This was the hardest project in our research so far. Not just because the pandemic but the constant fluctuation and lack of the appropriate researchers. During the period of this project five members of our group left the lab. Three of them went to non-academic area for higher salaries and two members went to abroad for better research opportunities (one to EMBL, EBI in Heidelberg (Germany), the other to Hebrew University, in Jerusalem (Israel). In the beginning of the project we strictly follow the research plan, but since it is a basic research the results of experiments reveal new ways for researchers that cannot be seen before.

In the beginning of the project, we investigated the various factors needed for crystallization of native and alternatively spliced transmembrane proteins, which is not just an interesting question in the era of structural genomics, but help to save many by choosing the appropriate protein form and parameters for structure determination of transmembrane proteins. We developed a method, called TMCrys in order to predict propensity of success for transmembrane proteins crystallization based on the data formerly deposited in TragetTrack, PDB and PDBTM databases. The new prediction method reaches the highest accuracy among the similar state of the art prediction methods and is unique in the field of supporting transmembrane protein crystallization projects. Later we developed a new server for the TMCrys method, in order to allowing scientist to predict the expected success rate of crystallization of a transmembrane protein. We published these work in the Bioinformatics [1-2].

By investigation of hundreds of NGS data required to reveal the alternative transcript products of transmembrane proteins, it also enabled us the identification of a large number of genetic variations in the human population. The phenotypic effects of these mutations range from neutral polymorphisms to severe 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 accepted in J Mol Biol [3].

While the alternative transcript products of various eukaryote genes were investigated, we found 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 [4].

While we work on the problem of transmembrane protein crystallization prediction, we recognized besides protein crystallization and structure determination by using x-ray diffraction method, a new technique, called Cryo-EM developed rapidly, and is used more frequently for transmembrane proteins. It is important from the view point of experimental data on the boundaries of membrane-embedded regions because these type of data is sparse. However, this information is present in cryo-electron microscopy 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 atomic 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. We published this work in the Bioinformatics and Methods of Molecular Biolology [5,6].

Regarding alternative splicing of transmembrane proteins, the main problem is the detection of the viability of the product of alternative spliced transmembrane protein. While a plethora of NGS data are available for alternative splice products, there are little evidence that all these mRNA result in stable

folded transmembrane proteins. Therefore, we developed a new experimental setup, where we can detect not just the existence of a transmembrane protein in a living cell, but the topology of the transmembrane proteins too, by a special chemical modification of extra-cellular available carboxyl groups of these proteins. We presented the results of the primary laboratory technique in the Scientific Reports [7].
For further analysis of alternatively spliced cell surface transmembrane proteins, we developed an alternative method for their detection. In order to generate more topology information for these proteins, a new step, a partial proteolysis of the cell surface has been introduced. This step results in new primary amino groups in the proteins that can be biotinylated with a membrane-impermeable agent while the cells still remain intact. Pre-digestion also promotes the emergence of modified peptides that are more suitable for MS/MS analysis. The modified sites can be utilized as extracellular constraints in topology predictions and may contribute to the refined topology of these proteins. The developed new technique was published in Scientific Reports too [8].
Besides developing high throughput techniques for detecting the various forms of transmembrane proteins on the cell surface, we were involved in two other projects to investigate special transmembrane proteins. First, the OATP3A1 proteins that has several alternative spliced product in human cells were characterized and we showed the presence of at least two cooperative substrate binding sites in OATP3A1. Besides providing the first sensitive probe for testing OATP3A1 substrate/inhibitor interactions, our results also help to understand substrate recognition and transport mechanism of the poorly characterized OATP3A1. We published this work in Biochemical Pharmacology [9]. Second, we were able to create a monoclonal antibody recognizing an extracellular epitope of human ABCC6. The monoclonal antibody recognized human ABCC6 in the liver of hABCC6 transgenic mice, verifying both specificity and extracellular binding to intact hepatocytes. This work was published in FEBS Letters [10].
Investigation of hABCC6 is important from the view point of tissue calcification as well. Calcification of various tissues is a significant health issue associated with aging, cancer and autoimmune diseases. Pseudoxanthoma elasticum is a rare genetic disease, a prototype for calcification disorders, resulting from the dysfunction of ABCC6. It is identified by excess calcification in a variety of tissues (e.g., eyes, skin, arteries) and currently it has no cure, known treatments target the symptoms only. We presented a new zebrafish (Danio rerio) model for Pseudoxanthoma elasticum. We showed that there are two functional and one non-functional paralogs for ABCC6 in zebrafish (abcc6a, abcc6b.1, and abcc6b.2, respectively). We created single and double mutants for the functional paralogs and characterized their calcification defects with a combination of techniques. Zebrafish deficient in abcc6a show defects in their vertebral calcification and also display ectopic calcification foci in their soft tissues. Our results also suggest that the impairment of abcc6b.1 does not affect this biological process. This results also show that the various version of a transmembrane protein presented either by paralog genes or alternative splicing may functional different ways. We published this work in Frontiers of Cell and Developmental Biology [11].
Genetic modification of transmembrane proteins is an important tool to investigate their structure-function relations as well as the effect of alternative splicing. We were involved in developing a combined theoretical and experimental technique for predicting the target selectivity of various Cas9 enzymes, used in Crispr technique for genome engineering. By the developed method the cleavage activity of various Cas9 enzymes can be predicted with high accuracy, paving the way for genetic modification of transmembrane proteins in genomes. This work was accepted in Nucleic Acids Research [12].
Alternative splicing was reported to alter the localization of transmembrane proteins in polarized cells. Therefore, we investigated the distribution of transmembrane proteins in epithel cells. First, we prepared a database containing experimentally verified mammalian transmembrane proteins with splicing information as well that display polarized sorting, focusing on epithelial polarity. In addition to the source cells or tissues, homology-based inferences and transmembrane topology (if applicable) were also calculated. The database, called PolarProtDb also offers a detailed interface displaying all information that may be relevant for trafficking and or splicing: including post-translational modifications (glycosylations and phosphorylations), known or predicted short linear motifs conserved across orthologs, as well as potential interaction partners. Data on polarized sorting has so far been scattered across myriads of publications, hence difficult to access. This information can help researchers in several areas, such as

scanning for potential entry points of viral agents like COVID-19. PolarProtDb shall be a useful resource to design future experiments as well as for comparative analyses. The database is available at http://polarprotdb.ttk.hu, and were published in Journal of Molecular Biology [13]. A prediction method, that can predict the effect of alternative splicing for the routing of transmembrane proteins was also developed and we make it publicly available at http://polarprotpred.ttk.hu. This method was published in Bioinformatics [14]. Since disordered regions have important role in maintaining protein-protein interactions and hence in routing of transmembrane proteins, an other prediction method was also developed, called MemDis. MemDis utilizes convolutional neural network and long short-term memory networks for predicting disordered regions in transmembrane proteins. In addition to attributes commonly used in intrinsically disordered regions (IDR) predictors, we defined several transmembrane protein specific features to enhance the accuracy of our method further. MemDis achieved the highest prediction accuracy on TMP-specific dataset among other popular IDR prediction methods. The method is available at http://memdis.ttk.hu and was published in International Journal of Molecular Sciences [15].

Paper published:

1. Varga, JK and Tusnády, GE (2018) TMCrys: predict propensity of success for transmembrane protein crystallization. Bioinformatics 34, 3126-3130 (doi: 10.1093/bioinformatics/bty342).
2. Vargja, JK and Tusnády, GE (2019) The TMCrys server for supporting crystallization of transmembrane proteins. Bioinformatics 35, 4203-4.
3. Dobson L, Mészaros B and Tusnády GE (2018) Structural Principles Governing Disease Causing Germline Mutations. J Mol Biol S0022-2836, 31101-X (doi: 10.1016/j.jmb.2018.10.005).
4. Baráth D, Jaksa-Czotter N, Molnár, J, Varga T, Balassy J, Szabó LK, Kirilla Z, Tusnády GE, Preininger E and Várallyay E (2018) Small RNA NGS Revealed the Presence of Cherry Virus A and Little Cherry Virus 1 on Apricots in Hungary. Viruses 10, E318 (doi: 10.3390/v10060318).
5. Farkas, B, Csizmadia, G, Katona, E, Tusnady, GE and Hegedus, T (2019) MemBlob database and server for identifying transmembrane regions using cryo-EM maps Bioinformatics, btz539.
6. Csizmadia, G, Farkas, B, Katona, E, Tusnády, GE and Hegedűs, T (2020) Using MemBlob to Analyze Transmembrane Regions Based on Cryo-EM Maps. Methods Mol Biol. 2112, 123-30.
7. Muller, A, Lango, T, Turiak, L, Acs, A, Varady, G, Kucsma, N, Drahos, L and Tusnady, GE (2019) Covalently modified carboxyl side chains on cell surface leads to a novel method toward topology analysis of transmembrane proteins Scientific Reports, 9, 15729.
8. Langó T, Pataki ZG, Turiák L, Ács A, Varga JK, Várady, G, Kucsma, N, Drahos, L and Tusnády GE (2020) Partial Proteolysis Improves the Identification of the Extracellular Segments of Transmembrane Proteins by Surface Biotinylation Scientific Reports 10, 8880.
9. Bakos É, Tusnády GE, Német O, Patik I, Magyar C, Németh K, Kele P and Özvegy- Laczka C (2020) Synergistic transport of a fluorescent coumarin probe marks coumarins as pharmacological modulators of Organic anion-transporting polypeptide, OATP3A1 Biochem Pharmacol. DOI: 10.1016/j.bcp.2020.114250.
10. Kozák E, Szikora B, Iliás A, Jani PK, Hegyi Z, Matula Zs, Dedinszki D, Tőkési N, Fülöp K, Pomozi V, Várady Gy, Bakos É, Tusnády GE, Kacskovics I and Váradi A (2020) Creation of the first monoclonal antibody recognizing an extracellular epitope of hABCC6 FEBS Lett 595, 789-798.
11. Czimer D, Porok K, Csete D, Gyüre Zs, Lavró V, Fülöp K, Chen Z, Gyergyák H, Tusnády GE, Burgess SM, Mócsai A, Váradi A and Varga M (2021) A New Zebrafish Model for Pseudoxanthoma Elasticum Front Cell Dev Biol 9, 628699.
12. Tálas A, Huszár K, Kulcsár PI, Varga JK, Varga É, Tóth E, Welker Zs, Erdős G, Pach PF, Welker Á, Györgypál Z, Tusnády GE and Welker E (2021) A method for characterizing Cas9 variants via a one-million target sequence library of self-targeting sgRNAs Nucleic Acids Res 49, e31.

13. Zeke A, Dobson L, Szekeres LI, Langó T and Tusnády GE (2021) PolarProtDb: A Database of Transmembrane and Secreted Proteins showing Apical-Basal Polarity J Mol Biol 433, 166705.
14. Dobson L, Zeke A and Tusnády GE (2021) PolarProtPred: Predicting apical and basolateral localization of transmembrane proteins using putative short linear motifs and deep learning. Bioinformatics , btab480.
15. Dobson L and Tusnády GE (2021) MemDis: Predicting Disordered Regions in Transmembrane Proteins Int. J. Mol. Sci. 22, 12270.