experimental description of memMoRF binding to lipid and protein partners important. For this objective, we utilized CFTR's disorder R domain, since its conformation has been studied in the absence and presence of phosphorylation and its interaction with the soluble intramolecular binding partner (NBD1) has been characterized. First, we overlaid purified R domain on PVDF membrane strips with immobilized, negatively charged lipids, either in the presence or absence of phosphorylation. However, we could not detect specific interactions. Two scenarios are possible: (1) Since IDR interactions are weak, overlaying may not be an adequate method to study this type of interactions. (2) R domain does not bind to lipids (has no MemMoRF regions). In order to further explore R domain binding to lipids we are setting up a fluorescent nanodisc system (Ren Q et al. Commun Biol 2022, 5 (1), 1-7, https://doi.org/10.1038/s42003-022-03443-4.). The specialty of this system is a split GFP at the

ends of the amphipatic protein MSP1D1, thus the fluorescence will depend on the remodeling of the membrane e.g. by immersing or interacting proteins, such as MemMoRF. Since it is known in our field that I engaged with CFTR disordered R domain, I was invited to contribute to a review paper (Geisler and Hegedus, FEBS Letters 2020).

A major objective of ours was also to generate experiment-based structural ensembles for IDRs and MemMoRFs. We explored and performed some experiments with CFTR R domain using CD and SAXS without remarkable results. Finally, we started collaboration with Gitta Schlosser (Eotvos University, Budapest), who is a mass spectrometry expert. In order to determine regions in disordered proteins, which are close to each other in space, we purify R domain and NBD1-R domain constructs (the first Nucleotide Binding Domain is an ordered domain of CFTR and is known to provide intramolecular interaction sites to the disordered R domain). We performed cross-linking with bifunctional cross-linkers, and detect the location of the cross-links using MS (XL-MS). We also generated complexes of NBD1 and R domain peptides using AlphaFold-Multimer to understand their interaction at a higher resolution (a few examples are shown in the figure on the right). Our objective is to integrate CL-MS experiments, structures of NBD1/R peptide complexes, and MD simulations with full length CFTR to understand the R domain regulation in the physiological context. As a required step before this, we studied the allosteric communication between different CFTR region (Soya *et al.* Nat. Comm. 2023, revised version submitted).



Complex of R domain peptides (red) and NBD1 (blue) were generated by AlphaFold-Multimer. Two examples, including phosphorylation sites 737 and 795, are shown.

We also plan to study NBD1-R domain interactions using single molecule force spectroscopy methods. We completed the first stage of this study and characterized the unfolding of NBD1 using atomic force spectroscopy (Padanyi *et al.* Computational and Structural Biotechnology Journal, 2022).

Dissemination of our newest results:

1. A manuscript on CoMemMoRFPred was submitted for publication
   Basu *et al.* CoMemMoRFPred: sequence-based prediction of MemMoRFs by combining predictors of intrinsic disorder, MoRFs and disordered lipid-binding regions
2. A manuscript is under preparation on pLMMoRFPred
   Csepi *et al.* Application and analysis of protein language models for predicting membrane interacting regions
3. A coauthored manuscript on the membrane interaction of the autophagy STX17 protein is under preparation (main author: G. Juhász)
4. The revised manuscript on CFTR allostery, written in collaboration with Gergely Lukacs, was sent back to Nat. Communication:
   Soya et al. Folding correctors can restore CFTR post-translational folding landscape by allosteric domain-domain coupling
5. Our manuscript is under review:
   Berta *et al.* SARS-CoV-2 Envelope protein alters calcium signaling via SERCA interactions; https://www.biorxiv.org/content/10.1101/2023.06.13.544745v1

Major achievements associated to the results of this project:

1. E. Suhajda received an UNKP fellowship to study interactions of CFTR NBD1 and R domain.
2. F. Bianka successfully defended her PhD thesis, Pazmany Peter Univ., 2023
3. T. Hegedus successfully defended his DSc thesis, Hung. Acad. Sci, 2023
4. T. Hegedus' application for "research professor" position was approved, Semmelweis Univ., 2023