THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY (PHD)

Characterisation of the Ty1 protease and the human trypsins' inhibitory profile

Lívia Diána Gazda

Supervisor: Dr. József Tőzsér

UNIVERSITY OF DEBRECEN

DOCTORAL SCHOOL of MOLECULAR CELL AND IMMUNE BIOLOGY

2023 Debrecen

Characterisation of the Ty1 protease and the human trypsins' inhibitory profile

By Lívia Diána Gazda, molecular biologist (MSc)

Supervisor: Prof. Dr. József Tőzsér

Doctoral School of Molecular Cell and Immunobiology

Members of the Examination Committee:

head:	Prof. Dr. István Balogh, PhD, DSc
members:	Dr. Attila Ambrus, PhD, DSc
	Dr. Eszter Csoma, PhD

The Examination takes place at : Faculty of Medicine, University of Debrecen, Theoretical Building, 5. Seminar room, 15th June 2023

Defense Committee:

Head:	Dr. István Balogh, PhD, DSc
Reviewers:	Dr. Eszter Boros, PhD
	Dr. Tibor Docsa, PhD
	Dr. Attila Ambrus, PhD, DSc
	Dr. Eszter Csoma, PhD

The PhD Defense takes place at the the Lecture Hall of Bldg. A, Department of Internal Medicine, Faculty of Medicine, University of Debrecen, 15th June 2023

1. INTRODUCTION

1.1 About proteases in general

Proteases are biocatalysts, that can be found in every living organism. They facilitate the cleavage of peptide bonds. They are vital for life, and involved in many biochemical pathways such as digestion, blot clotting, signal transduction and regulation of the immune system. Their abnormal function leads to pathological cases, for example tumour, inflammatory or neuro-degenerative diseases. They also play an important role in the life cycle of viruses and other microorganisms.

The most characterised proteases participate in digestion, such as the stomach secreted aspartic protease, pepsin, or the trypsin that is produced by the pancreas, and belong tot he group of serine proteases. The HIV protease is one of the well characterised viral proteases, that is also a therapeutic target of the retroviral therapy, therefore the number of studies of retroviral and retroviral-like proteases have been rising in the past decade.

1.2 Aspartic and serine proteases

Many aspartic proteases have been characterised in the past decades, such as pepsin, cathepsin D, cathepsin E and renin. The *International Union of Biochemistry and Molecular Biology*, IUBMB classifies aspartic proteases as hydrolyses, based on the type of catalysed reaction, within this group they are classified as endopeptidases where the active site contains aspartate (EC 3.4.23). These enzymes usually contain the following motif around their catalytic aspartate residue: Xaa-Xaa-Asp-Xbb-Gly-Xbb where Xaa is hydrophobic and Xbb can be serine or threonine. The group of aspartic proteases also comprises clinically important therapeutic target proteases, like the Human Immunodeficiency Virus 1 (HIV-1) protease (PR) or the Ty1 retrotransposon protease from *Saccharomyces cerevisiae.*

The group of serine proteases comprise many digestive and viral enzymes, for example trypsin and chymotrypsin. Based on the amino acid sequence and tertiary structure we can distinguish approximately 14 clans and 50 families within the group of serine proteases. Most of the proteases within the S1 family are synthesised in a precursor from that contains an N-terminal signal that is removed during limited proteolysis. The enzyme will be activated after the cleavage of the signal peptide. The main difference between the serine, threonine and cysteine proteases compared to the aspartic, glutamic and metallo proteases, that instead of a water molecule, an amino acid residue in the active centre acts as the nucleophile. Therefore, the acyl intermediates are only formed in the reactions that are catalysed by serine, threonine, and cysteine proteases. In serine proteases the OH group of a serine residue will act as nucleophile and attack the peptide bond in the substrate.

1.3 Inhibitors of aspartic and serine proteases

Pepstatin A is a universal aspartic protease inhibitor. It consists of 6 amino acids (Isovaleryl-Val-Val-Sta-Ala-Sta). Acetyl pepstatin A is derived from pepstatin A (Ac-Val-Val-Sta-Ala-Sta-OH; Ac: acetyl group), that binds to the active centre of the proteases. The OH group of the statine in the middle of the inhibitor forms a tetrahedral intermediate with the aspartates of the active centre, therefore during the hydrolysis the water molecule will be crowded out. Acetyl pepstatin A and pepstatin A are chemically almost identical, yet the acetyl pepstatin A is a more potent inhibitor of retroviral proteases than the pepstatin A in in vitro experiments.

Serine proteases and their inhibitors are one the most intensively studied protein complexes. Based on the mechanism of action of serine protease inhibitors we distinguish 3 groups: canonical, non-canonical and serpine-type inhibitors. Canonical inhibitors bind through a convex binding loop that is complementary to the active site. The mechanism of inhibition is similar to substrate binding. Serine protease inhibitor Kazal type 1 (SPINK1) belongs to the group of canonical inhibitors. It is expressed in the acinus cells of the pancreas and has a function of inhibition of premature trypsin activation in the organ that leads to self-digestion.

2. OBJECTIVES

While the protocol of *S. cerevisiae* Ty1 retrotransposon protease purification has been published for a while, the protease has not been characterised so far. Our research group aim was the complete biochemical characterisation of Ty1 protease. Following purification we aimed to determine the optimal conditions of the activity, the specificity of the protease, examine the inhibitory profile and perform in silico structural studies. We aimed to compare the properties of the protease to other retroviral and retroviral-like proteases to analyse the similarities and differences between them.

Chronic pancreatitis can be caused by the mutation of the PRSS, CTRC, SPINK1, or CPA1 genes. The mutations can lead to the disease via two different pathological pathways. One means the elevated activity of trypsin within the pancreas, the other one comprises the loss of function derived from the misfolding of the inhibitor. The trypsin inhibitor SPINK1 prevents the premature activation of trypsin, therefore mutations that affect SPINK1 might interfere with this pathway. Many mutations were identified by research groups in patients with pancreatitis and were suggested to play a role in the development of pancreatitis. However the mechanism behind this was not studied, therefore our group aimed to study the effect of sulfation of human anionic and cationic trypsins on the interaction with SPINK1. We also aimed to study the effect of mutations of SPINK1 on the binding and formation of inhibitor-enzyme complex.

3. MATERIALS AND METHODS

3.1 The analysis of Ty1 retrotransposon protease

3.1.1 Cloning, expression and purification of Ty1 protease

The pET11a vector coding for Ty1 Gag-PR-His_6 was a kind gift of Dr. J.F. Lawler research team (The Johns Hopkins University School of Medicine, Baltimore, Maryland, USA). The Ty1 protease coding sequence was cloned into pET11a vector by PCR. The full Gag-protease and native protease (PR) was expressed in BL21(DE3) *Escherichia coli* derived strain. The Ty1 Gag-PR-His_6 form was purified from the soluble fraction with Ni-chelate affinity chromatography, whilst the Ty1 PR- form was purified from the insoluble fraction with gel filtration. The purity of the proteins were analysed by 14% SDS-PAGE. The fractions containing the protease were dialysed against "yeast *in vivo*-like" buffer and were concentrated with 10K and 3K Amicon Ultra 0.5 ml centrifugal filter units. The protein concentration was determined by Bradford assay.

3.1.2 Ty1 identification by Western blot

The proteins were separated on 14% polyacrylamide gel and transferred to nitrocellulose membrane. The protease was detected with mouse anti-His primary antibody and HRP conjugated anti-mouse secondary antibody. The signal was developed with chemiluminescent substrate.

3.1.3 Activity assays with oligonucleotides

The oligopeptide substrates were dissolved in water and were incubated at a final concentration of 0.2–1.2 mM with Ty1 PR-His6 or Ty1PR (400–1600 nM) for 2-5 hours at 30°C. The cleavage products were separated by HPLC. The kinetic measurements were performed on Ty1 PR/IN cleavage site representing oligopeptide substrate (VPTIN*NVHTS). The kinetic parameters were calculated at less than 20% hydrolysis in the linear range.

3.1.4 Analysis of the effect of pH, ionic strength, temperature and urea concentration on the Ty1 activity

The impact of different reaction conditions on the Ty1 protease activity was studied with PR/IN cleavage site representing oligopeptide substrate (VPTIN*NVHTS). The optimal pH range of the protease was determined with Ty1 PR-His_6. The reactions were within the range of pH 6.5-9.0 and were incubated for 4 hours at 30°C. The effect of ionic strength on the activity was determined in the range of 0.5-2 M NaCl concentration. The optimal temperature was measured in the range of 18-37°C. The urea dissociation constant (UC_{50}) was determined in the range of 0.05-0.25 M urea concentration.

3.1.5 Inhibition of Ty1 protease

The activity of the protease was measured in the presence of HIV-1 inhibitors: atazanavir, nelfinavir, saquinavir, darunavir, amprenavir, lopinavir and tipranavir. The effect of universal retroviral protease inhibitors was determined with acetyl pepstatin and pepstatin on the proteolytic activity. The half maximal inhibitory (IC_{50}) concentration of acetyl-pepstatin was determined in the range of 100-1000 nM.

3.1.6 Expression and purification of fluorescent substrates

For the expression of recombinant fluorescent substrates we used a modified pDest-His_6-MBP-mTurquoise2 expression vector. The vector was cut with restriction enzymes, and the flexible region $(GGGGS)_4$ was cloned into the vector with two oligonucleotides and ligated with T4 DNA ligase. The constructs were transformed in TOP10 *E. coli* bacterial strain. The plasmids were purified with High-Speed Plasmid Mini Kit. The constructs were confirmed by sequencing. The sequences coding for the Ty1 protease cleavage sites were cloned into the linearised pDest-His_6-MBP-$(GGGGS)_4$-mTurquoise2 vector with BamHI and PacI restriction endonucleases. The linearised vector was identified in 1% agarose gel and purified with NucleoSpin Gel and PCR Clean-up kit. Finally 150 ng linearised plasmid was incubated with 200 ng oligonucleotide primer and ligated with T4 DNA ligase.

3.1.7 Expression and purification of fluorescent substrates

The fluorescent substrates were expressed in BL21(DE3) *E. coli* bacterial strain. The His_6-tagged recombinant fluorescent substrates were purified from the soluble fraction. The supernatants were incubated for 30 mins at RT with Ni-NTA magnetic beads. To separate the supernatant and beads we used Dynamag™-2 magnetic particle concentrator. Finally, the supernatant was removed and the beads were washed with washing buffer and resuspended in the reaction buffer.

3.1.8 Gel electrophoretic analysis of fluorescent substrates

The substrates were eluted from the beads and the buffer was exchanged to water with 10K Amicon Ultra 0.5 mL filter units. The reactions contained buffer, recombinant fluorescent substrate (1–3 mg/mL) and Ty1 protease (300–1200 nM). The control reactions contained storage buffer instead of the protease. The reactions were incubated at 30°C for 16 hours and stopped with Laemmli sample buffer. Before SDS PAGE the samples were denatured at 95°C. The substrates and cleavage products were separated by SDS PAGE. The SDS was washed out from the gel with water to allow the substrates to renature in the gel. After renaturation the band were visualized by blue light transillumination and the proteins were stained with Coomassie stain.

3.1.9 Kinetic measurements on recombinant fluorescent substrates

The kinetic measurements were performed using Ni-NTA magnetic beads. Different amounts of beads coated with the immobilised substrate were aliquoted in Eppendorf tubes. To separate the beads from the mobile phase a particle concentrator was used and the beads were

redissolved in the same volume of reaction buffer. The substrate concentration was determined by Bradford assay. The utilised maximum substrate concentration in the reactions was 0.08 mM and the protease was added in μM final concentrations. The reactions were incubated for 2 hours at 30°C and then stopped by the separation of the mobile phase and the beads with the immobilised substrate. The fluorescent intensity (RFU) of the liquid, mobile phase was measured with multimode microtiter plate reader with 400/10 nm excitation and 460/40 nm emission filters. The values were blanked, divided by the slope of the reference curve and plotted against the substrate concentration (μM). Kinetic parameters were determined at less than 20% substrate hydrolysis by Michaelis-Menten non-linear regression analysis using GraphPad Prism.

3.1.10 Homolog modelling of Ty1 protease

The PredictProtein server was used for the secondary structure prediction based on the sequence data of Ty1 protease (UniProtKB: Q07793). IUPred web server was used for disorder prediction. The homology modelling was performed by Modeller 9v13 software based on the xenotropic murine leukaemia virus-related virus (XMRV) (PDBID: 4EXH) and DNA damage-inducible protein 1 (Ddi1) (PDBID: 2I1A) protease sequences. Molecular visualisation was performed with PyMOL Molecular Graphics System software. The in silico analyses were performed by Dr. János András Mótyán.

3.1.11 Sample preparation and Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS)

The substrate coated Ni-NTA magnetic beads were incubated with Ty1 protease for 16 hours at 30˚C. After incubation the cleavage products were eluted with imidazole containing elution buffer. The eluted fractions were concentrated, and buffer exchanged to 50 mM Tris pH8.0 with 10K Amicon Ultra 0.5 ml centrifugal filter units. After buffer exchange the substrates were cleaved by *Tobacco Etch Virus*, TEV protease for 16 hours at 30˚C. The molecular mass of the final cleavage products was measured with MALDI-TOF MS.

The measurements were performed by Dr. Tibor Nagy (Department of Applied Chemistry) and carried out by a Bruker Autoflex Speed mass spectrometer using 2,5-dihydroxybenzoic acid (DHB) matrix in 50% aqueous acetonitrile with 0.1% TFA (1:1). The sample concentration was 100 mg/ml. 1 μl matrix and 1 μl sample was deposited on the plate.

3.2 Studies on SPINK1

3.2.1 Design of expression plasmids

For cell expression studies, the sequences coding for the wild type and mutant SPINK1 variants native and His10-tagged versions (SPINK1-His_{10}) were cloned into pcDNA3.1(-) with XhoI and BamHI restriction endonucleases by Dr. Andras Szabo. For protein purification purposes the wild type and mutant SPINK1 variants native and His10-tagged versions were previously cloned into SPINK1-minigene-1 construct. For expression in human cells and bacteria the human cationic trypsinogen, human anionic trypsinogen and tyrosylprotein sulfotransferase 2 previously were cloned in pTrapT7 and pcDNA3.1(-) plasmids.

3.2.2 Cell cultures and transfection

The human embryonic kidney (HEK293T) cells were transfected with pcDNA3.1(-) plasmid using Lipofectamine reagent and Opti-MEM medium. The cells were incubated for 15 hours at 5% CO_2 at 37°C. The supernatant was removed, and the cells were washed with PBS (pH 7.4), finally Opti-MEM medium was added to the cultures. The supernatant was removed after 48 hours and the SPINK1 protein was analysed on 15% SDS-PAGE and identified by Western-blot.

3.2.3 Western-Blot

Before electrophoresis the samples were boiled with 2x Laemmli sample buffer and 1 M DTT at 95°C for 15 minutes. The SPINK1-His_{10} protein was detected with Western-blot using PVDF membrane. After blocking with 5% marvel solution the membrane was incubated in 1:2000 horseradish peroxidase conjugated anti-His antibody for 1 hour at RT. The HRP activity was detected with chemiluminescent substrate. The native SPINK1 protein was also identified by Western using PVDF membrane and mouse monoclonal anti-SPINK1 antibody. For secondary antibody we used HRP-conjugated anti-mouse IgG antibody. The HRP activity was detected with chemiluminescent substrate.

3.2.4 The expression and purification of SPINK1 protein

The HEK293T cells were transfected with pcDNA3.1(-) SPINK1-His_{10} construct with polyethyleneimine (PEI). The cells were incubated for 15 hours in a CO_2 incubator, were washed with Opti-MEM. Finally, 20 mL Opti-MEM was added to the cells and were incubated for 48 hours. After incubation the supernatant was collected, and 20 ml Opti-MEM was added to the cells again. The SPINK1-His_{10} protein was purified with Ni affinity chromatography. The protein containing fractions were dialysed against or buffer exchanged with HiTrap desalting column 20 mM Tris-HCl pH8.0.

3.2.5 Expression, purification and activation of the trypsinogens

The non-sulphated recombinant cationic trypsinogen (Hu1) and anionic trypsinogen (Hu2) was expressed in *Escherichia coli* BL21(DE3) bacterial strain. The cells were disrupted with sonication and the soluble fraction was separated by centrifugation from the insoluble fraction. 1 mM L-cystine and 1 mM L-cysteine were added to the soluble fraction and were incubated overnight to help the folding of the trypsin isoforms. After incubation the trypsinogens were purified with ecotin affinity chromatography, The sulphated native trypsinogens were purified from human pancreatic juice with MonoQ anion exchanger column, followed with ecotin affinity chromatography. The non-sulphated and sulphated trypsinogens were expressed in HEK293T cells. The cells were transfected with 10 µg pcDNA3.1(-)-Hu1 or pcDNA3.1(-)-Hu2 plasmid and were grown in the presence of 50 mM sodium chlorate, in order to inhibit the endogen ATP-sulfurylase. The cells were cotransfected with 8 µg pcDNA3.1(-)-Hu1 with pcDNA3.1(-)-TPST2 or pcDNA3.1(-)-Hu2 with pcDNA3.1(-)-TPST2 plasmids, the latter coding for tyrosylprotein sulfotransferase 2 to promote the trypsinogen sulphation. The trypsinogens were purified from 200-400 ml supernatant with ecotin affinity chromatography. The purified inactive zymogen forms were activated with enterokinase and were titrated against ecotin.

3.2.6 Concentration determination assays

The HEK293T cell-expressed SPINK1-His10 protein concentration was determined by titration against trypsin. Half dilution series were prepared on 96-well plates with trypsin and reaction buffer. Every well contained 40 nM trypsin. The reactions were incubated for 30 minutes at RT. After incubation Z-Gly-Pro-Arg-p-nitroanilide substrate was added to the wells to start the reactions. The trypsin activity was plotted against the concentration the SPINK1 concentration was determined from the interception of the x-axis.

The bovine trypsin was titrated against p-nitrophenyl p-guanidinobenzoate. The trypsin inhibitor ecotin was expressed in *E. coli* BL21(DE3) periplasm and purified with trypsin affinity chromatography. The ecotin concentration was determined by titration against bovine trypsin and were used to determine the purified human trypsin isoforms concentration. The reactions contained 25-50 nM trypsin and exceeded by a minimum of 2 orders of magnitude the dissociation constant of ecotin. The concentration of the SPINK1 was also determined with active centre titration with 25-50 nM final trypsin concentration in the reactions. In the case of K41N SPINK1 the inhibitor binding was weak, therefore the concentration was determined with SDS-PAGE and densitometry using wild type SPINK1 as standard.

3.2.7 Assays to determine the equilibrium constant of the SPINK1 variants

The binding affinity of the wild type and mutant SPINK1 proteins to the trypsin isoforms were determined with the measurement of the dissociation constant (K_D) at equilibrium. A 96 well plate with fixed trypsin concentration containing wells were incubated with increasing concentrations of SPINK1 with a final volume of 0.2 ml and were incubated for 15 hours. The only exception was the K41N SPINK1 mutant that was incubated for 1 hour. The free trypsin

concentration was determined in the equilibrium by the addition of Z-Gly-Pro-Arg-AMC fluorescent substrate. The fluorescent intensity was measured in 96-well plate, at 380 nm excitatory and 460 nm emission wavelength. The free trypsin concentration was plotted against the total inhibitor concentration to determine the K_D values.

3.2.8 Association assays

The reactions contained 50 pM trypsin and 500 pM SPINK1 and were incubated at RT. The trypsin activity was measured with the addition of Z-Gly-Pro-Arg-AMC fluorogen substrate on 96 well plate, at 380 nm excitatory and 460 nm emission wavelength. The pseudo–first-order rate constant (kobs) was calculated from the slope of linear fits of semilogarithmic plots of ln(vt/v0) versus time of inhibition, using the equation (–kobs)(time) = ln(vt/v0), where vt is the residual trypsin activity and v0 is the maximal uninhibited enzyme activity. The second-order association rate constant (k_{on}) was calculated from the ratio of the pseudo–first-order rate constant and inhibitor concentration.

3.2.9 Dissociation constant measurements

The protease-inhibitor complex was obtained with mixing 11 nM SPINK1 and 10 nM trypsin and was incubated for 1 hour at 23°C. The reactions were started with the addition of Z-Gly-Pro-Arg-AMC substrate on 96 well plate, at 380 nm excitatory and 460 nm emission wavelength. The parabolic curves were fitted with the second-order polynomial function, the c coefficients were plotted against the initial complex concentration (pM) and the slopes of linear fits were used to calculate the dissociation constants (k_{off}) (s_{-1}). The turnover number of trypsins on the substrate was calculated using the same assay as for the dissociation constant determinations. The turnover number was calculated as the initial rates of substrate hydrolysis (RFU/s) was plotted against the enzyme concentration (pM). The slope of the curve gave the turnover number (RFU/s/pM).

4. RESULTS

4.1 The analysis of Ty1 retrotransposon protease

4.1.1 Cloning, expression and purification of Ty1 protease

The native Ty1 protease was cloned and expressed in pET11a bacterial expression plasmid and confirmed by sequencing. The native and His-tagged version of Ty1PR was expressed in *E. coli* cells. The native form was purified with gel filtration. The Ty1 Gag-PR-His_6 protein was purified with Ni-chelate chromatography, where the Ty1 Gag-PR-His_6 precursor (MWCO: 72 kDa) went through autoproteolysis resulting in the Ty1 PR-His_6 elution from the column (MWCO: 21 kDa). The processed Ty1 PR-His_6 and the native protease was involved in the biochemical characterisation.

4.1.2 Activity assay with oligopeptide substrates

To help the active folding of the Ty1 protease, we have tested several dialysis buffers that have been previously used, optimised for retroviral proteases. The folded protease and water dissolved oligopeptide substrate was added to the reactions and were incubated at 30°C. The reactions were stopped by the addition of TFA and the cleavage products were separated by HPLC on reversed phased column. The activity was calculated from the conversion of the substrate.

We have only observed activity when the protease was dialysed against yeast "in vivo like" buffer and the reactions were set up in PIPES- ("A" peptide buffer) or MES based ("B" peptide buffer). The ratio of the buffers was crucial for the proteolytic activity as we only observed activity in ¼ water, ¼ yeast "in vivo like" buffer, ½ PIPES or MES based buffer. Besides determining the optimal conditions for the proteolytic activity we have characterised the biochemical properties of the Ty1 protease.

The optimal ionic strength for Ty1 protease activity was within the range of 1.5-2 M. Similarly to human immunodeficiency virus type 1 (HIV-1), human T-lymphotropic virus type 1, -2, -3 (HTLV-1, -2, -3) and human foamy virus (HFV) proteases the ionic strength increased the activity >1 M NaCl concentration. The Copia transposon protease of *D. melanogaster* also achieved maximum activity at 2 M NaCl final concentration in the reaction.

The optimal pH was slightly basic at pH 7.7, which is higher than the usual, optimal pH for retroviral and retroviral-like protease activity. For example, HFV has a pH optimum of 6.6–6.8, whilst for the HIV-1 protease this is within the range on pH 4-6. The human ASPRV1 protease optimal range is also in the slightly acidic range pH 6.0-6.5. However, the optimal pH for the protease of human *paternally expressed gene 10*, similarly to Ty1 protease, also falls within the slightly basic range pH 6.9-7.4. Interestingly Copia transposon protease of *D. melanogaster* achieves maximum activity at pH 4, similar to HIV-1 protease.

The optimal temperature for Ty1 protease activity was 30°C. This was observed in the case of other retroviral-like proteases, for tobacco vein mottling virus (TVMV) the optimal temperature of activity was observed at 34°C.

On the contrary, HFV and HIV-1 protease show the highest activity at 37˚C. In the case of Ty1 protease, the observed lower temperature optimum correlates with the previously published results of Lawler et al. They have described the lower transposition efficiency between 32–36°C, which could be explained by the Ty1 proteases heat sensitivity.

The Ty1 protease showed high sensitivity against the chaotropic agent urea. The urea dissociation constant (UC_{50}) was very low at 0.05 M. This value is 10-fold less than the UC_{50} of ASPRV1 (UC_{50}=0.54 M), and substantially lower than the UC_{50} of HIV-1 protease (UC_{50}=1.47 M), however it is similar to the XMRV protease UC_{50}=0.2 M.

The lower UC_{50} implies lower dimer stability, which will be discussed later in the *in silico* analysis. The retroviral (HIV-1, HTLV-1, BLV and MMLV) proteases usually show elevated activity on natural cleavage site analogue oligopeptide substrates, but in the case of the native and Ty1 PR-His_6 we have detected low activity on the oligopeptide substrates. Our research group has previously described low activity of HFV and Gag-encoded *Avian myeloblastosis* virus (AMV) protease. The Ty1 protease and Ty1 PR-His_6 both cleaved the VPTIN*NVHTS synthetic oligopeptide substrate representing the PR/IN cleavage site of the precursor polyprotein of the Ty1 retrotransposon. The kinetic measurements were performed using the same, PR/IN cleavage site representing oligopeptide. The catalytic constant (k_{cat}) was determined for the native and the His-tagged versions of Ty1 protease, showing no discrepancy, meaning that the hexahistidine tag does not alter the active folding.

The reported low specificity constant of HFV (0.007 $mM^{-1}s^{-1}$) was similarly low in the case of Ty1 protease too (0.028 $mM^{-1}s^{-1}$). We have also tested the other Ty1 retrotransposon cleavage sites, IN/RT (IHLIA*AVKAV) and Gag/PR (TARAH*NVSTS) representing synthetic oligopeptide substrate, but we could not observe activity.

4.1.3 Kinetic measurements on fluorescent protein substrates

For the kinetic assays we have used a previously reported recombinant fluorescent protein-based assay. The substrates contained an N-terminal His_6 tag to promote immobilisation, a maltose binding protein (MBP) to improve folding, and a C-terminal fluorescent protein (mTurquoise2) that enables the detection of fluorescent intensity. Furthermore the substrates contained one of the Ty1 retrotransposon cleavage sites and a control TEV protease cleavage site. The substrates also contained an insertion of a flexible linker $(GGGGS)_4$ sequence prior to the fluorescent protein tag to make the cleavage site more accessible. We have examined the activity of the Ty1 protease on 7 natural cleavage site representing fluorescent protein substrate in different buffer composition.

The uncleaved and cleaved fragments were separated on SDS-PAGE. Nonspecific cleavage sites were not observed at the endpoint nor during the incubation period. The recombinant protein substrates that were incubated with Ty1 protease we have only observed one fluorescent fragment meaning that the protease only cleaves at the engineered site and has no cryptic cleavage sites in the substrate. The samples were also analysed by MALDI-TOF mass spectrometry.

We determined the kinetic parameters of the Ty1protease by fluorometric assay. The catalytic constant for the substrates that contained 10 amino acid long Ty1 cleavage sites were all comparable and the same magnitude, the highest value was measured on the PR/IN cleavage

site representing substrate in reaction buffer (50 mM MES, 100 mM Tris, 50 mM Na-acetate, 150 mM NaCl, 75 mM KCl, 12.5 mM NaH_2PO_4, 61.25 mM Na-glutamate, 12.5 mM $MgSO_4$, 0.125 mM $CaCl_2$, 0.05% Tween20, pH 8.0). The catalytic constants were higher on the substrates that contained 20 amino acid long Ty1 cleavage sites instead of 10. However in reaction buffer "A" (10 mM PIPES, 75 mM NaCl, 0.25% Nonidet P-40, 5% glycerol, 75 mM KCl, 12.5 mM NaH_2PO_4, 61.25 mM Na-glutamate, 12.5 mM $MgSO_4$, 0.125 mM CaCl2, 0.05% Tween20, pH 7.0) the k_{cat}/K_m constants were lower on the 20 amino acid long Ty1 cleavage site containing substrates in the case of IN/RT and Gag/PR sites. The reason behind this discrepancy could be the differences between the composition of the buffers and/or in the case of the extended substrate sequences there might be alternative, non-productive binding events that can affect the measured parameters. The role of surface residues in substrate binding was described recently for HIV-1 and HTLV-1 proteases, this interaction surface was named as substrate binding groove or S-groove. The S-groove is able to interact with those residues of the polyprotein which connect the functional domains (e.g. the matrix (MA) and capsid (CA) proteins in HIV-1 polyprotein). While the active sites of HIV-1 and HTLV-1 protease bind the P5-P5' substrate residues, their S-grooves enable interactions with P12-P6 and P6'-P12' residues of the interdomain linkers, which provides extended substrate-binding surface for the enzymes. The overall binding energy was found to be higher in the case of interactions with P12-P12' residues as compared to binding of only P4-P4' residues at the active site.

When comparing the catalytic constants of the 10 and 20 amino acid substrates in the case of Ty1 protease we did not observe substantial difference, and it was insufficient to prove the presence of surface binding sites in the Ty1 protease. In order to further examine this, we have designed not only the 10 amino acid long (VPTIN*NVHTS) and 20 amino acid long (PSNISVPTIN*NVHTSESTRK) PR/IN cleavage site, but we have also designed a mutant version (GGGGGVPTIN*NVHTSGGGGG) where the amino acids in the P10-P6 and P6'-P10' positions were changed to glycine. We were interested in if changing the other cleavage site residues would disrupt the side chain-mediated enzyme-substrate interactions at these sites. The k_{cat}/K_m catalytic constants were comparable for the mutant and wild type 20 amino acid long PR/IN substrates. The substitutions to glycine caused only slight differences in the measured k_{cat}/K_m values, and these were not significant.

Despite the fact that the HIV-1 protease was found to involve surface amino acids in the substrate binding, our results did not prove that a substrate-groove was involved in the substrate binding in the case of Ty1 protease. The substitution of P10-P6 and P6'-P10' substrate residues caused no significant changes in the measured catalytic constants, even when measured in different buffers. We propose that because we did not observe change or abolishment of substrate binding, the surface binding sites of Ty1 protease - if they exist - may have only weak interactions with the P10-P6 and P6'-P10' substrate residues.

4.1.4 Analysis of cleavage positions in the recombinant fluorescent protein substrates

Our research team have found previously that the separation of cleavage products by SDS-PAGE may indicate the presence of alternative cleavage sites in the recombinant substrates. However the control cleavage site of TEV PR in a His6-MBP-mTurquoise2 fusion protein has not been used for the determination of cleavage position of the studied protease before we deployed it in the case of Ty1 protease.

Our aim was to confirm that the Ty1 protease has no alternative nonspecific cleavage sites within the recombinant fluorescent protein substrate. To prove this, we cleaved the substrates with Ty1 and TEV protease simultaneously. The small product fragments were analysed by MALDI-TOF mass spectroscopy for every substrate variant. The reactions were performed in reaction buffer "B". As the reaction buffer contained Tween20 detergent – that can have a negative effect on the mass spectroscopy – this which was removed with buffer exchange to 50 mM Tris (pH 8.0) buffer. The buffer exchange was required for the MALDI-TOF measurements, the Tween20 buffer c. For buffer exchange we used 10 kDa centrifugal filter units. The removal was successful, the polyethylene glycol-derivative did not impair or interfere with the detection of small proteolytic fragments.

After the digestion of the recombinant fluorescent protein substrates with Ty1 and TEV proteases the molecular masses of proteolytic fragments were determined by MALDI-TOF mass spectroscopy. The calculated m/z values ([M+H]+) were compared to the measured values which corresponded well. The SDS-PAGE and MALDI-TOF MS results were in agreement, proving that the Ty1 protease cleaves only at one specific site in the recombinant fluorescent substrates and there are no alternative cleavage sites anywhere else in the maltose binding protein or $(GGGGS)_4$ flexible linker.

4.1.5 Inhibition studies of Ty1 protease

The potential inhibitors of Ty1 protease were tested. The inhibitors applied in antiretroviral therapy can be found in the database of the *Food and Drug Administration* (FDA). We have studied the impact on the proteolytic activity of the following inhibitors: atazanavir, nelfinavir, saquinavir, darunavir, amprenavir, lopinavir and tipranavir. We have also tested a potent inhibitor of HIV-1 PR and two universal aspartic protease inhibitors, acetyl-pepstatin and pepstatin A.

None of the inhibitors showed inhibitory potential at >100 nM final concentration, except acetyl-pepstatin which inhibited the proteolytic activity. While amprenavir, tipranavir, atazanavir, darunavir, lopinavir and DMP-323 were reported to inhibit the activity of XMRV protease, none of the them, nor nelfinavir or saquinavir had an effect on the activity of the Ty1 protease. The HIV-1 protease inhibitors showed no inhibitory potential in other retroviral-like protease studies either, they did not inhibit the human ASPRV1 or PEG10 protease. The XMRV protease activity was slightly inhibited by acetyl-pepstatin and pepstatin A, however pepstatin A did not inhibit the activity of Ty1 protease to any extent. We have determined the following values on Ty1 protease: IC_{50}=367.5 nM and a K_i=296 nM. In the case of HIV-1 protease these values are lower (K_i=13.15 nM, IC_{50}=1.18 nM) meaning the inhibitory potential is higher, acetyl-pepstatin has a higher impact on HIV-1 protease activity. The inhibitory potential is comparable to the XMRV protease (K_i=712 nM, IC_{50}=1290.2 nM). However, in the case of the copia retrotransposon protease interestingly, similar values were measured (Ki=15 nM) to the HIV-1 protease. In the case of Ty1 protease, not only one of the retroviral protease/aspartic inhibitors showed inhibitory potential, but the measured values confirmed that even acetyl-pepstatin is a weak inhibitor of Ty1 protease. Acetyl-pepstatin previously were reported to bind to XMRV PR in a unique mode, simultaneously two molecules of the inhibitor can bind to the active site of the protease in a head-to-head orientation. Crystallographic studies

in the future might help to investigate and reveal whether the binding of the inhibitor in the case of Ty1 PR resembles that of the XMRV.

4.1.6 In silico studies

The structure of the Ty1 protease has not been solved experimentally to date, therefore it was predicted with homology modelling. The protease domain is longer than the usual retroviral and retroviral-like protease domain, at 181 amino acids long. Therefore, the N- and C-terminal regions are extended, substantially longer than in the case of retroviral and retroviral-like proteases. The structural and functional role of these regions were not characterised. The slightly longer N- and C-terminal regions were observed in other retroviral proteases, for example in the case of MMLV protease, however, the terminal regions were only several residues longer than that of HIV-1 PR. For Ty1 protease it is around 30-40 amino acid extension in both regions. Moreover, the extended regions do not affect the activity of the MMLV protease, but the precise processing of the N-terminal region prior to the dimerisation domain of HIV-1 protease has been proven to have an enhancing effect on the activity. MMLV and XMRV proteases show 98% sequence identity and only differ in 2 amino acid residues, therefore they can be compared to the Ty1 protease.

The overall arrangement structural elements of the Ty1 protease and the in silico predictions showed a good agreement with the structure of other retroviral and retroviral-like proteases. The consensus D-S/T-G-A active-site motif of retroviral proteases also corresponds with the D-S-G-A sequence in Ty1 PR. Ty1 PR was also predicted to share its general fold with the retroviral proteases.

The structure prediction of the N-terminal region (N1-H56) did not predict α-helixes or β-sheets in the region. Similarly, to the N-terminal extension in Mo-MuLV protease, the disorder prediction also indicated the unstructured nature of this N-terminal extension. It suggests that in both proteases this region is flexible and has unknown conformation. However, near the catalytic motif of the protease (A71-H75) an α-helix was predicted that may correspond to the additional helical insert previously observed in equine infectious anaemia virus (EIAV) and the DNA damage-inducible protein 1 (Ddi1) proteases.

Based on the results of predictions, the dimer interface of the homodimeric Ty1 protease in the C-terminal region, contains only β-sheets that are connected by loops. Moreover, each monomer was predicted to have four β-sheets, in contrast with equine infectious anaemia virus (EIAV) and HIV-1 proteases, where these C-terminal β-sheets show alternation, in the Ty1 protease this phenomenon was not observed. Because none of the known retroviral or retroviral-like homodimeric aspartic proteases have eight-stranded dimeric interfaces we suggest that the Ty1 protease dimer interface may be similar to Ddi1 protease, also consists only six β-sheets, therefore we used the structure of the Ddi1 protease as a template to model the dimer interface of Ty1 protease.

Modelling an eight-stranded dimer interface without a proper template would have been challenging, and would have only resulted in a rough estimation of the possible involvement of a fourth β-sheets in the dimer formation. Ty1 protease also showed high sensitivity to the urea concentration when compared to the HIV-1 protease, which can be explained by the organisation of the β-sheets in the dimer interface. While HIV-1 protease contains alternating

β-sheets in the N- and C-terminal region, XMRV and Ty1 proteases contain β-sheets only in the C-terminal region, which are not alternating. Both of them showed lower dimer stability, therefore higher sensitivity to the urea concentration than HIV-1 protease. This implies that stability of the C-terminal dimerisation region, which contain no alternating β-sheets is lower than of that of HIV-1 protease, proving that where β-sheets are altered, there is interaction between these structural elements that strengthen the dimer stability. The catalytic motif also might play a role in the dimer stability. Also the interactions between serine or threonine residues of the consensus active centre (D-S/T-G-A) of the retroviral aspartic proteases, called the fireman's grip, stabilise the dimerisation. It has been reported that those retroviral protease dimers that contain serine instead of threonine are less stable. This was confirmed by the T26S mutation of HIV-1 protease, that showed lower specific activity compared to the wild type, and by the S25T HFV mutant that was less sensitive to the urea concentration. In Ty1 protease we also find serine in the active centre.

In order to examine N- and C-terminal regions, we modelled the structure of the whole Ty1 protease. There is no structural data of the extended regions regarding any aspartic protease, therefore the predictions without a template are only approximate. The predicted model of the full-length protease was used to support the interpretation of the in vitro results. The kinetic measurements were performed on recombinant fluorescent protein substrates in order to study the involvement of the putative substrate-groove residues in substrate binding. The recognition of the amino acid residues of P10-P6 and P6'-P10' in the substrates did not prove the existence of the S-groove beyond all doubt. The possible conformational changes of Ty1 protease imply that the surface amino acid residues are not that exposed, accessible as in the case of HIV-1 protease, however as we mentioned the predicted model's confidence is quite low at the terminal regions, and does not provide enough evidence. However, steric hindrance of the surface amino acid residues and active site might be possible by the N- and C-terminal extensions based on the putative model. We propose that in the Ty1 protease the binding surface of substrate residues P10-P6 and P6'-P10' has a different structure than the S-groove of HIV-1 protease or it is absent.

Without more accurate structural models or molecular dynamical calculations we are unable to confirm, if the Ty1 protease surface residues are inaccessible or the involvement of the N- and C-terminal regions are underlied. The hydrophobicity index of the Ty1 and Ty3 proteases' natural cleavage sites (for P10-P10' residues) have already been studied and determined. Remarkable differences were revealed between the cleavage sites compared to retroviral proteases, but the structures have not been analysed before. We used the predicted model of Ty1 protease to study the substrate binding pockets. The P4-P1 site amino acid composition was determined by alignment based on structural data, for Ty1 protease. For HIV-1 protease it has already been determined, therefore, the binding pockets for Ty1 protease have been mapped based on the corresponding positions and identification of residues in HIV-1 protease.

Similar to HIV-1, the S1 binding pocket of Ty1 protease consists mostly of hydrophobic residues. Retroviral proteases preferably bind hydrophobic residues in the P1 site of the substrates, but in the case of Ty1 protease the residues at the P1 and P1' sites are both usually hydrophilic. Based on the model structure, the S2 site is also hydrophobic, corresponding to the hydrophobicity of P2 and P2' sites, where usually a valine or isoleucine is positioned. At position S3 usually we find hydrophilic residue which is in agreement with the P3 and P3'

positions' residues, where we can mostly find hydrophilic residues in the substrate sequences. The P4 and P4' positions are neither hydrophilic nor hydrophobic. The S4 has been identified as hydrophobic, however, it is important to note, that this site is exposed to the surface and not well-defined. The distribution of hydrophilic, hydrophobic and charged amino acids in the Ty1 protease substrate binging cavity mostly did not differ from the HIV-1 protease, except the Ty1 protease S3 cavity, that contains no charged amino acids. Even if the binding site residues composition is in agreement with the profiles of cleavage sites, based purely on the binding cavity composition the specificities cannot be estimated accurately. To determine the enzyme specificity, extended in silico calculations are required on the enzyme-substrate complexes, deploying modified substrate series for detailed investigation.

4.2 Studies on SPINK1

4.2.1 Missense SPINK1 variants with preserved secretion

The secretion studies were performed by Alexandra Demcsák Alexandra and Miklós Sahin-Tóth. Our previous experiments and data from literature implies that most of the mutations of SPINK1 inhibitor cause secretion deficiency. However, 7 SPINK1 mutations (N34S, N37S, K41N, I42M, P55S, R65Q and Q68R) did not affect the expression and secretion of the inhibitor. The experiments were performed using both C-terminal His_{10}-tagged and native SPINK1 variants, in order to rule out the impact of the His_{10}-tag on the proteins folding. We found that the His_{10}-tag in general enhanced the secretion of SPINK1 proteins. The expression of variants K41N and Q68R slightly diminished compared to the wild type. The level of expression for the other variants did not change remarkably when compared to the wild type. The enhanced secretion of Q68R variant was reported earlier, however, our results have not confirmed this. The diminished secretion of R65Q variant has also been reported, but we could not reproduce the results, observing the opposite. We have also measured the inhibitory potential of the variants on trypsin in the supernatant of transfected cells. We could not detect inhibition on trypsin by K41N SPINK1. With further experiments we were aiming to find answers to whether the secreted SPINK1 variants promote or enhance the risk of chronic pancreatitis.

4.2.2 Modelling of trypsin-SPINK1 complex

The modelling of trypsin-SPINK1 complex was performed by András Szabó. To date there is no available structural data on the complexation of trypsin with SPINK1 inhibitor, therefore our research team has built a model where the native and sulphated human cationic trypsin was superpositioned with the complex of recombinant SPINK1 K41Y or I42E variant with bovine chymotrypsinogen A. Finally, we have restored the reactive site Lys41-Ile42 amino acid residues. The residues in the seven examined variants imply that the K41N and I42M mutations might have a negative effect on the strength of the peptide bond in the reactive loop Lys41-Ile42 (P1 and P1' positions). The affected side residues in the other variants N34S, N37S, R65Q and Q68R, probably do not directly interact with the trypsin residues, therefore they are less likely to disturb inhibitor binding. Interestingly, in the model structure the sulphate

group of the trypsin Tyr154 residue is proximal to the side residue Pro55, suggesting that the sulphation might affect the binding of P55S variant.

Moreover, the SPINK1 Tyr43 side residue, in the reactive loop P2' position, also appears to be sterically close to the sulphate group of the trypsin Tyr154. These predictions suggest that the mutation at P2' position of SPINK1 weakens the interaction with trypsin. We aimed to test these suggestions.

4.2.3 The wild type and N34S SPINK1 binding to human trypsins

The question is open, whether the N34S mutation changes the interaction between SPINK1 and trypsin, therefore we have investigated this mutant thoroughly. For the experiments we have used recombinant SPINK1-His_{10}, as detailed in the material and methods. Previous studies have reported that the inhibition of the trypsin with the wild type and N34S SPINK1 is comparable, however the experiments had limitations, being only semi-qualitative at best. We have measured the equilibrium binding of N34S SPINK1 to *Escherichia coli* expressed recombinant trypsin, that goes through no posttranslational modification, therefore it is not sulphated.

The wild type SPINK1 shows strong binding to the cationic (K_D 1.1 pM) and anionic (K_D 0.3 pM) trypsins. Similar values were measured in the case of N34S mutant as well, for cationic trypsin K_D=1.5 pM, for anionic trypsin 0.4 pM. As the structure prediction implies, these values were higher when measured with sulphated trypsins purified from human pancreatic juice. For the wild type SPINK1 we measured K_D of 62.2 pM for cationic, 36.7 pM for anionic trypsin, in the case of N34S mutant we have measured 32.3 and 16.7 pM.

The results indicate that the sulphation plays an important role in the interaction between SPINK1 and trypsins. The N34S mutant inhibitory profile suggests that the mutant protein inhibits the trypsins slightly more, however it is not a remarkable difference. We can conclude, that the inhibitory potential of N34S SPINK1 is not weaker compared to the wild type SPINK1. In biological systems the association and dissociation rate constant of the inhibitor can be more important than the equilibrium binding, therefore besides the K_D values, we have measured the K_{on} and K_{off} with the wild type and the N34S SPINK1 mutant, for cationic and anionic sulphated and non-sulphated recombinant trypsins. The results were similar for both inhibitors. When we compare sulphated and non-sulphated recombinant trypsins SPINK1 binding, association to the non-sulphated recombinant trypsins is faster, whilst for sulphated trypsins the dissociation rate is higher. The higher dissociation rate constants explain the higher K_D values, which we determined during the measurements of equilibrium binding constants. The K_D values were calculated from the reaction rate constant and were compared to the equilibrium binding strength. The results were all within the pM or lower concentration range in the case of non-sulphated recombinant trypsins. It is important to note, however, that the method of measurement is not reliable in the sub-picomolar K_D range. The calculated K_D values of sulphated and non-sulphated recombinant trypsins were two- to five-fold lower than the measured values. Because two experimentally different methods were used to determine the KD values, this is within acceptable variation limits.

4.2.4 Temporary inhibition of trypsins by wild type SPINK1 and N34S mutant

SPINK1 is a so-called temporary inhibitor, being inactivated over time and releasing the free trypsin. As the first step of release, trypsin cleaves peptide bound in the SPINK1 reactive loop between Lys41-Ile42, that is followed by the cleavage of the peptide bound between Arg67-Gln68, and cleavage of other sites in the inhibitor. When we compared the release of sulphated and non-sulphated recombinant trypsins from SPINK1 inhibition, we found that the sulphated trypsins were released faster than the non-sulphated recombinant trypsins.

This can be explained with the higher K_D values, as in the case of sulphated trypsins the inhibitor dissociation from the trypsins is faster. We confirmed the degradation of SPINK1 with Western blot, which was in agreement with the rising trypsin activity in the samples. The kinetics of the temporary inhibition was comparable for the wild type and N34S mutant, and the release was slightly slower for the mutant SPINK1.

4.2.5 Sulfation of trypsins weakens SPINK1 binding

We observed that the sulphation of tyrosine 154 of human trypsins weakens the binding of SPINK1 inhibitor and facilitates its release. To study these we have purified the sulphated trypsin isoforms from human pancreatic juice and expressed the recombinant non-sulphated versions in *E. coli*. To rule out any confounding effect we have also expressed the sulphated and non-sulphated trypsins in HEK293T cells, as it is described in the materials and methods. The trypsins were purified from the conditioned medium and determined the equilibrium binding constants with wild type and N34S mutants. The K_D values were always higher for sulphated trypsins than for non-sulphated trypsins, supporting the suggestion that the sulphation weakens the inhibitor binding. Interestingly, when compared the trypsins purified from *E. coli* expressions and non-sulphated trypsins purified from human pancreatic juice, to the HEK cell expressed trypsins, the K_D values were higher for the non-sulphated HEK293T cell expressed trypsins, and were lower for the sulphated HEK293T cell expressed trypsins. This results in seemingly less difference between the sulphated and non-sulphated isoforms of trypsins, but it does not change the observation that the sulphation has a negative effect on the SPINK1 binding, therefore weakening it.

4.2.6 Role of SPINK1 tyrosine 43

Molecular modelling suggested that the Y43 of SPINK1 is proximal to the sulphate group of the trypsin, thus it might affect the strength of binding. To confirm this, tyrosine 43 was mutated to alanine or arginine and the equilibrium binding constants were determined with sulphated and non-sulphated trypsins. Mutant SPINK Y43A showed twenty-fold, the Y43R mutant showed a hundred-fold reduction of binding measured with non-sulphated cationic trypsin when compared to the wild type inhibitor. In contrast, when the experiments were

performed with sulphated cationic trypsin the binding strength was comparable for Y43A mutant and wild type SPINK1, whereas Y43R mutant bound three-fold stronger. Because the Y43A mutant showed reduced binding to the non-sulphated trypsin compared to the wild type, we cannot conclude that tyrosine 43 of SPINK1 has an important role in stabilising the binding of the inhibitor to the trypsin through tyrosine 154, probably it is only stacking against it. In contrast, in the case of sulphated trypsin this interaction is absent, thus the KD values for Y43A mutant and wild type SPINK1 are similar. However, mutation to arginine improves the interactions with sulphated trypsin and reduces with non-sulphated trypsin, probably through an electrostatic interaction between the sulphate groups and guanidinium. Taken together the results, mutations of Tyr43 seem to strengthen the notion that this side chain has a role in trypsin binding and is a key determinant of the weaker SPINK1 binding to sulphated trypsins.

4.2.7 Binding of SPINK1 variants to native human trypsins

Besides the mutant N34S we have expressed, purified and studied 6 more SPINK1 mutants that showed no discrepancy of secretion. We measured their equilibrium binding to sulphated human anionic and cationic trypsins purified from pancreatic juice. From a clinical or pathological perspective, these interactions are relevant, therefore we aimed to characterise them. Mutant K41N only bound poorly to the above trypsin isoforms, we measured KD values in the micromolar range, that caused 20,000 to 30,000-fold reduction in affinity. The mutation I42M in the reactive site reduced binding affinity by three- to seven-fold. P55S mutant showed only a slight decrease of 1.6- to 3.4-fold reduction. N34S, N37S, R65Q, and Q68R mutants exhibited KD values that were smaller or similar to the wild type SPINK1. Altogether we propose, that only rare mutation in the reactive site K41N and I42M can directly affect the inhibitor binding to trypsins significantly.

Our observations imply, that the diminished affinity and binding of SPINK1 to the human trypsins is not a frequent phenomenon in chronic pancreatitis, and possibly associated with rare mutations in the reactive site. More likely, that the main pathogenic mechanism behind the SPINK1 mutations, is the loss of trypsin inhibition, caused by reduced expression, secretion and/or insufficient folding.

5. SUMMARY

During my studies I had the opportunity to study two proteases, their properties and their inhibition. We expressed and purified the protease from the *S. cerevisiae* Ty1 retrotransposon. We characterised the His-tagged and untagged form of the Ty1 protease using HPLC and fluorometric enzyme activity assay methods. The biochemical properties of the protease are similar to the retroviral and retroviral like proteases. We have measured and calculated the kinetic parameters of the protease (K_M, k_{cat}, k_{cat}/K_M) on artificial oligopeptides and fluorophore fused protein substrates mimicking natural cleavage sites. An extended surface binding site of HIV1 protease has been discovered recently, therefore with modified substrates we were studying the putative surface substrate binding grooves of Ty1 protease. We have not found proof of the presence of surface substrate binding sites. We have studied the effect of ionic strength, temperature, pH and urea concentration on the enzyme activity, the properties of Ty1 protease were comparable to the properties of other retroviral and retroviral-like proteases. Finally, we have performed studies with HIV1 and other aspartic protease inhibitors. Based on our results the Ty1 protease is naturally resistant against most of the HIV-1 inhibitors.

We have also studied the inhibitory potential of seven SPINK1 variants (N34S, N37S, K41N, I42M, P55S, R65Q, and Q68R) on sulphated and non-sulphated human trypsin isoforms. We determined the association and dissociation rate constants and calculated the equilibrium dissociation constant for the wild-type and mutant SPINK1 variants. We measured the K_D of the wild type SPINK1 with the recombinant (non-sulphated), and native (sulphated) human cationic and anionic trypsins. In the case of cationic trypsin we observed a 50-fold increase, whilst for the anionic trypsin we measured more than 120-fold increase of the K_D values. We studied the temporary inhibition of the sulphated and non-sulphated forms of human cationic trypsin by N34S SPINK1 and found no significant difference between the mutant and the wild type of the inhibitor. Moreover, we have studied the effect of two artificial mutations (Y43A and Y43R) and their interaction with the Y154 residue of sulphated and non-sulphated cationic trypsin. The inhibitory effect was dependent on the sulphation of the trypsin isoforms. Altogether, we suggest that the SPINK1 mutations do not affect the inhibitory potential significantly, and that the SPINK1 mutations probably contribute to the development of pancreatitis through the impaired expression of the gene and/or by the reduced folding and secretion of the inhibitor.

6. ACKNOWLEDGEMENTS

I would like to thank Prof. Dr. József Tőzsér the head of the Department of Biochemistry and Molecular Biology for his support during my PhD years.

I would like to thank Prof. Dr. Fésüs László former head of the department for enabling me to carry out my studies at the Department of Biochemistry and Molecular Biology.

My sincere thanks goes to Dr. András Szabó and Dr. János András Mótyán for their professional support in the published papers and doctoral thesis. I would like to thank Dr. Krisztinának Joóné Matúz for introducing me into laboratory work and scientific experiments. I would like to thank Szilvia Janics-Pető laboratory technician for her support in the laboratory throughout the years.

I would like to thank Alexandra Demcsák and Miklós Sahin-Tóth their contribution towards the preparation of the publications.

I would like to thank Dr. Ferenc Tóth, Dr. Mária Golda and Dr. Mohamed Mahdi for their professional support.

I would like to thank Dr. Zsófia Szojka, Norbert Kassay, Dr. Márió Miczi, Dr. Vanda Toldi and Katalin Nagy for their friendship and professional advice. I would like to thank Dr. Tibor Nagy for his professional support with the mass spectroscopy measurements.

Finally, I would like to acknowledge the opportunity and friendly environment provided by the employees of Department Biochemistry and Molecular Biology.

My PhD studies were financed by GINOP-2.3.2-15-2016-00044 “PHARMPROT teaming” project and by the Higher Education Institutional Excellence Programme of the Ministry of Human Capacities in Hungary, within the framework of the Biotechnology thematic programme of the University of Debrecen and by the GINOP-2.3.3-15-2016-00021 project, co-financed by the European Union and the European Regional Development Fund.

Moreover it was financed by the National Research Development and Innovation Office of Hungary grants PD120960 and FK127942, the National Institutes of Health grants R01 DK117809 and R01 DK058088, the OTKA Bridging Fund of University of Debrecen the János Bolyai Research Scholarship of the Hungarian Academy of Sciences grant (BO/00514/19/5), and the ÚNKP-20-5 New National Excellence Program of the Ministry for Innovation and Technology of Hungary from the source of the National Research, Development and Innovation Fund.

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/121/2022.PL
Subject: PhD Publication List

Candidate: Lívia Diána Gazda
Doctoral School: Doctoral School of Molecular Cellular and Immune Biology

List of publications related to the dissertation

1. Szabó, A., Toldi, V., **Gazda, L. D.**, Demcsák, A., Tőzsér, J., Sahin-Tóth, M.: Defective binding of SPINK1 variants is an uncommon mechanism for impaired trypsin inhibition in chronic pancreatitis.
   J. Biol. Chem. 296, 1-13, 2021.
   DOI: http://dx.doi.org/10.1016/j.jbc.2021.100343
   IF: 5.157 (2020)
2. **Gazda, L. D.**, Matúz, K., Nagy, T., Mótyán, J. A., Tőzsér, J.: Biochemical characterization of Ty1 retrotransposon protease.
   PLoS One. 15 (1), 1-24, 2020.
   DOI: http://dx.doi.org/10.1371/journal.pone.0227062
   IF: 3.24

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

List of other publications

3. Bozóki, B., Mótyán, J. A., Miczi, M., **Gazda, L. D.**, Tőzsér, J.: Use of Recombinant Fusion Proteins in a Fluorescent Protease Assay Platform and Their In-gel Renaturation.
JoVE. 143, 1-15, 2019.
DOI: http://dx.doi.org/10.3791/58824
IF: 1.163

4. Bozóki, B., **Gazda, L. D.**, Tóth, F., Miczi, M., Mótyán, J. A., Tőzsér, J.: A recombinant fusion protein-based, fluorescent protease assay for high throughput-compatible substrate screening.
Anal. Biochem. 540-541, 52-63, 2018.
DOI: http://dx.doi.org/10.1016/j.ab.2017.11.001
IF: 2.507

Total IF of journals (all publications): 12,067
Total IF of journals (publications related to the dissertation): 8,397

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.

23 March, 2022

DEBRECENI EGYETEM * Egyetemi és Nemzeti Könyvtár *