Summary
In this two years project we works on intra and extracellular covalent labeling of transmembrane proteins on living cell surface. For intracellular labeling, we optimized experimental parameters for the sulfo-NHS-acetate labeling and mild permeabilization. Unfortunately, after sulfo-NHS-acetate preparation lot of extracellular lysines remained intact, and these unmodified lysines reacted with sulfo-NHS-biotin after mild detergent treatment. Thus we could not use these way of intracellular labelling for topology determination. Additionaly, we worked with rat's synaptosomes prepared in the Research Group of Proteomics at the Institute of Biology, ELTE and determined the extent of intracellularly labeled lysines. Unfortunately, the synaptosomes were not closed properly, therefore the amount of intracellularly labeled lysines were above 30%, therefore, we could not use these data for topology determination too. Although these two directions were unsuccessful for covalent labeling of transmembrane proteins in order to improve topology prediction methods, we were able two develop other techniques to improve our earlier experiments. It was the covalent modification of the carboxyl groups on the accessible cell surface, followed by the isolation and digestion of these proteins. We utilized two membrane-impermeable activating agents and one biotinylating reagent in a two-step reaction. At first, the reaction parameters were optimized on a single protein (BSA) then extended to the surface of living cells. Both flow cytometry and confocal microscopy measurements confirmed the integrity of the cells. During the experiments, 135 new topological positions were identified for 38 TMPs in HL60 cell line by 16 nanoLC-MS/MS runs and from BSA digestions, 29 amino acid carboxyl side chains were detected from the accessible surface of the protein. The determined peptides were mapped to the primary sequences of TMPs and the labeled sites were utilized as extracellular constraints in topology predictions that contribute to the refined low-resolution structure data of these proteins. In a second approach, a partial proteolysis of the cell surface was introduced to the original method. Two different serine proteases (trypsin and chymotrypsin) were utilized for the partial digestion of the HL60 cell surface. The optimal enzyme concentrations and digestion time periods were determined by flow cytometry experiments, then these parameters were used for confocal microscopy experiments to verify that the labelling agent does not penetrate the plasma membrane. The labeled membranes were isolated, SDS-PAGE was used to test the reproducibility of the preparation process, and finally the samples were solubilized and digested. ~97% of identified peptides by nanoLC-MS/MS were biotinylated. The labelled peptides were mapped to the human proteome and by filtering for at least 3 peptides for each type of experiment, altogether 223 extracellular positions were identified in 75 TMPs in HL60 cell line by 25 nanoLC-MS/MS runs. The essence of the new experimental method is the pre-digestion of the cells with the proteolytic enzymes such as trypsin and chymotrypsin prior to the Sulfo-NHS-SS-biotin labelling of the cell surface, giving rise to previously undetectable positions of TMPs that become available for the biotinylation reaction in either of two ways. 37% of all N-terminally labeled peptides was only detected in trypsin and chymotrypsin pre-digested samples, indicating the formation of previously non-existent N-terminals. This work was published in Scientific Reports too.
Besides developing these cell surface labeling techniques, we created a new computational method for supporting crystallization of transmembrane proteins and made it available for the scientific community on a web server, called TMCrys server. In the last 10 years, several prediction methods were developed to enhance the success of structure determination by estimating the chance of successful experiments. However, almost all of them mix globular and TM proteins leading to predict TMPs as 'hard to crystallize' (or somewhat equivalent) without the ability to distinguish between crystallizable and non-crystallizable TMPs. The only TMP-specific method is MEMEX but being created in 2008, the data used is outdated. We introduced the TMCrys method to aid the process of structure determination of TMPs. Since the algorithm of TMCrys requires installing some libraries and software packages hereby we introduce the TMCrys server, providing a graphical user interface for the prediction via our HPC to facilitate the usage of the method. This work was published in Bioinformatics.

In cooperating Tamás Hegedűs (SE), we created a new computational method for identifying transmembrane regions of proteins whose structure were determined by cryo-EM method. While experimental data on the boundaries of membrane-embedded regions are sparse, this information is present in cryo-EM density maps and it has not been utilized yet for determining membrane regions. We developed a computational pipeline, where the inputs of a cryo-EM map, the corresponding atomistic structure, and the potential bilayer orientation determined by TMDET algorithm of a given protein result in an output defining the residues assigned to the bulk water phase, lipid interface and the lipid hydrophobic core. Based on this method, we built a database involving published cryo-EM protein structures and a server to be able to compute this data for newly obtained structures. This work was published in Bioinformatics as well as in a book chapter.

During the experimental works, by investigation of hundreds of transmembrane proteins' sequences required to identify possible modification sites of transmembrane proteins, enabled us to identify a large number of genetic variations in the human population. The phenotypic effects of these mutations range from neutral for polymorphisms to severe for some somatic mutations. Disease causing germline mutations represent a special and largely understudied class with relatively weak phenotypes. In this analysis a large amount of disease-causing mutations were analyzed and contrasted to polymorphisms from a structural point of view. Our results delineate the characteristic features of disease causing mutations starting at the global level of partitioning proteins into globular, disordered and transmembrane classes, moving towards smaller structural units describing secondary structure elements and molecular surfaces, reaching down to the smallest structural entity, post-translational modifications. This work has been published in J Mol Biol.

In additional to these studies we cooperated with István Simon and Bálint Mészáros to reveal how folding and binding intertwine by studying sequential, structural and functional properties of protein complexes. We found that strikingly, the properties of protein interactors in terms of sequence and adopted structure are defined not only by the intrinsic structural state of the protein itself but also to a comparable extent by the structural state of the binding partner. The three different types of interactions are also regulated through divergent molecular tactics of post-translational modifications. This not only widens the range of biologically relevant sequence and structure spaces defined by ordered proteins but also presents distinct molecular mechanisms compatible with specific biological processes, separately for each interaction type. We published this work in Journal of Molecular Biology.

We bumped into another interesting question while investigated the sequences of transmembrane proteins, namely the footprints of various fruit viruses' RNAs. Fruit trees, such as apricot trees, are constantly exposed to the attack of viruses. As they are propagated in a vegetative way, this risk is present not only in the field, where they remain for decades, but also during their propagation. Metagenomic diagnostic methods, based on next generation sequencing (NGS), offer unique possibilities to reveal all the present pathogens in the investigated sample. Using NGS of small RNAs, a special field of these techniques, we tested leaf samples of different varieties of apricot originating from an isolator house or open field stock nursery. As a result, we identified Cherry virus A (CVA) and little cherry virus 1 (LChV-1) for the first time in Hungary. We presented this work in Viruses.

Altogether we published nine manuscript in this two year (plus one year prolonged) project with the support of NKFIH.