SHORT THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY (PHD)

Structural studies on the cysteine protease of Venezuelan equine encephalitis virus

by Gyula Hoffka

Supervisor: János András Mótyán, PhD.

UNIVERSITY OF DEBRECEN
DOCTORAL SCHOOL OF MOLECULAR CELLULAR AND IMMUNE BIOLOGY

DEBRECEN, 2024.

Structural studies on the cysteine protease of Venezuelan equine encephalitis virus

By Gyula Hoffka, Chemistry MSc.

Supervisor: János András Mótyán, PhD.

Doctoral School of Molecular Cellular and Immune Biology, University of Debrecen

Head of the **Defense Committee**:	Gábor Szabó, PhD., DSc.
Reviewers:	Dóra Karancsiné Menyhárd, PhD.
	Ferenc Erdődi, PhD., DSc.
Members of the Defense Committee:	Balázs Veres, PhD.
	István Timári, PhD.

The PhD Defense takes place at the Lecture Hall of Bldg. A, Department of Internal Medicine, University of Debrecen, at 1 p.m. on 10th of December, 2024.

1. Introduction

1.1. The Venezuelan equine encephalitis virus

One of the greatest challenges in modern medicine is defending ourselves against viruses and the treatment of diseases resulting from viral infections. In contrast to bacterial diseases, viruses cannot be controlled by antibiotics. In addition to severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), which has been responsible for the pandemic that has shaken the world in recent years, there are many further dangerous viruses. These include the Venezuelan equine encephalitis virus (VEEV) which is a member of the *Togaviridae* family of Group IV(+) viruses, especifically that of the New World group of the alphavirus genus. Viruses of this genus are generally characterised by an ss(+)RNA genome, a globular shape, and glycoproteins in the cell membrane that are essential for receptor-mediated recognition. New World alphaviruses (VEEV, EEEV and WEEV) can cause encephalitis in humans, with symptoms ranging from mild to severe, occasionally resulting in death. VEEV infection typically results in flu-like symptoms, however, it can cause death, with a mortality rate of approximately 1%. Additionally, about 14% of cases exhibit neurological symptoms. As of today, currently there are no antiviral drugs approved by the US Food and Drug Administration (FDA) to treat the infection, and the development of potential vaccines is still in progress. These include a vaccine that is based on the use of virus-like particles which has already undergone clinical trials, as well as a DNA-based vaccine which has demonstrated effectiveness in macaques. Climate change may also contribute to the potential spread of the virus in the future, potentially leading to the emergence of VEEV in regions where it was not previously present. Global travel networks may also ease its spread, enabling the virus to appear anywhere in the world. Considering all these factors, an antiviral drug or vaccine against VEEV would be of a high priority.

1.2. The nsP2pro protein

The VEEV genome contains two open reading frames (ORFs). The first ORF encodes a polyprotein containing non-structural proteins (nsP) and the second encodes structural proteins. After translation, the polyprotein is cleaved into separate nsP1, nsP2, nsP3 and nsP4 non-structural proteins via proteolytic cleavage. The non-structural protein 2 (nsP2) plays a crucial role in the life cycle of the virus. Several important regions have been identified in alphaviral nsP2 proteins: the N-terminal region (Gly1-Ile456) was found to have ATPase and

GTPase activity, as well as RNA helicase activity, while the C-terminal region (Met457-Cys794) is composed of a papain-like cysteine protease domain followed by an S-adenosyl-L-methionine (SAM) dependent RNA methyltransferase (SAM MTase) domain. The non-structural protein 2 protease of VEEV, nsP2pro (EC 3.4.22.B79) is a cysteine protease being essential for the life cycle of the virus. The protease domain is linked to the SAM MTase domain which has been shown to play a role in proteolysis: the Arg662, Lys705 and Lys706 residues are involved in the formation of the substrate-binding cleft, therefore, facilitate substrate recognition. The catalytic diad of nsP2pro is composed of a cysteine and a histidine residue, Cys477 and His546, the substrate binding site is located between the protease and SAM MTase domains. The most important role of nsP2pro is the cleavage of the nsP1234 protein at the nsP1/nsP2, nsP2/nsP3 and nsP3/nsP4 sites. To denote the substrate-binding subsites and the substrate residues binding to these sites we can apply the nomenclature introduced by Schechter and Berger in 1967. Accordingly, the substrate residues are referred to by increasing numbers starting from the clevage position towards the N terminus (P1, P2, etc.) and towards the C terminus (P1', P2', etc.). The substrate-binding site of the enzyme is divided into subsites, each responsible for binding one of the substrate residues, these sites are referred to as S2, S1, S1', S2', etc., according to the substrate numbering.

The autoproteolytic cleavage site sequences of alphaviruses show a high degree of similarity, typically containing a residue having a smaller side-chain at the P1 position, followed by a glycine at the P2 position. The residue at the P3 position is conserved, while P4 is more variable. A higher variability is observed for P1'-P4' residues.

Among the substrate-binding subsites of VEEV nsP2pro, the Asn475, Val476 and Ala509 residues play a crucial role in the formation of the S1 subsite, while the main-chain amides of Val476 and Cys477 form an oxyanionic cavity. The S2 subsite resembles the so-called "glycine-specific motif" that is found usually in cysteine proteases, as it is formed by a highly conserved Trp547 and the catalytic His546 before it, and is responsible for binding the glycine of the P2 position. The SAM MTase domain plays a crucial role in the formation of the S3 subsite which is formed by Ile698 and Met702, flanked by Ala509 and His510 residues.

VEEV nsP2pro is an important drug target owing to its critical role in the viral life cycle, as well as its activity against the interferon defence mechanism, amongst others. The protease is able to recognize and cleave the TRIM14 protein – which has a role in the immune response of infected host cells – at a cleavage site which is similar to those of the viral polyprotein. These sequences are known a short stretches of homologous host-pathogen protein sequences

(SSHHPS). These observations suggest that VEEV nsP2pro can be considered as an important target for drug design.

1.3. The crystal structure of alphavirus nsP2pro proteins

Among the alphaviruses, the structures of the protease domain was determined only in the case of VEEV, the Chikungunya virus (CHIKV) and the Sindbis virus (SINV). The most structural information are available for VEEV nsP2pro. The first VEEV nsP2pro structure was described in 2006 (PDB ID: 2HWK), for which the protein containing the 457-794 region of nsP2 was expressed and crystallised, followed by an X-ray crystallography experiment that was able to determine the structure of the protease and SAM MTase domains (Asp468-Ser787) with a resolution of 2.45 Å.

In a study published in 2016, the structure of VEEV inhibited by the nsP2pro inhibitor was determined (PDB ID: 5EZS). Subsequently, in a study published in 2017, a mutant VEEV nsP2pro was also investigated which enzyme exhibited relatively lower enzymatic activity. The enzyme contained the N475A mutation, the crystal structure was determined for the Gln471-Glu791 region with a resolution of 2.10 Å (PDB ID: 6BCM). An alternative conformer of the N-terminal region was observed in the structure, in which the N-terminus (Gln471-Cys477) appeared to bind into the active site, occupying the substrate binding site. The authors of the study referred it to as a "self-inactivated" form ($6BCM^{self\text{-}inactivated}$), while the main conformer was considered to be the active conformer ($6BCM^{active}$). According to the electron density map, the active and inactive conformers were present in the structure with a distribution of 60 and 40%, respectively. Recently, in a study published in 2023, the structure of VEEV nsP2pro was determined co-crystallized with an inhibitor that was developed by using *in silico* methods (PDB ID: 8T8N).

1.4. Crystal structure determination

Multiple experimental methods are available to study protein structures, the most widely applied method is X-ray crystallography. The method requires the expression, purification and crystallisation of the examined protein. The resulting single crystal is then irradiated with X-rays and the electron density map is calculated based on the diffraction images. The position of the individual atoms is determined based on the electron density map, into which the atoms are inserted, followed by the building of the three-dimensional protein structure. The determined structures of the proteins are deposited in the RCSB Protein Data Bank (PDB) database.

1.5. Molecular mechanics

A major challenge for the experimental methods is to study the real-time behaviour of the individual protein molecules and their conformational changes. The molecular mechanics methods provide the possibility to study changes in the dynamical properties of individual molecules, as well as ensembles of molecules. The basic principles of the method follow classical, deterministic Newtonian mechanics, accordingly, molecules are described as model atoms connected by springs. The individual atoms have a defined charge, mass and size, and their movement is determined by their interactions with the other atoms. The type of the bonds between atoms is predetermined, which precludes the possibiltiy to model electron transfer with molecular dynamics. Molecular dynamics methods are of significant importance for structural studies.

The molecules to be investigated are described by a properly parameterised force field, which is a function of several terms that, along the position of individual atoms, serves as the basis for the determination of energy as a function of atom positions and atom types. Typically, bond distances, bond angles, dihedral angles, Lennard-Jones potential and Coulumb interaction are included as members of the formula describing the force field. The topology of each individual molecule with a given electron structure, including proteins, water molecules, ions, etc., must also be given. The topology cannot be changed during the simulations.

A prerequisite of the application of the method is a suitable structural model, preferably an experimentally defined structure, or if necessary, a structure prepared by homology modeling or by machine learning methods. It should be noted, that crystal structures, as well as homology models or structures estimated by AI-based methods (*e.g.* AlphaFold) represent the structure of the given protein in the crystal phase rather than in a living cell or an *in vitro* experimental environment.

To be able to run simulations, we need to prepare the protein structure, check whether a monomer or dimer assembly is required, remove unnecessary molecular fragments (such as inhibitors), check for alternative conformers and adjust the protonation states of each protonatable side-chain, and if necessary, apply end-capping groups to the protein chain endings, or adding missing atoms.

Modern simulations are almost exclusively run in the presence of an explicit solvent for which the selection of a water model and the addition of solvent molecules is required. In general, molecular dynamics simulations are performed by applying periodic boundary conditions which is most often a simulation box with a cubic or truncated octahedral geometry.

Ions are added to neutralise the system, and the ionic strength can be adjusted for physiological models.

The first step of the simulations is usually a minimization of the structure, to eliminate high-energy conformations arising from geometric distrotions. It is followed by the molecular dynamics simulation, initiated with preset parameters and randomly generated starting velocities of atoms. During the simulation, the software computes the forces acting on the atoms in each step based on the force field, and updates the position of the atoms according to the simulation time interval. Significant control is provided over the parameters of the system, including the specification of the simulation conditions as well as the properties of the ensemble. The most commonly used conditions are the so-called NVT, or canonical and NPT, or isothermal-isobaric conditions. In the formulas, N represents the atomic number, V the volume, T the temperature, P the pressure, in the cases of N and V, are kept close to a preset constant. or in the case of P and T, to an approximate value. During the simulation, the algorithm updates the system at specified intervals, with the possibility of specifying the sampling period: the time interval after which the system outputs the coordinates obtained.

The starting crystal or modelled structure can significantly differ from the structure that is actually present in a physiological solution. By applying the *simulated annealing* method, the actual structure in the solution can be deduced, by heating the model system to a temperature higher than the physiological temperature, equilibrating at that temperature, followed by cooling to a low temperature, repeating the process over multiple cycles.

To study the complex of a protein and its ligand (*e.g.* substrate), it might be necessary to apply constraints because the substrate is likely to react with the enzyme shortly after binding, whereas in a molecular dynamics simulation, a reaction cannot be modelled, and it may be necessary to keep the substrate artificially in the active site.

1.6. Analysis of simulation results

One way to examine the quality of the molecular dynamics simulations run is by calculating the root mean square deviation (r.m.s.d.) values. The method can be applied to check the accurate reproduction of the behaviour of proteins with a more rigid, defined structure, obtained from crystal structures. However, it is important to note in case of highly flexible, dynamic proteins or protein fragments (*e.g.* disordered proteins), as in their case, convergence is achieved when the fluctuation are within a range, with the average value being unchanged.

Multiple methods are available to study the dynamic properties of proteins, including the measurement of various geometric parameters, which can be used to confirm the presence of various interactions.

It is of crucial importance to study the presence of hydrogen bonds which provide one of the strongest secondary interactions. Simulations provide an excellent opportunity to study the hydrogen bond interaction networks. A crystal structure represents only a particular state of the protein, more importantly an average of the individual molecules in the crystal, and does not even represent the structure in the solvent phase, while a simulation of the sufficient quality can reveal very valuable details about the hydrogen bond interactions present in the aqueous medium.

During the simulations, certain regions of the protein may have higher conformational flexibility, therefore, the long trajectories usually result in a number of structures that show various degrees of differences. It may be necessary to compare and classify the resulting structures and to select the representative structure(s) by so-called clustering.

1.7. Application of simulation sotwares

Multiple molecular dynamics software are available, several of them are free, including Amber, GROMACS and NAMD. For the presented work, we have applied the Amber16 software, that contains several programs and scripts with different functions. It consists of two parts, AmberTools and the Amber computational software package. AmberTools contains all of the scripts needed to prepare a protein, run simulations and evaluate the results. It includes the parameters for each force field, the *tleap* script for preparing structures, and the *antechamber* program for parameterising further molecules. The simulations can be run with the *sander* program while the results can be analyzed by using the *cpptraj* script. The *pmemd* script, which is more powerful than the computational programs available in AmberTools, is included in the Amber computational package and it is required for running simulations significantly faster, taking advantage of graphics processing unit (GPU) acceleration. All of the presented calculations were run with Amber16 software, using the GPU-accelerated version of the *pmemd* algorithm. The preparation of the structures and the analysis of the results were performed using the scripts available in AmberTools.

Molecular dynamics simulations usually require large computational resources, and generate large amounts of data. Generally, a supercomputing (HPC) environment - that uses several high-performance CPU cores and GPUs - is required for running multiple simulations

in parallel. In this environment, the users are supported by pre-installed softwares, parameterised according to the resources of the HPC. In addition, expensive softwares may be available under academic licences. Modern MD programs are highly optimised to run on GPUs, speeding up the work many times over traditional CPU runs. We can highlight the role of the CUDA programming interface for achieving such speed, which most molecular dynamics software that uses GPU acceleration rely on.

2. Aims

The aim of the studies presented in the doctoral thesis was to investigate the structural features of the non-structural protein 2 protease of VEEV (nsP2pro) using computational chemistry methods, with a special emphasis on the inter- and intramolecular interactions at the active site. The structural investigations were part of experimental studies aimed at determining the specificity of the enzyme through *in vitro* activity measurements, as well as resolving the enzyme's structure using X-ray crystallography.

I. Investigation of the substrate specificity of VEEV nsP2pro

In order to support the experimental specificity studies, our aim was to model a complex of VEEV nsP2pro with an oligopeptide substrate, and then analyze the complex with molecular dynamics simulations, as well as to examine the structural and energetic properties of substrate variants for the identification of the structural features determining the amino acid preferences.

II. Crystal structure analysis of VEEV nsP2pro

To carry out an in-depth investigation of the active and self-inactivated conformational states observed in the crystal structures, we aimed to map the structural features that determine the stability of the enzyme, by comparing the available crystal structures and performing molecular dynamics simulations, with a particular focus on the effects of mutations, as well as the interaction networks of the active site.

3. Materials and methods

All of the calculations were carried out by using HPC resources provided by the Governmental Information-Technology Development Agency (GITDA, formerly NIIF).

3.1 Investigation of the substrate specificity of VEEV nsP2pro

The initial structure used for the specificity assay was a complex model structure of the VEEV nsP2pro and its natural oligopeptide substrate. The protein structure included the 469-767 residues, it provided by Dr. Patricia M. Legler (Center for Bio/Molecular Science and Engineering, U.S. Naval Research Laboratory).

The sequence of the original model and the experimentally investigated protein were aligned by using the Clustal Omega software. The peptide substrate was modified to represent the SFV nsP1/nsP2 cleavage site (EYHAGA↓GVVETP), which was also used in the *in vitro* experiments. Two C-terminal residues of nsP2pro were modified by using FoldX software, to resemble the experimentally investigated protein. We have applied end-capping groups at the termini of the enzyme and substrate, as well. The protonation state of protonatable amino acid side-chains was set according to PROPKA 3.1 calculations, and the proton position of neutral histidines was determined using Chimera software. The Cys477 active site residue was modelled to have a deprotonated side-chain, while the catalytic His546 residue had a protonated side-chain. The preparation and addition of the hydrogens was performed by using the *tleap* script of Amber16. The ff14SB force field was used to describe both the protease and the substrate, and the water molecules were described using the TIP3P model.

The Amber16 software was used to run the calculations. After an initial minimization, molecular dynamics simulations were performed under NPT conditions with 1 fs step size. First, we heated the system to 300 K during a 2 ns simulation, followed by equilibration at 300 K for 2 ns. The Langevin dynamics method was used to control the temperature. For bonds containing hydrogen as one of the atoms, the SHAKE algorithm was applied. The system was further heated to 400 K during a 0.5 ns simulation and then equilibrated at 400 K for a further 0.5 ns. This was followed by cooling to 5 K during a 1 ns simulation using the simulated annealing method.

Then, we heated the system again to 400 K in a 0.5 ns simulation, carrying out a further 0.5 ns equilibration simulation at that temperature. We repeated this simulation cycle two times again, but in the second case, the system was heated only to 300 K. An additional 0.5 ns long

equilibration simulation was run at this temperature. For the complete duration of the simulations, a 50 kcal mol^{-1} $Å^{-2}$ constraint was applied between the sulphur atom of the catalytic diad Cys477 and the substrate P1 alanine carbonyl oxygen. A bond of the same value was applied between the nitrogen atom of the catalytic donor His546 and the main alkane nitrogen atom of the substrate P1' glycine.

The convergence was determined based on the r.m.s.d. values obtained from the final equilibration step, considering the average structure as a reference. The DynaMut web server was used to prepare the complexes containing the modified substrates. In total, one P5, four P4, two P2, two P1, 20 P1' and two P2' mutant substrates were investigated. Based on the definition of folding free energy, the ≥0 and the <0 kcal/mol changes (ΔΔG) were considered to be stabilizing and destabilizing, respectively. The LigPlot+_v2.1 software was used for two-dimensional visualization of interactions, and the PyMOL software was used for spatial visualization of structures.

3.2. Determination of the VEEV nsP2pro crystal structure

The expression and crystallization of the recombinant protein, as well as the resolution of the structure was performed by George T. Lountos, Danielle Needle, Alexander Wlodawer and David S. Waugh. The examined protein contained the Arg463-Thr785 region of VEEV nsP2, with the resolved structure encompassed almost the full-length protein (Asn472-Thr785). The protein contained K741A and K767A surface entropy-reducing mutations in the SAM MTase domain of nsP2pro, while the cysteine protease domain contained no modifications. The crystallised protein was subjected to X-ray diffraction analysis and the resulting crystal structure is available in the Protein Data Bank database, with identifier 8DUF.

3.3. Molecular dynamics study of the VEEV nsP2pro crystal structure

Three nsP2pro structures were analyzed by molecular dynamics: our novel crystal structure that the contains the K741A and K767A surface mutations (PDB ID: 8DUF), as well as the structures of the wild-type (PDB ID: 2HWK) and the N475A mutant nsP2pro (PDB ID: 6BCM). All non-protein atoms present in the crystals were removed except for water molecules. The studied 8DUF structure included the Asn472-Ala768 residues. The active and self-inactivated structures were investigated separately. The PROPKA 3.1 software was used to determine the protonation states of the titratable amino acid side-chains. We have added acetyl

or N-methylamide groups to the termini by using GaussView 6. Hydrogens were added to the structures by using the *tleap* module of AmberTools16. The catalytic residues were in their charged form (deprotonated cysteine and protonated histidine).

We added explicit solvent water molecules to the systems by using the Split charge method for adjusting the number of sodium and chloride ions while considering the physiological ion concentrations and the charge of each protein. The ff14SB force field was used to parameterize the protein while the water molecules were described using the TIP3P model.

The Amber16 software was used to carry out the calculations, following an initial minimization the step size was set to 1 fs, and the systems were first heated to 300 K in a 1 ns simulation using NVT conditions, followed by a 1 ns NVT equilibration at 300 K. This was performed by using a positional restraint of 10 kcal mol^{-1} $Å^{-2}$ on protein atoms. An additional equilibration was also run under NPT conditions for 0.5 ns. After the equilibration, the systems were heated to 400 K during a 0.5 ns simulation. At this temperature, we further equilibrated the systems for 1 ns, then cooled them to 5 K during a 2.5 ns simulation. After this, the systems were heated again to 400 K, equilibrated, and cooled to 5 K, according to the protocol of simulated annealing. The durations of the calculations were the same. A 5 kcal mol^{-1} $Å^{-2}$ positional binding constraint was applied for the protein atoms. We have run two further cycles, using positional binding of 3 kcal mol^{-1} $Å^{-2}$ and 1 kcal mol^{-1} $Å^{-2}$ respectively, for the main-chain atoms of the protein only. Finally, the systems were equilibrated at 5 K for 1.5 ns without applying any constraints, heated to 300 K in a 1.5 ns simulation, and equilibrated at 300 K for a further 5 ns. Then, so-called "production run" calculations were performed for the evaluation that were run at 300 K for 50 ns without constraints. Sampling of the coordinates was performed at every 5 ps. For each structure we investigated, we ran three separate simulations each.

The r.m.s.d. values were calculated for the structures obtained from the production run simulations, the average coordinates of every tenth structure were used as reference. The network of hydrogen bonds formed by the N-terminal regions was also investigated, the three parallel trajectories were evaluated together. We have defined significant hydrogen bonds as those that are present in at least 10% of the trajectory, assuming a high likelihood of their occurence. We have performed multiple distance measurements to determine the dynamic properties of the proteins. Each independent simulation was treated separately, and the resulting distances were examined as function of simulation time, analyzing each independent trajectory individually. We have also investigated the active site interactions of the self-inactivated

conformers by clustering the computated structures. Following clustering, the centroid of the cluster with the highest percentage ratio was used as the represetnative structure. The interactions were mapped by using the LigPlot+ software. The previously described binding model of the enzyme-substrate complex was used for comparison.

The molecular dynamics of the crystal structures were carried out with a longer protocol, applying *simulated annealing*, to map the interactions of the N-terminals. While for the substrate specificity study, our aim was to equilibrate a representative substrate-enzyme complex, to use it as a starting point to study different variants with the DynaMut software.

4. Results

4.1. The substrate specificity of VEEV nsP2pro

Our aim was to examine the specificity of VEEV nsP2pro with both experimental and computational methods. In our work, amino acid preferences of the S5, S4, S2, S1, S1′ and S2′ substrate binding pockets were investigated *in vitro* by using recombinant protein substrates representing the wild-type nsP1/nsP2 of SFV cleavage site (EYHAGA↓GVVETP) and its P5, P4, P2, P1, P1′ and P2′ variants. In total of 31 different substrates were designed and examined *in vitro* (work of Dr. Beáta Bozóki). In order to explain the differences between the experimentally determined cleavage efficiencies and kinetic parameters, *in silico* analysis of the structures was performed. For this purpose, I have prepared a model structure of the enzyme in complex with an oligopeptide substrate (twild-type SFV nsP1/nsP2 cleavage site) that was used in the experimental assays, as well. The prepared model was then investigated by molecular dynamics simulations. The r.m.s.d. was calculated for the last steps of the simulation (so-called "production run"), and we did not observe significant fluctuations. The structure obtained from the last step was considered as a representative complex of the VEEV nsP2pro enzyme and the oligopeptide substrate.

The equilibrated enzyme-substrate complex structures was used to prepare the series of structures containing modified substrates by using DynaMut software (work of Dr. János András Mótyán), the model complexes were used to investigate the enzyme-ligand interactions and determine the structural determinants of the VEEV nsP2pro's amino acid preferences.

The P5 tyrosine was modified to glutamine. For this variant, DynaMut calculations showed that the mutation weakens the enzyme-substrate interaction ($\Delta\Delta G$=-0.222 kcal/mol) which was in agreement with the experimental results. In the case of the P5-Gln-containing substrate, the less efficient *in vitro* cleavage might be due to the fact that the wild-type tyrosine forms two hydrogen bonds at the S5 subsite (side-chain and main-chain-mediated interactions with the Ser371 and Ser701 residues), which interactions were not detected in the case of the P5-Gln mutant substrate.

The wild-type histidine at the P4 site was modified into several different amino acids. The $\Delta\Delta G$ values calculated with DynaMut showed good correlation with the experimentally determined values, which showed higher proteolytic activity for all variants as compared to the wild-type. At the S4 subsite, polar enzyme-substrate interactions are predominant, the S4-Lys706 side-chain is able to form a salt bridge with the S4-Gln side-chain. Consistently, in our

experimental studies, cleavage of P4-Gln was found to be more efficient compared to the wild-type substrate. In agreement with this, the VEEV nsP1/nsP2 and nsP2/nsP3 natural cleavage sites also contain glutamine at this position. The lowest activity was observed for the P4-Gly mutant, suggesting that side-chain-mediated polar interactions at the S4 subsite are indeed important for substrate recognition.

The wild-type cleavage site contains glycine in P2 position. Mutation of this residue to alanine or valine has destabilizing effects due to the loss of a main-chain hydrogen bond. The folding energy difference calculated with DynaMut was -0.536 kcal/mol for P2-Ala and 0.495 kcal/mol for P2-Val. Significantly lower activity values were determined for both mutants *in vitro* than for the wild-type, and based on the structural analysis the binding of valine and alanine P2 side-chains is less favorable because these residues have higher volume than glycine.

The wild-type P1 alanine residue forms hydrophobic interactions with the S1 residues. In the case of the P1 glycine variant, a moderate change was predicted (0.188 kcal/mol). This was consistent with the results of experimental studies which revealed highly similar activities for the wild-type (P1-Ala) and mutant (P1-Gly) substrates. A more significant change in enzyme-substrate interactions was predicted for the P1-Val mutant (1.207 kcal/mol). Accordingly, the mutant substrate was cleaved by the enzyme only to a negligible extent. Our results are in agreement with observations from previous studies, suggesting the binding of the branched apolar side-chains is less favorable to the S1 subsite due to their lower flexibility.

The backbone of the wild-type glycine at the P1' position can form a hydrogen bond with the side-chain of Arg662 residue. All possible P1’ variant substrates were investigated *in vitro*, the determined activities showed the best correlation with the hydrophobicity of the entire cleavage site, owing to the fact that the enzyme forms a hydrogen bond with the main-chain of the substrate in this position. Based on our results, it is likely that the amino acid preference of the enzyme P1' is not entirely independent of the side-chain. Apart from glycine, the highest activity values were observed for polar (Thr and Ser) or aromatic (Tyr, Trp, Phe) residues. All naturally occuring cleavage sites of nsP1/nsP2, nsP2/nsP3 and nsP3/nsP4 of alphaviruses contain Gly, Ala and Tyr residues at the P1' position, like VEEV. The highest activity has been determined for the P1'-Tyr variant where additional hydrophobic interactions can occur due to the aromatic side-chain, and the hydroxyl group of tyrosine may form an additional hydrogen bond with the main-chain of Leu665. The hydrophobic interactions may be formed by Phe or Trp side-chains, as well, which is in agreement with the relatively high activities observed for the substrates containing aromatic residue in P1’ position. The calculated folding energy

differences did not show significant correlation with the experimentally determined relative specific activity values, owing to the fact that in this position, a „break" can be observed in the substrate. The introduction of a larger side chain can lead to severe structural distortions, and the resulting larger energy changes are less predictable with DynaMut

The nsP1/nsP2, nsP2/nsP3 and nsP3/nsP4 autoproteolytic cleavage sites of most alphaviruses contain hydrophobic/apolar residues in P2' position (Pro, Leu, Ile or Val), while for certain alphaviruses, including VEEV, the nsP1/nsP2 cleavage site contains a polar Ser at this position. In the experimental studies, my colleagues have mutated the P2' valine of the wild-type nsP1/nsP2 of SFV cleavage site to serine or to proline. Previous studies have revealed that the contribution of the S4-S1' subsites to substrate recognition is significant. The substitution of wild-type valine with serine did not result in significant differences of the kinetic parameters, the k_{cat}/K_M was 0.06 $mM^{-1}s^{-1}$ in both cases. The P2' side-chain is exposed to the solvent, thus, mainly the main-chain atoms of the substrate interact with the enzyme. The calculated change of the folding energy was -0.068 kcal/mol which was considered to be more moderate. As compared to this, the calculated folding energy difference for P2'-Pro was more remarkable (0.427 kcal/mol). The experimentally determined specific activities were highly similar in the case of the wild-type as well as the P2'-Pro and P2'-Ser variants. The calculated and experimentally determined values for the S2 pocket can be explained by the fact that the S2' pocket is less well-defined and more opened towards the solvent as compared to the S4-S1' pockets. Previous studies have suggested that P2'-P6' residues are exposed to the solvent and the main-chain-mediated interactions are predominant in positions P1'-P6'.

4.2. Determination and analysis of the VEEV nsP2pro crystal structure

We have determined the structure of VEEV nsP2pro with X-ray crystallography. The expressed recombinant protein contained the Arg463-Thr785 residues of the VEEV nsP2. The SAM MTase domain contained two surface mutations (K741A and K767A) in order to improve diffraction via reducing surface entropy. The cysteine protease domain did not contain any modifications. The structure of the crystallized recombinant protein (Arg463-Thr785) was resolved along almost its entire length (Asn472-Thr785), but a part of the Tyr764-Tyr774 surface loop (Arg769-Tyr774) was not visible in the electron density map, presumably due to the relatively higher flexibility of this region. The resolution of the defined structure was

relatively high: 1.46 Å. In the crystal structure determined from the electron density map, the N-terminus exhibited two different conformations (8DUF^{active} and 8DUF$^{self\text{-}inactivated}$).

In case of the 8DUF^{active} conformer, the structure contained the residues from the Lys473 residue, whereas for the 8DUF$^{self\text{-}inactivated}$ from Asn472. At the N-terminus, residues until Cys477 had alternative conformers. The N-terminus appeared to be exposed to the solvent in case of the 8DUF^{activate} conformer, while in the case of 8DUF$^{self\text{-}inactivated}$ conformer it flipped to the active site and fitted to the interdomain cleft.

The self-inactivated conformer that contains the N-terminus occupying the S5-S1' substrate binding sites can be considered as catalytically inactive. This conformational state corresponded to that previously observed for the N475A active site mutant enzyme (6BCM$^{self\text{-}inactivated}$). The self-inactivated state was not observed for the wild-type enzyme (PDB ID: 2HWK).

Based on the comparison of the known VEEV and CHIKV nsP2pro structures we found that the N-terminus was exposed to the solvent in the active conformations, although, the conformation of the N-terminus was different if the enzyme contained the N475A mutation (6BCM^{active}).

In the 8DUF crystal structure, the Arg769-Tyr774 residues did not appear in the C-terminal region, and additionally, differences were observed for this region as compared to the structures lacking mutations of the SAM MTase domain (5EZS, 2HWK, 6BCM), suggesting that the Tyr764-Tyr774 loop has an increased flexibility as a result of the mutations.

In the case of CHIKV nsP2pro, the β-strand which contains the catalytic His548 is able to prevent substrate binding by reducing the distance between the protease and SAM MTase domains, and a similarly low distance was observed in case of the N475A mutant VEEV nsP2pro. However, we did not observe a smaller distance in the case of 8DUF, suggesting that the N475A mutation is unlikely to play a role in the formation of the self-inactivated form.

4.3. Molecular dynamics study of VEEV nsP2pro

4.3.1. Examination of the molecular dynamics simulations

During the "simulated annealing" phase of the protocol, the geometry of the structures was evaluated by visualisation. After the equilibration phase, we have verified the "production" calculations by determining the r.m.s.d. values, using the average coordinates of every tenth structure as reference.

The results showed that there were no significant conformational changes in any of the conformers of any of the structures, which was a prerequisite for further investigations. A higher fluctuation rate was observed for the self-inactivated structure of the N475A mutant (6BCM$^{self\text{-}inactivated}$) as compared to its active conformer (6BCM^{active}), suggesting a higher flexibility of the self-inactivated structure of the mutant. We have also examined the final N-terminal conformations of the production runs. In case of 6BCM^{active}, an intermediate structure, falling between the active and self-inactivated conformers, was observed in one of the simulations.

4.3.2. Interactions of the VEEV nsP2pro N-terminal region

We have mapped the interactions formed by the N-terminal region of nsP2pro for all of the studied structures. The hydrogen bonds showed no significant differences between the active and self-inactivated conformers of the N475A mutant (6BCM) and the enzyme containing the wild-type Asn475 active site residue and the K741A/K767A mutations (8DUF). The catalytic residues of opposite charges (Cys477 and His546) formed hydrogen bonds in all trajectories. It is important to highlight the differences of the 475th residue. In the 8DUF structure, i.e. in the presence of the wild-type Asn475 residue, the active conformer formed four hydrogen bond interactions (8DUF$^{MD\text{-}active}$), out of which two side-chain mediated and one main-chain hydrogen bond were formed with Asp507, and one side-chain hydrogen bond with Arg662 residue. In all cases, the side-chain-mediated hydrogen bonds appeared in less than half of the complete trajectories. The main-chain hydrogen bond was present in 78% of the trajectories, indicating a stronger interaction.

In the case of 8DUF, the self-inactivated form (8DUF$^{MD\text{-}self\text{-}inactivated}$) showed a very different interaction network: the side-chain of Asn475 formed a hydrogen bond with both the 472nd and 473rd N-terminal residues. Their presence was 11% and 24% respectively, indicating a highly fixed and internally ordered structure of the N-terminus. Lys473 and Ala474 also formed additional hydrogen bonds with other residuess in the protein, and Asn475 formed two hydrogen bonds with His510, with a significantly higher prevalence in the simulations (25% and 34%).

In the 6BCM$^{MD\text{-}active}$ structure, the methyl group of the alanine side-chain of the N475A mutant was unable to form any hydrogen bond interactions. The hydrogen bond between the alanine and the side-chain of Asp507 had a significantly lower prevalence (13%), suggesting a weaker interaction. Neither Lys473 nor Ala474 formed hydrogen bonds, indicating lower

stability of the structure. In the 6BCM$^{MD\text{-}inactivated}$ structure, however, the lack of side-chain-mediated hydrogen bonds does not appear to have significantly affected the main-chain hydrogen bonds of N475A. The asparagine side-chain at the 475 position can form multiple hydrogen bonds which may play a significant role in the stabilization of both active and self-inactivated conformers. Due to the N475A mutation, the loss of side-chain hydrogen bonds results in significant differences in the hydrogen bond interaction networks of the N-terminus, which is even more pronounced in the active conformer.

4.3.3. The β-hairpint and Asn545 residue

The Asn545 residue is located in a loop of the cysteine protease domain which is referred to as β-hairpin. This hairpin was found to have a relatively lower B-factor in both self-inactivated conformers. To determine the interdomain distances, we measured the distance between the center of mass of Asn545 and Leu665 residues. We hypothesized that any change in this distance might indicate the change of the relative distance between the cysteine protease and SAM MTase domains because these residues are located at the domain interface. We observed similar fluctuations of the domain distance in the case of the active and self-inactivated conformers, as well. In all cases, the distance was between 5-10 Å, suggesting that the N475A mutation does not significantly alter the mobility of this loop.

4.3.4. Distance between Asn545 and Val476 residues

The movement of the β-loop was previously observed in the N475A mutant structure, indicating that the hydrogen bond interaction between N545-O and V476-NH atoms may be responsible for the relatively lower flexibility of the β-loop in the self-inactivated conformer. In our 8DUF simulations, a hydrogen bond between the main-chain oxygen of Asn545 and the main-chain of Val476 was observed. This observation was confirmed by the examination of the intramolecular interactions of the N-terminus, as we found that the terminal region interacts with the β-hairpin (Asn545) in the self-inactivated conformer, but no interaction was observed in the active conformer.

Based on molecular dynamics simulations, the hydrogen bond interaction between Asn545 and Val476 was even more pronounced in the self-inactivated conformer of the N475A mutant structure (31%) as compared to the wild-type Asn475 (13%), indicating a stronger interaction.

We have also calculated the distance between the center of mass of Asn545 and Val476 residues in the trajectories. In the 8DUF structure, the distance calculated for the active conformer was between 10-15 Å, while the distance for the self-inactivated conformer showed significantly lower fluctuation (5-8 Å). The values calculated for the $2HWK^{MD\text{-}active}$ structure were in good agreement with those calculated for the $8DUF^{MD\text{-}active}$. The values obtained for the $8DUF^{MD\text{-}self\text{-}inactivated}$ and the $6BCM^{MD\text{-}self\text{-}inactivated}$ structures were similar, suggesting structural similarity between the self-inactivated conformers, even in the presence of the N475A mutation. Although, the values calculated for the $6BCM^{MD\text{-}activate}$ showed a larger fluctuation in several trajectories as compared to the enzymes lacking the N475A mutation, the distance was even similar to that of the self-inactivated conformers, suggesting that the N475A mutation possibly destabilizes the active structure. Furthermore, the consequence of the decrease in the distance may be a more easy conversion of the active conformer into the self-inactivated.

4.3.5. Interactions of the active site

The $8DUF^{self\text{-}inactivated}$ conformer is equivalent to the self-inactivated conformer of the N475A mutant ($6BCM^{self\text{-}inactivated}$), in which the N-terminal region is flipped into the interdomain cleft and the ^{472}AKANVC^{477} residues bind into the active site. In both cases, Cys477 occupies the S1' subsite while the other N-terminal residues (472-476) bind to the S5-S1' subsites. In case of the self-inactivated conformer, the N-terminal region binds to the active site resembling the binding mode of the natural substrate, thus, it can be considered as a pseudo-substrate in the self-inactivated conformers. We have mapped the main interactions formed by the pseudo-substrate at the active site (intramolecular interactions) which were compared to those of an oligopeptide substrate (intermolecular interactions).

To investigate the interactions between the oligopeptide substrate and the enzyme at the S2-S7 sites, we used the model complex of VEEV nsP2pro and an oligopeptide substrate representing the VEEV wild-type nsP1/nsP2 cleavage site, provided by Dr. Patricia M. Legler. In order to compare the binding of the pseudo- and the oligopeptide substrates, we mapped the interactions of the P2-P7 residues (^{2}QEAGAG^{7}) of the peptide substrate. We found that the number of hydrogen bonds formed between the enzyme and the substrate is relatively low which is in agreement with the previously described findings that mainly van der Waals interactions are responsible for the substrate binding at the S4-S1’ sites.

The comparison of the inter- and intramolecular interactions revealed that there are no significant differences in the interactions between the self-inactivated conformers of the wild-type and the N475A mutant enzymes, indicating their high similarity. The binding of the pseudo- and oligopeptide substrates showed fundamental similarities, residues of both "substrates" occupy the S5-S1' subsites and their interaction networks are also highly similar. Besides the similarities, significant differences were also observed for the inter- and intramolecular interactions which might be considered while designing inhibitor molecules.

4.3.6. Accessibility of the active site

Previous molecular dynamics studies have found that the substrate representing the P6-P6' residues of the VEEV nsP3/nsP4 cleavage site (RFDAGA↓YIFSSD) can undergo conformational changes, and the P1'-P6' residues can be exposed not only towards the protease domain but also towards the SAM MTase domain. The conformational flexibility of the enzyme may play a significant role in substrate binding and may be related to the availability of the active site, which is important for the binding of a substrate or the N-terminus. An other study also suggested that the interaction between Asn475 and Arg662 residues may play a significant role in the stabilization of nsP2pro.

We investigated the distance between the center of mass of these residues and found that there was a non-significant but systematic difference between the two conformers in the case of the wild-type Asn475 residue-containing structures (8DUF and 2HWK). A slightly higher distance was observed for self-inactivated conformation. However, the value obtained for the active conformer of the N475A mutant enzyme (6BCM$^{\text{MD-active}}$) was higher, it was even comparable with the values calculated for the 8DUF$^{\text{MD-self-inactivated}}$ structure, and showed remarkably larger fluctuations. The N475A mutation did not change the conformation of Arg662, but caused weakened interaction between the two domains, which was revealed only by the simulations rather than by the crystal structure. Our results imply the importance of the interaction between Asn475 and Arg662 residues in the stabilization of the structure.

4.3.7. Stability of the catalytic dyad

In general, the distance between the catalytic residues is an enzyme-specific value that is characteristic for the stability of the active site. Accordingly, we investigated the distance between Cys477 and His546 residues. For 8DUF and 2HWK structures, this distance was found

to be close to 5 Å, with only minor fluctuations. However, the degree of fluctuation was higher for the active conformer of the N475A mutant (6BCM$^{MD\text{-}active}$), the distance was even 10 Å in some simulations, indicating a remarkably lower conformational stability of the active site. This may be of significant importance for the activity of the enzyme: an enzyme having a destabilised active site may be unable to hydrolyse peptide bonds as efficiently as the wild-type.

The loss of the N475 residue's interactions upon the N475A mutation may cause such conformational changes at the interface of the SAM MTase and protease domains which can potentially interfere with substrate binding, as it was implied by the altered interactions between Asn545 and Val476 residues as described above. An increased distance between the SAM MTase and protease domains might facilitate the binding of either a substrate or the N-terminus to the active site. However, in the N475A mutant, we have observed pronounced fluctuations of the distance between the catalytic residues, indicating a higer probability for the binding of the N-terminus (since cleavage does not occur). The N475A mutant may have a higher probability to form the self-inactivated conformation, and an unfavourable structural change in the active site may have a negative impact on nsP2pro's activity. In contrast to this, the molecular dynamics analyses suggested that the self-inactivated conformer is less stable in the enzyme that contains the wild-type active site residue (Asn475) but modified surface residues K741A and K767A mutations).

4.3.8. Effects of the K741A és K767A surface mutations

The structural coordinates of VEEV nsP2pro (PDB ID: 8DUF) were determined based on an enzyme containing K741A and K767A surface mutations which were introduced in order to reduce surface entropy. These residues of VEEV nsP2pro correspond to the R755 and R781 residues of the SINV nsP2 SAM MTase domain. The R781 residue is located on the surface of the SAM MTase domain of SINV, at the interface of the nsP2 protein and the 'macro' and zinc-binding domains of nsP3. The R755 residue may be involved in interactions between nsP2 and nsP3. The R718A mutation of SINV nsP2pro has previously been shown to cause no change in viral viability. It is unknown whether the replication cycle of VEEV might be affected by these surface mutations because no literature data are available and none of the previously studied enzymes contain these mutations. Furthermore, the work presented in this PhD thesis was aimed to study the structure of the recombinant nsP2pro protein *in vitro* and *in silico*, thus, our studies did not include the investigation of the life cycle in cellular systems. But, it is important to note

that the similarities between the VEEV and SINV proteins may imply the possible effects of K741A and K767A mutations. The residues in equivalent structural positions may have highly similar roles in the protein function. Consequently, the positively charged Lys741 and Lys767 side-chains of VEEV may also mediate interactions between nsP2 and nsP3, similar to SINV nsP2pro. Our studies did not provide information about these interactions owing to the fact that the protein we studied represented only the enzymatic domain of nsP2pro but not the full-length nsP2 polyprotein or the nsP23 precursor protein. *In vitro* activity assays may reveal how the K741A/K767A mutations interfere with enzymatic properties.

Our structural analyses did not provide direct evidence for the contribution of the K741A and K767A mutations to the formation of the self-inactivated conformation of VEEV nsP2pro. These mutations have not been introduced previously by any of the studied VEEV nsP2pro proteins. In addition, in this work the mutations were designed in order to improve structure determination rather than to study their effect on viral replication. The structure of parts of the surface loop (Arg769-Tyr774 residues) located in the proximity of the modified positions was not visible in the electron density map, which indicated that the mutations increase the flexibility of this loop. The examination of crystal contacts have revealed that the side chain of Lys767 contributes to the stabilization of the loop in the crystal structure through a salt bridge. Although, the Lys741 and Lys767 residues are located far from the active site and are not involved in ligand binding, it can be hypothesized that their modification - due to the alteration of certain allosteric interactions - may still contribute to the appearance of the self-inactivated form or to its stabilization. This hypothesis is supported by the observation that the appearance of the self-inactivated conformer was not previously observed for the wild-type enzyme which contained no mutations either at the active site or at the enzyme surface (2HWK).

5. Summary

The goal of the research - presented in my doctoral thesis - was to study the Venezuelan equine encephalitis virus (VEEV) non-structural protein 2 protease (nsP2pro) by using computational approaches, especially molecular dynamics methods. We aimed to investigate the structural features that determine the specificity and stability of the enzyme and have correlation with the formation of the self-inactivated conformation.

To investigate the substrate specificity, I prepared a model of the nsP2pro in complex with an oligopeptide substrate representing the SFV nsP1/nsP2 natural cleavage site sequence, this substrate was used in the *in vitro* specificity study, as well. The enzyme-substrate complex was investigated with molecular dynamics simulations, followed by the modification of the substrate in an equilibrated complex in various positions (form P5 to P2') *in silico.* We have determined how the modification of the substrate affects the enzyme-substrate interactions. The results of the computational analyses were used to interpret those of the experimental specificity assays and to identify the structural features that determine amino acid preferences.

We have determined the structure of VEEV nsP2pro using X-ray crystallography. We found that the N terminus of the enzyme - lacking any mutation at the active site but containing two in its surface - flipped into the active site. This so-called self-inactivated conformation closely resembled that of the previously described N475A active site-mutant enzyme. The structural requirements of the self-inactivated conformation were investigated by using molecular dynamics simulations, based on the structural coordinates determined in this work or published previously. We examined the hydrogen bond interaction network of the N-terminus and the changes in the distances between the most important residues. In addition, the similarities and differences of the interaction networks were also determined in the case of binding of the N terminus (self-inactivated conformer) or an oligopeptide substrate to the active site. The simulations revealed the interactions that are responsible for the formation of the self-inactivated conformation and also appear in case of the wild-type Asn475 residue.

Our results help better understanding of the structural determinants of the specificity and stability of VEEV nsP2Pro, and can potentially be used in the investigation of other homologous viral proteases as well as can aid the design of protease inhibitors.

6. Ackowledgement

I would like to express my gratitude to my supervisor, **Dr. János András Mótyán** for his invaluable assistance throughout my research and thesis work, and for his selfless support in the preparation of numerous grant proposals and presentations, which have very significantly contributed to my professional development.

I would like to express my thanks to professor **Dr. József Tőzsér** both for the professional support as well as for the opportunity to join the Laboratory of Retroviral Biochemistry, to participate in many research projects, and to present the results at multiple conferences.

I would like to thank **Dr. Beáta Bozóki** for the experimental investigation of VEEV nsP2pro's substrate specificity which was crucial for the completion of my doctoral thesis.

I express my thanks to professor **Dr. David S. Waugh** as well as to **Dr. George T. Lountos**, **Dr. Danielle Needle** and **Dr. Alexander Wlodawer** for the determination of the VEEV nsP2pro's crystal structure.

I would like to thank the members of the Laboratory of Retroviral Biochemistry, **Dr. Mohamed Mahdi**, **Dr. András Szabó**, **Dr. Mária Golda** and **Irene Wanjiru Kiarie** for the collaborative work and for the possibility to contribute to further amazing projects. I want to express my thanks the other members as well, for the help provided by **Noémi Miltner, Viktor Ambrus**, as well as **Ágota Nagyné Veres** and **Szilvia Janics-Pető**. I am especially grateful to **Balázs Tibor Kunkli** for the professional assistance, in-depth discussions and for the success of our successful grant.

I would like to express my gratitude to **Prof. Dr. Shina Caroline Lynn Kamerlin**,both for the EBSA bursary opportunity as well as for the help and advices she provided in the field of computational chemistry.

I would like to thank my previous mentors, **Dr. Attila Mándi** who has introduced me to the fundamentals of computational chemistry, and professor **Dr. Mónika Fuxreiter** for the profesional help. I would like to express my gratitude in memory of **Dr. István Komáromi**† who introduced me to a multitude of methods of protein modelling.

I want to thank professor **Dr. László Fésüs László**, **Dr. Dóra Karancsiné Dr. Menyhárd** and **Dr. Krisztina Fehér** for the professional advice provided during the complex exam, which significantly contributed to increasing the quality of my doctoral thesis.

I want to thank **Dr. Attila Fekete** for his advice in the field of computational chemistry and for his kind patience regarding the usage of the supercomputers. I would like to express my thanks for the many advices to **Dr. Márton Miskei.** I want to thank **Dr. Péter Béla Hajdu** for his advices in the field of physics. I owe my thanks to professor **Dr. Tibor Kurtán**, for the help he provided. I want to thank **Ilona Tornyi** for her advices on bioinformatics methods.

I want to thank **Dr. Boldizsár Tüű-Szabó**, **Norbert Duró** and **Dr. Renáta Kertiné Ferenczi** for the help during my earlier research works.

I express my thanks to **Lajos Tölli** for the assistance with computer devices.

I want to thank **Zsuzsanna Oláh**, **Orsolya Szabó**, **Anett Nagy** and **Ildikó Vida** for the help with administrative tasks.

I would like to express my gratitude to my family. I am especially grateful for my **mother** for her kind support and patience during my studies. I would like to thank my **aunt** for her help. I want to express my thanks to my **sister** and **brother**.

With sincere gratitude, I remember my **father** who always supported me in my studies and encouraged me to face any difficulties that arose. I will always cherish his memory.

The work was supported by the following projects:
Data Intensive and Open Science program (to Gyula Hoffka), Thematic Excellence Programme (TKP2021-EGA-20), MTA Post-covid Programme (POST-COVID$_{2021\text{-}16}$), „Idea to Life" Scientific Idea Development Programme (EFOP-3.6.1-16-2016-00022) and EBSA Bursary (to Gyula Hoffka). Ministry of Culture and Innovation ÚNKP-23-5-DE-486 (to András János Mótyán) and ÚNKP-23-3-II-DE-456 (to Hoffka Gyula) New National Excellence Programme supported by the National Research, Development and Innovation Office. András János Mótyán was supported by the Bolyai János Scholarship (BO/00110/23/5) of the Hungarian Academy of Sciences.

We acknowledge KIFÜ for awarding us access to resource based in Hungary.

7. Appendix

Conference talks

1. **Hoffka G**, Komáromi I. A mikrobiális transzglutamináz reakció vizsgálata hibrid QM/MM (ONIOM) módszerekkel.
   KeMoMo-QSAR 2017 Symposium, Szeged, 2017.06.01-02.

2. **Hoffka G**, Fuxreiter M. Multiscale simulations on enzymatic catalysis.
   11th Molecular, Cell and Immune Biology Winter Symposium, Debrecen, 2018.01.04-05.

3. **Hoffka G**, Mándi A, Herczeg M, Borbás A, Fuxreiter M, Komáromi I. Hidrogén-szulfit addíció in silico vizsgálata.
   KeMoMo-QSAR 2018 Symposium, Szeged, 2018.05.24-25.

4. **Hoffka G**, Fuxreiter M. Conformational heterogeneity propels enzyme evolution.
   12th Molecular, Cell and Immune Biology Winter Symposium, Debrecen, 2019.01.10-11.

5. **Hoffka G**, Fuxreiter M. A konformációs heterogenitás szerepe az enzimek evolúciójában.
   KeMoMo-QSAR 2019 Symposium, Szeged, 2019.06.06-07.

6. Fuxreiter M, **Hoffka G**, Mones L. The role of conformational heterogeneity in the evolution of enzymes.
   27th National Meeting of the Hungarian Biophysical Society, Debrecen, 2019.08.26-29.

7. **Hoffka G**, Fuxreiter M. Role of conformational heterogeneity in enzymatic catalysis
   13th Molecular, Cell and Immune Biology Winter Symposium, Debrecen, 2020.01.09-10.

8. **Hoffka G**, Fuxreiter M. Multiscale simulations on the role of conformational heterogeneity.
   1st Molecular, Cell and Immune Biology Summer Symposium, Debrecen, 2021.05.14.

9. **Hoffka G**, Fuxreiter M. Theoretical studies on the underlying mechanism of enzyme evolution.
   15th Molecular, Cell and Immune Biology Winter Symposium, Debrecen, 2022.01.09-10.

10. **Hoffka G**, Mótyán JA, Tőzsér J. Computational studies on the self-inactivated conformation of the wild-type Venezuelean equine encephalitis virus (VEEV) protease.
    2nd Molecular, Cell and Immune Biology Summer Symposium, Debrecen, 2022.06.03.

11. **Hoffka G**, Lountos GT, Tőzsér J, Mótyán JA. Venezuelai ló-láz encephalitis vírus proteáz öninaktivált konformerének vizsgálata molekuladinamikai szimulációkkal.
Bioinformatics 2022 conference, Budapest, 2022.11.11.

12. **Hoffka G**, Mahdi M, Tőzsér J, Mótyán JA. Structural analysis of the binding of nirmatrelvir to the SARS-CoV-2 main protease.
Bioinformatics and Data Science in Genomic Studies, Debrecen, 2022.11.24-25.

13. **Hoffka G**, Mahdi M, Tőzsér J, Mótyán JA. Combined computational study of the binding of nirmatrelvir to SARS-CoV-2 main protease: Insight into resistance mechanism.
EBSA-2023 Congress, Stockholm, 2023.07.21.-08.04.

14. **Hoffka G**, Mahdi M, Tőzsér J, Mótyán JA. Multiscale computational study on the binding of nirmatrelvir to SARS-CoV-2 main protease: possible pathways to resistance.
29th National Meeting of the Hungarian Biophysical Society, Budapest, 2023.08.28-21.

15. **Hoffka G**, Kamerlin SCL. The role of conformational dynamics in the evolution of designed Kemp eliminase.
FEBS 2023 Advanced Course: Computational Approaches to Understanding and Engineering Enzyme Catalysis, Zagreb, 2023.09.25-29.

16. **Hoffka G**, Mahdi M, Tőzsér J, Mótyán JA: Multiscale computational study on the protonation states of the SARS-CoV-2 main protease catalytic dyad.
17th Molecular, Cell and Immune Biology Winter Symposium, Debrecen, 2024.01.15-16.

17. Kairie I, Linkner T, Szojka Z, **Hoffka G**, Kunkli B, Tőzsér J, Mahdi M. Efficacy of integrase strand transfer inhibitors & the capsid inhibitor lenacapavir against HIV-2 and the cellular effects of HIV-2's viral protein X.
17th Molecular, Cell and Immune Biology Winter Symposium, Debrecen, 2024.01.15-16.

Poster presentations

1. **Hoffka G**, Lountos GT, Tőzsér J, Mótyán JA. Crystallographic and molecular dynamics simulations shed light on the selfinactivated conformation of the Venezuelan equine encephalitis virus (VEEV) protease
29th National Meeting of the Hungarian Biophysical Society, Budapest, 2023.08.28-21.

UNIVERSITY AND NATIONAL LIBRARY
UNIVERSITY OF DEBRECEN
H-4002 Egyetem tér 1, Debrecen
Phone: +3652/410-443, email: publikaciok@lib.unideb.hu

Registry number: DEENK/302/2024.PL
Subject: PhD Publication List

Candidate: Gyula Hoffka
Doctoral School: Doctoral School of Molecular Cellular and Immune Biology
MTMT ID: 10080018

List of publications related to the dissertation

1. **Hoffka, G.**, Lountos, G. T., Needle, D., Wlodawer, A., Waugh, D. S., Tőzsér, J., Mótyán, J. A.: Self-inhibited state of Venezuelan equine encephalitis virus (VEEV) nsP2 cysteine protease: a crystallographic and molecular dynamics analysis.
J. Mol. Biol. 435 (6), 1-20, 2023.
DOI: http://dx.doi.org/10.1016/j.jmb.2023.168012
IF: 5.6 (2022)

2. Bozóki, B., Mótyán, J. A., **Hoffka, G.**, Waugh, D. S., Tőzsér, J.: Specificity Studies of the Venezuelan Equine Encephalitis Virus Non-Structural Protein 2 Protease Using Recombinant Fluorescent Substrates.
Int. J. Mol. Sci. 21 (20), 1-26, 2020.
DOI: http://dx.doi.org/10.3390/ijms21207686
IF: 5.924

List of other publications

3. Golda, M., **Hoffka, G.**, Cherry, S., Tropea, J. E., Lountos, G. T., Waugh, D. S., Wlodawer, A., Tőzsér, J., Mótyán, J. A.: P1' specificity of the S219V/R203G mutant tobacco etch virus protease.
Proteins. [Epub ahead of print], 2024.
DOI: http://dx.doi.org/10.1002/prot.26693
IF: 2.9 (2022)

4. Mótyán, J. A., Mahdi, M., **Hoffka, G.**, Tőzsér, J.: Potential Resistance of SARS-CoV-2 Main Protease (Mpro) against Protease Inhibitors: lessons Learned from HIV-1 Protease.
Int. J. Mol. Sci. 23 (7), 1-20, 2022.
DOI: http://dx.doi.org/10.3390/ijms23073507
IF: 5.6

UNIVERSITY AND NATIONAL LIBRARY
UNIVERSITY OF DEBRECEN
H-4002 Egyetem tér 1, Debrecen
Phone: +3652/410-443, email: publikaciok@lib.unideb.hu

5. Ambrus, V. A., **Hoffka, G.**, Fuxreiter, M.: Asymmetric dynamic coupling promotes alternative evolutionary pathways in an enzyme dimer.
Sci. Rep. 10 (1), 1-9, 2020.
DOI: http://dx.doi.org/10.1038/s41598-020-75772-5
IF: 4.379

6. Tüű-Szabó, B., **Hoffka, G.**, Duró, N., Fuxreiter, M.: Altered dynamics may drift pathological fibrillization in membraneless organelles.
Biochim. Biophys. Acta. Proteins Proteom. 1867 (10), 988-998, 2019.
DOI: http://dx.doi.org/10.1016/j.bbapap.2019.04.005
IF: 2.371

7. Ferenczi, R., Illyés, T. Z., Király, S. B., **Hoffka, G.**, Szilágyi, L., Mándi, A., Antus, S., Kurtán, T.: Evaluation of Different Synthetic Routes to (2R,3R)-3-Hydroxymethyl-2-(4-hydroxy- 3-methoxyphenyl)-1,4-Benzodioxane-6-Carbaldehyde.
Curr. Org. Chem. 23 (26), 2960-2968, 2019.
DOI: http://dx.doi.org/10.2174/1385272823666191212113407
IF: 1.933

Total IF of journals (all publications): 28,707
Total IF of journals (publications related to the dissertation): 11,524

The Candidate's publication data submitted to the iDEa Tudóstér have been validated by DEENK on the basis of the Journal Citation Report (Impact Factor) database.

24 May, 2024

DEBRECENI EGYETEM
Egyetemi és Nemzeti Könyvtár