THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY (PHD)

Biochemical characterization of the human PEG10 and ASPRV1 retroviral-like aspartic proteases

by Mária Golda

Supervisor: Prof. Dr. József Tőzsér

UNIVERSITY OF DEBRECEN
DOCTORAL SCHOOL OF MOLECULAR CELL AND IMMUNE BIOLOGY

DEBRECEN, 2021

Biochemical characterization of the human PEG10 and ASPRV1 retroviral-like aspartic proteases

By Mária Golda, MSc, molecular biologist

Supervisor: Prof. Dr. József Tőzsér, PhD, DSc

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

Examination Committee

Head:	Prof. Dr. Gábor Szabó, PhD, DSc
Members:	Prof. Dr. János Minárovits, PhD, DSc
	Dr. Tamás Emri, PhD, DSc

The Examination will be organized online at 9:00 am 15th of June, 2021

Defense Committee:

Head:	Prof. Dr. Gábor Szabó, PhD, DSc
Reviewers:	Dr. Gyöngyi Gyémánt, PhD
	Dr. Krisztián Kvell, PhD
Members:	Prof. Dr. János Minárovits, PhD, DSc
	Dr. Tamás Emri, PhD, DSc
	Dr. Gyöngyi Gyémánt, PhD
	Dr. Krisztián Kvell, PhD

The PhD defense will be organized online at 11:00 am 15th of June, 2021. Participation requires registration. For registration and further information please email to golda.maria@med.unideb.hu.

1. INTRODUCTION

1.1. Aspartic proteases

The aspartic proteases are one of the most important groups of the proteolytic enzymes. Based on the international classification applied by International Union of Biochemistry and Molecular Biology (IUBMB) - which classifies the enzymes according to the type of catalyzed reaction - the aspartic proteases were registrated with the catalog number EC 3.4.23.

Most of the aspartic proteases are synthesized in inactive proforms (precursors or zymogens) which undergo a multiple-step maturation process resulting in the release of the active forms. For their catalytic activity there is a need for the cooperation of the two aspartates within the active site of the enzyme. The generally accepted scheme of hydrolytic process catalyzed by the aspartic proteases was first described in 1987. According to the concept, one of the catalytic aspartates is protonated (proton donor), while the other one is deprotonated (proton acceptor), which activates the nucleophil water molecule between the aspartates and attacks the carbon atom of the carbonyl group in the peptide bond of the substrate. Concurrently, the oxygen of the carbonyl group is protonated by the other aspartate, which establishes a tetrahedral intermediate state. The protonation and rearrangement of amide group of the substrate results in the dissolution of the tetrahedral state, thereafter, the two products are released.

In case of the retroviral proteases, the catalytic residues (D-S/T-G) are connected to each other by network of hydrogen bonds, which is referred to as *fireman's grip*. The role of the residues forming catalytic triplet (D-T-G) in human immunodeficiency virus type 1 (HIV-1) was also confirmed with experimental methods. The Asp25 and Asp25' residues of the motif directly interact with the substrate/inhibitor, hence their modification lead to the inactivation of the enzyme. The strong hydrogen bonds formed between Thr26 and Thr26' residues play a pivotal role in the dimer stabilization by the *fireman's grip*. The glycine residues of the triplet (Gly27 and Gly27') have a role in the recognition and binding of the substrate, which enables the catalysis.

The unified names of the substrate binding pockets and the substrate residue were standardized in 1967. The residues at the N-terminal side of the scissible

bond are numbered as P3, P2, P1, *etc*., while the residues of the C-terminal side are numbered as P1', P2', P3', *etc*. The respective substrate binding sites of the enzyme are referred as S1, S2, S3, *etc.* and S1', S2', S3', *etc*.

The literature regards pepstatin as a general inhibitor of the aspartic proteases. The pepstatin molecule and its salt derivatives were found to be proper pepsin inhibitors. As a reversible peptide-mimetic inhibitor, the pepstatin was proved to be a highly effective cellular aspartic protease inhibitor in *in vitro* experiments. However, in case of the inhibition of viral proteases this inhibition was less effective. Simultaneously with the discovery of pepstatin A, a novel aspartic protease inhibitor was isolated from the *Streptomyces naniwaensis* species (S-PI, as *Streptomyces*-pepsin inhibitor). The experiments showed that S-PI was practically the N-acetyl form of the pepstatin A. Acetyl-pepstatin can be considered as the effective inhibitor of the HIV-1 and HIV-2 proteases. Although the pepstatins are not used in the clinical practice, the better insight into their inhibitory mechanisms highly contributed to the discovery of novel drugs against human patogenic viruses. The majority of the first- and second-generation HIV protease inhibitors are peptide-mimetic, which bind to the active site of the enzyme in order to inhibit the hydrolysis of Gag ang Gag-Pol fusion proteins and eventually stop the virus maturation. The tipranavir and darunarvir are exceptions, these are non-peptide-mimetic inhibitors. It is worth mentioning that, the rapid resistance developing against the HIV protease inhibitors singificantly restricted the application of peptide-mimetic inhibitors in the clinical practice.

1.2. Retroviral-like aspartic proteases in the eukaryotic genom

Whilst the endogenous aspartic proteases possess a bilobar structural composition - which is a result of evolutionary gene duplication (pepsin) - the genes of retropepsin aspratic proteases code only for one catalytic domain and are able to form homodimers with high probability. Such retroelement is the Ddi1 protein (*DNA- damage inducible 1*) being also in the eukaryotic genom. Although, the role of the yeast Ddi1 (yDdi1) protein as extra-proteasomal ubiquitin receptor was characterized, the function of the strongly conserved aspartic protease domain remained to be elucidated.

The human genome encodes two Ddi1-like proteins, the Ddi homolog 1 (*human* Ddi1, hDdi1) and the Ddi homolog 2 (*human* Ddi2, hDdi2) proteins. Based on its genomic organization, hDdi2 is supposed to be an ortolog of the original yDdi1, from which the hDdi1 was formed by retrotransposition. They contain both an ubiquitin-like domain (*ubiquitin-like*, UBL) and an aspartic protease domain, but the role of human Ddi1-like proteins in the transport and targeting of proteins to proteasomal degradation is not clarified to date.

The nuclear receptor-interacting protein 2 (NRIP2), which is also known as neuronal-interacting factor (NIX-1) is exclusively expressed in neurons of the central nervous system. Between its nuclear receptor binding motifs (LQTLL, LLQAL) there is a D-T-G aspartic protease motif, whose role in cellular function is unrevealed.

1.2.1. *peg10* (paternally expressed gene 10)

The *peg10* is a retrotransposon derived gene integrated into the human genome (7q21.3). It is expressed exclusively from paternal chromosome, the transcription from maternal allele is inhibited by DNA methylation.

Due to the presence of alternative splicing two transcript variants exist (*peg10*-A 6573 bp, *peg10*-B 6584 bp), but not only the alternative splicing is responsible for the expression of different PEG10 protein variants in various tissue types. Besides the alternative translational start sites, the rare -1 ribosomal frameshift translation in the human genom, and the posttranslational autoproteolysis may also result in the appearance of PEG10 proteins with different molecular weight.

In case of *peg10,* the retrotransposon and retroviral specific -1 frameshift translational mechanism was observed. Due to a -1 frameshift translational mechanism, two overlapping open reading frames exist which are referred as RF1 and the RF2. The retrotransposon origin of *peg10* LTR was proved by its rare translational mechanism, as well as the conserved domain organization. There are long terminal repeat (LTR) sequences at the termini of LTR-retrotransposons, and contain overlapping open reading frames (ORF1 and ORF2), the *gag* encodes CA and NC structural proteins, while the *pol* codes for protease (PR), reverse transcriptase (RT), and integrase (IN) enzymes.

Although, *peg10* lacks LTR regions, it is classified as a member of Ty3/Gypsy family of retrotransposons, due to containing a bilobal CA-like domain, a zinc finger (CX2CX4HX4C) and a strongly conserved aspartic protease motif (D-S-G), and a truncated RT domain. Unlike retroviruses, retrotransposons lack the gene encoding the envelope proteins (*env*) and their life-cycle is devoid of extracellular steps.

The PEG10 was found to be located in the cytoplasm and nucleus, as well, the localization was supposed to depend on the presence of putative binding partners. The potentially most important binding proteins identified to date include the E3 ubiqutin-protein ligase (SIAH1) and activin-like kinase 1 (ALK1), latter one acts as a cell membrane receptor and belongs to the TGF-β-1 family.

The literature data about the main characetristics of PEG10 aspartic protease domain are still limited. While it has been proved that PR_{PEG10} is an active protease, the enzymatic and functional features have not been elucidated to date.

1.2.2. ASPRV1 (*retroviral-like aspartic protease 1*)

The skin is the first defense line between our body and the external environment. For the maintenance of its integrity and homeostatis the activation of proteolytic cascades and the processing of target proteins are needed. The *retroviral-like aspartic protease 1* (ASPRV1) or otherwise referred as to *Skin-specific Aspartic Protease* (SASpase) is also a member of such skin cascades, it is specifically expressed in *stratum granulosum* layer of the epithelium. The translated ASPRV1 is processed in the skin into smaller fragments, which are named according to their molecular size. The full-length protein that contains a putative transmembrane domain has a 37 kDa molecular weight (SASP37). The pro-form is considered to be SASP28 (28 kDa) and its autoprocessing results in the release of the shortest form SASP14 (14 kDa).

The only known natural substrate of the mature SASP14 is profilaggrin (pro-FLG), which is composed of filaggrin (FLG) monomer repeats and has >400 kDa molecular weight. During its maturation, pro-FLG is cleaved at the GSFLY↓QVST intermonomer linker sequences, leading to the release of FLG

monomers. Thereafter, the monomers are further modified and degraded into small peptides and amino acids, which play pivotal role in moisturization and protection of the skin from external environmental factors.

Previous studies focused on the expression and function of ASPRV1 in epidermis, but recently ICAM1^{+}macrophage-like neutrophil cells of the mouse and human immune systems were also found to express ASPRV1. Its mRNA was found to be the most abundant in blood neutrophils as compared to other types of leukocytes, and elevation of its level was detected in the brain lesions of patients suffering from severe multiple sclerosis.

2. Aims

In my doctoral work we aimed the biochemical characterization of two retroviral-like human proteins.

The regulation of the PEG10 expression plays a cardinal role in the embryonic development and in the postnatal period, as well. The expressional changes of PEG10 may lead to severe diseases, however, in the literature the majority of the publications designate the protein only as PEG10, whilst the significance of the strongly conserved aspartic protease domains in the $RF1/RF2_{PEG10}$ poliprotein still remained unknown. Our goal was to reveal the enzymatic activity of PEG10 and investigate its biochemical characteristics in cellular systems.

The function of human ASPRV1 is essential in skin. The enzyme formed by autocatalysis is responsible for the hydrolysis of the pro-FLG into fillagrin monomers, its disorder is known to be associated with some skin-related diseases. Although, the role of ASPRV1 protein is relatively elucidated, its significance in other cells (such as neutrophil granulocytes) is less unravelled. Furthermore, the information about the structure and biochemical features of the enzyme are also limited.

The specific aims of the work were the followings:

- We aimed the generation of expression constructs to study the retrotransposon-derived PR_{PEG10} protein, including determination of the optimal conditions (*e.g.*, pH optimum) and strategy (cis-trans activity) of protease activity, investigation of the effects of protease inhibitors, and determination of the effects of modified PEG10 proteins on cell viability and proliferation.
- We planned characterization of the enzymatic features of GST-tagged SASP28 and SASP14 enzymes, with a special emphasis on the susceptibility towards HIV protease inhibitors, specificity and kinetic parameters, and we aimed to prove the phenomenon of autoactivation as well as the autoproteolysis by studying wild type and mutated GST-SASP28 enzymes.

3 Materials and methods

3.1. Characterization of PEG10

3.1.1. *In silico* analysis

The secondary structure of PR_{PEG10} was predicted by Jpred4 server, the IUPred webserver was used for disorder prediction. Homolog model of $RF1/RF2_{PEG10}$ (346-477 region) was prepared by using Modeller9v13 program, the structures were visualized using PyMOL Molecular Graphics System program. The prediction of ubiquitination sites in PEG10 protein were performed by UbiSite, BDM-PUB and UbPred online web servers (work of Dr. Mótyán János András).

3.1.2. Modification and cloning of PEG10 sequences

The coding sequence of human *peg10* cDNA sequence was cloned into pQE-TriSystem expression plasmid, the empty plasmid was the kind gift of Prof. Dr. Zoltán Papp (Division of Clinical Physiology, Institute of Cardiology, Faculty of Medicine, University of Debrecen).

For the cloning, HindIII and XhoI restriction enzymes were used. As the *peg10* sequence contains and internal XhoI restriction site (661-666 bp), it was necessary to carry out its synonym point mutation before cloning. A thrombin cleavage site was designed to facilitate removal of the polyhistidine tag from the fusion protein. The insert was ligated into the expression plasmid by T4 DNA ligase. To ensure the translation of the full-length $RF1/RF2_{PEG10}$ sequence, insertional mutagenesis was performed in the frameshift site, the translated framesift mutant protein was referred as $_{fs}RF1/RF2_{PEG10}$. D369A, D370A, and S371A mutations of PR_{PEG10} were generated using a QuikChange mutagenesis kit according to the manufacturer's instructions. Sequencing was performed using gene- and vector-specific primers by the BigDye Terminator v3.1 Cycle Sequencing Kit.

3.1.3. Expression of recombinant PEG10 protein in eukaryotic cells

Human embryonic kidney 293 (HEK293T) cells were transiently transfected with PEG10 expression constructs at 60 % confluency using a polyethylenimine (PEI) transfection protocol. We used empty pQE-TriSystem plasmid DNA as a mock control.

3.1.4. Lysis of HEK293T cells

Cells were trypsinized and collected by centrifugation 24 h after transfection. The pellet was resuspended in phosphate-buffered saline containing a protease inhibitor cocktail. The pellets were lysed by sonication and then centrifuged, the proteins were detected in the supernatants by Western-blot.

3.1.5. Production of recombinant substrate in bacterial cells

For the expression of the His_6-MBP-$_{fs}$RF1/RF2$_{PEG10}$ (1-345)-mTurquoise2 recombinant fusion protein substrate, the coding sequence of 1–345 residues of the full-length $_{fs}$RF1/RF2$_{PEG10}$ protein was amplified from the pGEX-4T-3 expression construct using PCR. Both the amplicon and pDest-His_6-MBP-mTurquoise2 plasmid were digested with PacI and NheI restriction endonucleases followed by ligation. The substrate was expressed in BL21(DE3) *Escherichia coli* cells. After cell lysis, the protein substrate was purified by Ni-NTA magnetic agarose beads, and the buffer was changed using 10K Amicon Ultra centrifugal filter tubes. The purified recombinant protein was applied as the substrate of PR$_{PEG10}$.

3.1.6. *In vitro* analysis PR$_{PEG10}$ activity

After cell lysis and centrifugation, supernatants were dialyzed overnight against storage buffer. Protease inhibitor coctail was also added to cell lysates to inhibit the proteolysis of the protein of interest by serine and cysteine proteases present in the cellular extract. The proteolytic activity of $_{fs}$RF1/RF2$_{PEG10}$ was assessed by incubating the total cell lysates of transfected cells overnight. For the “0 h” sample, the collected cell lysate was frozen immediately at −20°C. To investigate the susceptibility of a recombinant

protein substrate for cleavage by PR_{PEG10}, the total cell lysate of transfected HEK293T cells expressing $_{fs}RF1/RF2_{PEG10}$ protein was incubated with His_6-MBP-$_{fs}RF1/RF2_{PEG10}$ (1-345)-mTurquoise2 fusion protein in storage buffer for 20 h at 37°C.

3.1.7. Determination of pH optimum

Effect of pH on PR_{PEG10} activity was determined in META buffer. Following transfection of cells, cell lysate was dialyzed against storage buffer (pH 7.0) and then incubated at 37 °C for 20 h in a 1:1 volume ratio of a series of META buffers (in 5.0-8.0 pH range).

3.1.8. Western-blot analysis

Samples were loaded onto a 14 % SDS-polyacrylamide gel, followed by electrophoresis and transfer of proteins onto a nitrocellulose membrane. The membrane was blocked by dry milk in TBS buffer at room temperature, followed by overnight incubation with the antibodies. The membranes were washed with TTBS and were then incubated with polyclonal anti-mouse or anti-rabbit secondary antibodies at room temperature. The membranes were subsequently washed a further three times with TTBS. Proteins were detected using SuperSignal West Pico chemiluminescent substrate.

3.1.9. *In vitro* inhibition of PR_{PEG10}

Pepstatin A, lopinavir, nelfinavir, saquinavir, darunavir, and tipranavir - dissolved in DMSO - were used for the inhibition of PR_{PEG10}. The reaction mixtures contained the lysate of $_{fs}RF1/RF2_{PEG10}$-transfected cells (dialyzed against storage buffer) and the HIV protease inhibitors (10 μM final concentration), and were incubated at 37 °C for 20 h. Only DMSO was added to the control sample. The reactions were terminated by the addition of SDS loading dye, and the mixtures were boiled at 95 °C for 10 min and centrifuged briefly before Western-blot analysis.

3.1.10. Cell viability assay

HEK293T and HaCaT cells were seeded in 96-well plates 16 h before transfection. Lipofectamine LTX&PLUS™ reagent was used for transfection according to the manufacturer's instructions, at 70-80 % confluence. Cell viability was determined by adding 1 mM MTT reagent to the cells that were incubated at 37 °C for 4 h. Then, MTT formazan crystals were dissolved using DMSO, and the absorbance was measured at 544 nm using an automatic microplate reader. HEK293T cells transfected with a pQE-TriSystem plasmid were used as mock control, whereas the efficiency of transfection was determined by measuring fluorescence of cells transfected with a pQE-TriSystem-GFP plasmid.

3.1.11. Cell proliferation assay by flow cytometry

HEK293T and HaCaT cells maintained in DMEM were plated in 6-well plates 24 h before transfection. Cells maintained in Opti-MEM medium lacking antibiotics were transfected with $_{fs}$RF1/RF2$_{PEG10}$ or D370A-$_{fs}$RF1/RF2$_{PEG10}$ DNA, using Lipofectamine LTX&PLUS™ reagent according to the manufacturer's protocol. For mock control, pQE-TriSystem was used to transfect cells. After incubation, 3 mL of fresh medium supplemented with 10 % FBS was added to the transfected cells, which were then further incubated for 36 h. The success of transfection was verified by determining the fluorescence of control cells that have been transfected with pQE-TriSystem-GFP plasmid. Cell proliferation was analyzed using flow cytometry using mouse-anti-human Ki67-FITC antibody.

3.2. Characterization of ASPRV1

3.2.1 *In silico* analysis

Secondary structure prediction was performed based on the sequence of SASP14 using PredictProtein, JPred4, DSC, SOPMA and GOR4 web servers. Hydropathy index values were obtained from the literature. SWISS-MODEL Workspace was applied for template search, Modeller 9v13 for homology modeling. The volumes of substrate-binding cavities were calculated based on a previously described method using the SiteID module of Sybyl program

package (run on Silicon Graphics Fuel workstations). Molecular visualizations were made by PyMol Molecular Graphics System. Stability analysis was performed using I-Mutant 2.0 web server. Multiple structure alignment was performed by using mTM-Align web server. (work of Dr. Mótyán János András).

3.2.2. Cloning and mutagenesis of SASpase

SASP37, SASP28, and SASP14 sequences were amplified from pCMV6-XL4-asprv1 vector by PCR. The amplified inserts were digested with BamHI and EcoRI restriction endonucleases, the inserts were subcloned into pGEX-4T-3 expression vector using Quick Ligation Kit. SASP28-A189K/N190I and SASP28-A167G/L168G/A189K/N190I auto-processing site mutants were generated by QuickChange Lightning Multi Site-Directed Mutagenesis Kit. Success of cloning and mutagenesis was proved by sequencing in all cases. (work of Katalin Nagy)

3.2.3. Expression and purification of GST-SASpase proteins

The GST-SASP28 and -SASP14 proteins were expressed in *Escherichia coli* BL21(DE3) competent cells. Cells were harvested by centrifugation and the pellets were lysed in lysis buffer followed by sonication. The solubilized fusion proteins were purified using an ÄKTA prime instrument. Fractions were dialyzed against store buffer. Protein concentrations were determined by Pierce BCA Protein Assay Kit. Coomassie Brillant Blue and PageBlue Protein Staining solution was used for gel staining. Densitometry was performed using the GelAnalyzer software.

3.2.4. Synthetic oligopeptide substrates

Oligopeptide substrates representing the naturally occurring matrix/capsid (MA/CA) cleavage site of HIV-1 (VSQNY↓PIVQ) and its P2- and P3-modified variants were in-house-stocks. Oligopeptide substrates representing the wild-type and modified pro-FLG linker sequences were ordered from BioBasic. All peptides were dissolved in distilled water, except the P4-phosphorylated GSFLY↓QVSTH peptide, which was dissolved in 50 % DMSO.

3.2.5. Activity measurements with GST-SASP14 enzyme

Activity assay mixtures containing enzyme, buffer and substrate were incubated at 37 ° C for 1 hour and stopped by the addition of TFA. Substrates and cleavage products were separated using reversed-phase chromatography column on Merck Hitachi HPLC system, using automatic injector. Increasing water/acetonitrile gradient was used for separation in the presence of 0.05 % TFA. The peptides were detected at 206 nm, followed by integration of chromatographic curves.

3.2.6. Determination of kinetic parameters and optimal conditions

The following activity assays were performed using VSQLY↓PIVQ peptide representing P2-Leu variant of HIV-1 MA/CA cleavage site as substrate. After activity measurements, the kinetic parameters were determined by fitting the data to Michaelis-Menten equation, using non-linear regression. We calculated the catalytic constant (k_{cat}) based on the active enzyme concentration. Kinetic parameters were determined at less than 20 % substrate hydrolysis, and data were evaluated using GraphPad Prism 5.01 program.

To investigate dependence of GST-SASP14 activity on pH and ionic strength, cleavage reactions were performed in META buffer. For the determination of optimal pH, the buffer contained 2 M NaCl, and pH ranged from 5.0 to 9.0. To study effects of ionic strength on protease activity, NaCl concentration ranged from 0 to 2 M (at pH 6.0). (work of Katalin Nagy.)

To determine urea dissociation constant, the activits of GST-SASP14 was measured in META buffer containing urea in 0-2 M final concentration range.

3.2.7. Inhibition of SASP14 by HIV-1 protease inhibitors

Pepstatin A, acetyl-pepstatin, indinavir, tipranavir, saquinavir, nelfinavir, darunavir, lopinavir, and amprenavir were in-house stocks. Acetyl-pepstatin was dissolved in acetic acid (50 %), while all other inhibitors in dimethyl sulfoxide (DMSO). VSQLY↓PIVQ oligopeptide was used as substrate of GST-SASP14. For screening of inhibitors, final concentration of DMSO was <1 %, all inhibitors were applied at ≥10 μM final concentration. To determine inhibitory constant and the concentration of the active GST-SASP14, indinavir

was applied in 0–10 μM range of final concentration. Statistical significances were calculated by using GraphPad QuickCalcs t test calculator.

3.2.8. Autoactivation

To investigate the autoactivation of ASPRV1, GST-SASP28 enzyme was pre-incubated at 37 °C for different times, followed by analysis of each sample by SDS-PAGE to determine the ratio of processed and unprocessed forms. All samples were supplemented with VSQLY↓PIVQ oligopeptide substrate, as well, and were incubated for 37°C. After addition of TFA, the substrate and cleavage products were separated by an HPLC-based method.

3.2.9. Cleavage site identification by HPLC-(+)ESI-TOF

For all HIV-1 MA/CA oligopeptide substrates that were efficiently cleaved by GST-SASP14, the molecular weights of the substrates and products were determined by HPLC coupled to electrospray ionization time-of-flight mass spectrometry (HPLC-(+) ESI-TOF). The HPLC-MS measurements were carried out by Waters 2695 Separation Module. VDSphere PUR 100 C18-M-SE reverse phase column was applied with gradient elution. A MicroTOF-Q type Qq-TOF MS instrument equipped with ESI ion source was used in positive ion mode. (The HPLC-MS analysis of the prepared samples was performed by Dr. Tibor Nagy at the Department of Applied Chemistry of the University of Debrecen.)

3.2.10. Cleavage site identification by MALDI-TOF-MS

To determine cleavage positions within the wild-type and autoproteolytic site mutant GST-SASP28 enzyme proteins, samples were incubated in reaction buffer (37°C) overnight. After the incubation, the samples were dialyzed against distilled water and then concentrated using Eppendorf Concentrator plus system. Samples were prepared using C18 loaded ZipTip pipette tips, and after loading and drying on the sample plate, MALDI-TOF MS measurements were performed on a Bruker Autoflex Speed mass spectrometer in linear mode. The cleavage position was identified by determining the molecular weight of the released SASP14 (work of Dr. Tibor Nagy.)

3.2.11. Statistical analysis

The GraphPadQuickCalcs free web calculator was used for statistical analysis (http://graphpad.com/quickcalcs/ttest1).

4 RESULTS AND DISCUSSION

4.1. Studies on PEG10

4.1.1. *In silico* analyses

Structural information about PR_{PEG10} was only limited (neither crystal- nor solution-structures have been determined to date), therefore, we applied *in silico* methods to investigate structural features. Based on the analysis, both the conserved active site motif and the arrangement of secondary structural elements implied high similarity with retroviral and retroviral-like proteases.

The disorder prediction showed that the region adjacent to the frameshift site of PR_{PEG10} (247-348 aa) has no ordered structure, and the region located in the proximity of the catalytic D-S-G-A motif has α-helical conformation. Predictions implied that C-terminus of PR_{PEG10} domain contains three β-strands connected by short loops, indicating that the homodimeric protein has a six-stranded dimer interface. A structure of the homodimer PR_{PEG10} was built by homology modeling, PR_{PEG10} was found to share its main structural characteristics with retrovirus-like Ddi1 and Ddi2 proteins. The model structure was used to investigate the specificity of PR_{PEG10} and to estimate its autoproteolytic cleavage site.

4.1.2. Detection of PEG10 proteins

PEG10 proteins were expressed in transiently transfected HEK293T cells. For the identification of the variants in cell lysates, $RF1_{PEG10}$-specific (1-325 aa) antibody was used, which was not sufficient to detect $RF2_{PEG10}$. The lysate of HEK293T cells was used as a control in Western-blot experiments. The quantitative comparison of different protein forms was not possible because of the polyclonal nature of the antibody. We observed no differences between the levels of D370A-$_{fs}RF1/RF2_{PEG10}$ and S371A-$_{fs}RF1/RF2_{PEG10}$ proteins' expression.

4.1.3. Validation of the proteolytic activity of PR_{PEG10}

In case of $_{fs}RF1/RF2_{PEG10}$, we observed appearence of two new bands (~45 kDa and ~37 kDa) in the Western-blot experiments, which correspoonded to

those proteolytic fragments (PF1 and PF2, respectively) which were released via autoprocessing of PR_{PEG10}. None of the variants (D370A-$_{fs}$RF1/$RF2_{PEG10}$ and S371A-$_{fs}$RF1/$RF2_{PEG10}$) carrying mutations within the catalytic triads showed proteolytic activity. To exclude the possibility that PEG10 is processed by other enzymes of the total cell lysate, the catalytically inactive D370A-$_{fs}$RF1/$RF2_{PEG10}$ protein was incubated with the lysate of non-transfected HEK293T cells. Our results proved that the proteolytic fragments were released as a result of catalytic activity of PR_{PEG10}, but not *via* processing by cellular proteases.

4.1.4. Estimation of proteolytic cleavage site

There are no information available in the literature about amino acid preferences and autoproteolytic cleavage site sequences of PR_{PEG10}, and in our experiments were also unable to identify cleavage positions *in vitro*. Therefore, the homology model of PR_{PEG10} was used to obtain information about the enzyme's specificity and together with the results of *in vitro* experiments estimate the cleavage site. Our *in vitro* results implied that $RF1_{PEG10}$ and PF1 proteolytic fragments show 4-5 kDa difference of their molecular weights. Based on this difference and structural analysis, the autoproteolytic cleavage site was predicted to be located between the CA-like domain and the zinc finger motif of $RF1_{PEG10}$.

4.1.5. PEG10 ubiquitination

Our *in vitro* studies implied post-translational modification of PEG10, so we used *in silico* algorithms, which proved the putative ubiquitination of the protein. To confirm the modification *in vitro*, the D370A-$_{fs}$RF1/$RF2_{PEG10}$ protein expressed in HEK293T cells was purified by Ni-chelate affinity chromatography and then Western-blot was carried out using anti-ubiquitin antibody. As a control, lysate of non-transfected HEK293T cells was used. Our assay clearly demonstrated the ubiquitination of D370A-$_{fs}$RF1/$RF2_{PEG10}$.

4.1.6. Trans-activity of PR_{PEG10}

The purified His_6-MBP-$_{fs}$RF1/$RF2_{PEG10}$-mTurquoise2 substrate was incubated with the total cell lysate of HEK293T cells containing $_{fs}RF1/RF2_{PEG10}$ protein, the proteins were detected by Western-blot (anti-MBP and anti-PEG10 antibodies). We proved that the expressed PEG10 protein is able for self-processing (cis-activity), the recombinant substrate was also proved to be susceptible for cleavage by TEV protease. Despite the intramolecular cleavage (self-proteolysis), the PR_{PEG10} was not able to cleave the recombinant substrate. The lack of the trans-activity may be explained with a self-inactivation mechanism, which phenomenon was also observed in case of the alphavirus capsid protease.

4.1.7. Effect of pH on autoproteolysis

The effect of pH on the autoproteolysis of $_{fs}RF1/RF2_{PEG10}$ was investigated. We found that $_{fs}RF1/RF2_{PEG10}$ is able for self-provessing in 6.0-8.0 pH range, but the activity was the highest close to neutral pH (6.9-7.4), which is highly similar to the pH optimum of human foamy virus (HFV) protease (pH 6.6).

4.1.8. Effect of HIV protease inhibitors on autoproteolysis of PEG10

We tested the inhibitors potential of some protease inhibitors against PRPEG10, such as the classical aspartic protease inhibitor pepstatin A, first- (nelfinavir and saquinavir) and second-generation (lopinavir) peptide-mimetic HIV PR inhibitors, furthermore, second generation HIV PR inhibitors (darunavir and tipranavir) with a „flexible" structure were also tested. We observed no inhibitory effect of the tested inhibitors on autoproteolysis, which means that PEG10 has natural resistance against these inhibitors.

4.1.9. Effect on cellular viability and proliferation

The viability of the HEK293T cells expressing $RF1_{PEG10}$ was significantly increased as compared to the control cells, and in case of $RF1/RF2_{PEG10}$ no detectable changes were obtained in HEK293T and HaCaT cell lines. When we monitored the results of the various PEG10 overexpressing cells, the significant increase was also maintained in the $RF1_{PEG10}$ transfected HEK293T samples in

comparison with that of RF1/RF2$_{PEG10}$ samples. In order to prove the fact, that PR$_{PEG10}$ may have an important influence on the cell viability, plasmids with the catalytically active ($_{fs}$RF1/RF2$_{PEG10}$) and inactive sequences (D370A-$_{fs}$RF1/RF2$_{PEG10}$) were transfected. If the results were compared with the mock control, the viability of the catalytically inactive PR$_{PEG10}$ protein (D370A-$_{fs}$RF1/RF2$_{PEG10}$) expressing cells did not change. However, the $_{fs}$RF1/RF2$_{PEG10}$ containing active PR$_{PEG10}$ protease in both cell lines strikingly reduced the ratio of the viable cells. Taken our results together we can suspect that PR$_{PEG10}$ can regulate the biochemical processes influencing the viability of RF1$_{PEG10}$ in the examined cells, which can be further investigated at the molecular level.

The effect of the transfection performed with different plasmids on the cell growth was carried out by identification of Ki67 protein expression. The RF1$_{PEG10}$ and RF1/RF2$_{PEG10}$ plasmid transfection into cells kept in optimal growth conditions did not induce any change compared to control group in either PEG10 variants. The comparison of frameshift mutants with control group also did not reveal significant changes, but if the $_{fs}$RF1/RF2$_{PEG10}$ and D370A-$_{fs}$RF1/RF2$_{PEG10}$ samples were examined based on their Ki67 positive cell ratio, then the significant difference was observed in HEK293T cells. The $_{fs}$RF1/RF2$_{PEG10}$ protein unlike the the catalytically inactive PR$_{PEG10}$ had an inductive effect on the cell growth. In case of HaCaT cells no substantial difference was found in cell proliferation test, which may be explained by the 24 h duplication time of the cells.

4.2. Characterization of ASPRV1

4.2.1. In silico analysis

The structural information about SASP14 were also limited, therefore, the structural characteristics were investigated with *in silico* methods. Like in case of the PR$_{PEG10}$, the SASP14 structural analysis also implied the presence of a helical insert adjacent to the catalytic motif and three C-terminal β-strands in the protease domain, indicating that the homodimer has a six-stranded dimer interface. The enzyme contains a D-S-G-A sequence, which corresponds to the

conserved active site motif of retroviral proteases (D-S/T-G-A), and the overall fold of SASP14 is similar to those of retroviral and retroviral-like proteases.

4.2.2. Design of SASPase mutations

In order to improve stability of SASP14 enzymes and prevent oxidation of the sidechain, the Met2 residue was mutated to Ile. Based on the sequence-based *in silico* analysis (I-Mutant 2.0), the mutation increases the stability of the enzyme but potentially has no effect on activity as this residues is not involved in substrate binding and dimerization. In our experiments the M2I mutant SASP14 was considered as the wild-type. In case of the GST-SASP28 form, we introduced mutation of the known autoproteolytic sites to study their effect on self-processing. In case of the A189K/N190I double mutant the P1 and P2 residues of the N-terminal SASP14 processing site were modified, whilst mutations of the P1 and P1' residues of the alternative cleavage site were introduced in the case of A167G/L168G/A189K/N190I quadruple mutant.

4.2.3. Expression and purification of GST-fusion proteins

The pre-proform of ASPRV1 (SASP37) was not studied due to the expression difficulties caused by the transmembrane region. The SASP14 and SASP28 proteins were expressed in BL21(DE3) bacterial cells in their GST-fused forms. After expresion, the proteins were purified by affinity chromatography, the purities of protein fractions were determined by SDS-PAGE, then the fraction of highest purity were dialyzed against storage buffer.

4.2.4. Kinetic parameters of GST-SASP14

We defined the kinetic parameters of the GST-SASP14 (K_M=1.26 mM; k_{cat}=0.034 s^{-1}; k_{cat}/K_M=0.027 $mM^{-1}s^{-1}$), which values indicated a lower catalytic efficiency of GST-SASP14 as compared to HIV-1 protease.

4.2.5. Effect of pH and ionic strength on enzyme activity

The pH optimum of the GST-SASP14 enzyme (pH_{opt}=6.27±0.02) was found to be closer to the physiological pH of the *stratum granulosum*, while in more acidic environment - such as in the upper layers of *stratum corneum* - its activity

was lower. Like the HFV and HIV-1 proteases, the GST-SASP14 also exhibited an enhanced enzyme activity by increasing the ionic strenght.

4.2.6. Determination of the urea-dissociation constant (UC_{50})

Urea dissociation constant (UC_{50}) means the urea concentration which reduces the activity of the active, homodimeric enzyme with 50 %. The activity of GST-SASP14 protease was measured at increasing urea concentration. We observed lower dimer stability for GST-SASP14 as compared to HIV-1 PR, which can be explained in part by the differences of their dimer interfaces. The proteases which have alternating N- and C-terminal β-strands in their dimer interface (HIV-1 PR) have significantly higher densitiy of intermonomer interactions as compared to those proteins whose interface contains non-alternating strands (*e.g.* XMRV, Ddi, and PEG10 proteases).

4.2.7. Examination of substrate specificity

The amino acid preferences of S2 and S3 substrate binding sites of GST-SASP14 were studied by using synthetic oligopeptide substrates representing the P2- and P3-modified variants of HIV-1 MA/CA (VSQNY↓PIVQ) cleavage site. We did not observe hydrolysis in case of the wild-type substrate (VSQNY↓PIVQ) after 1 hour incubation, and only negligable conversion was observed after 16 hours. None of the P3-modified oligopeptide substrates were cleaved efficiently even after 16 hour incubation, the P3-Val and -Asp mutants were not cleaved, while the P3-Gln, -Lys and –Gly mutants were also less efficient substrates, the cleavage efficiency was sginificantly lower as compared to that of the P2-Leu mutant.

To define the cleavage positions within the oligopeptide substrates, the substratesand products were analysed by HPLC-ESI-TOF. The retention time and mass/charge ratio (m/z) values were obtained for all peptides. The results showed that the substrates were cleaved between P1-Tyr and P1'-Pro residues in all cases, which was identical with the cleavage position of HIV-1 PR, alternative cleavage sites were not identified.

4.2.8. Effect of phosphorilation on substrate hydrolysis

Based on literature data, the pro-FLG can be phosphorilated at the intermonomeric cleavage sites, which may have a role in the prevention of the proteolysis between the FLG units, inhibiting protein maturation. We assumed that phosphorilated substrates can not be cleaved by SASPase, therefore, the cleavage of the substrate representing the natural FLG processing site) and its P4-phosphorilated form was also investigated. The efficient hydrolysis of the wild-type GSFLY↓QVSTH sequence was observed, but the P4-phosphorilated oligopeptide (GpSFLY↓QVSTH) was not processed.

4.2.9. Testing the effects of HIV-1 protease inhibitors

The effects of different inhibitors were tested, and the screening resulted in the indinavir as the most potent inhibitor, the inhibitory constant was determined for indinavir (K_i=0.62 μM). While examining the acetyl-pepstatin, we observed a phenomemon described earlier for HIV-1 PR. Using the acetyl-pepstatin in 10 μM final concentration we found increase of enzyme activity, while higher inhibitor concentration caused significant inhibition. This phenomenon was explained in case of HIV-1 by stabilization of the dimer by the inhibitor. In case of the SASP14 enzyme this mechanism remained unknown, therefore, additional experiments are needed for determination of the molecular background.

Like the PR_{PEG10}, the SASP14 was also not inhibited by amprenavir, darunavir, lopinavir, nelfinavir, saquinavir and tipranavir. We found that SASP14 contains more residues which correspond to the primary or secondary resistance mutations of the HIV-1 PR in the corresponding positions. From these only three residues are relevant to form the binding site in HIV-1 PR: the I47 in S2 and S4 pockets, the Q58 in the pocket S4, and the V82 in pocket S1. The SASP14 V66 and A80 wild-type residues were correspondent with the HIV-1 PR Q58V and G73A secondary resistance mutations, in spite of this in our experiment we observed inhibition of SASP14 with indinavir. Thus, the natural resistance to many inhibitors cannot be explained exclusively by the presence of HIV-1 resistance mutations at some positions.

4.2.10.Autoproteolysis of GST-SASP28 enzyme forms

Our results showed that the generated mutations (A189K/N190I and A167G/L168G/A189K/N190I) did not abolish completely the ability of GST-SASP28 for autoprocessing, however, it was substantially reduced. Based on the results of SDS-PAGE and MALDI-TOF-MS analyses we can state that the mutations did not cause shift of the cleavage position to the alternative cleavage site or to other position(s). According to our results the wild-type GST-SASP28 protein had more considerable autoproteolitic efficiency (80 %) than the mutants, which showed ~10 % conversion even after 20 hour incubation.

4.2.11.The validation of the autoactiation

The self-processing – leading to release of the SASP14 enzyme form - was ensured by the pre-incubation of the GST-SASP28 precursor. Before incubation the GST-SASP28, while after the incubation the SASP14 was the predominant enzyme form. Following the pre-incubation, the oligopeptide substrates were added to the samples, then the product formation was investigated by HPLC analysis. We found that the substrate conversion was proportional to the time of pre-incubation, it increased as a function of the amount of released SASP14. Our results imply that autoprocessing causes autoactivation, which can be directly concluded from the comparison of the catalytic activities of the two enzyme forms.

5 SUMMARY

For the characterisation of PR_{PEG10}, we designed expression constructs coding for the full-length polyprotein containing a frameshift mutation ($_{fs}RF1/RF2_{PEG10}$) and its active site-mutants (D370A-$_{fs}RF1/RF2_{PEG10}$ and S371A-$_{fs}RF1/RF2_{PEG10}$). Western-blotting utilizing $RF1_{PEG10}$-specific antibody was used to detect the presences of the proteins and their activity in transiently transfected HEK293T cells. The differences between the calculated (based on sequence) and observed (based on Western-blot) molecular weights indicated the ubiquitination of the protein, which was verified by using a ubiquitin-specific antibody. The analysis of the autoproteolysis implied a neutral pH optimum of the PR_{PEG10} (pH 6.9-7.4). Although, we observed intramolecular cleavage of PR_{PEG10} (the autoproteolysis resulted in the release of PF1 and PF2 products), we have not detected intermolecular interaction using a recombinant protein substrate. We were unable to identify the cleavage position *in vitro*, thus, it was estimated *in silico*. Despite testing multiple HIV protease inhibitors, we observed no inhibitory effect in any case. The effect of PR_{PEG10} on the viability and proliferation of HEK293T and HaCaT cell lines was also studied. The overexpressed PR_{PEG10} was found to decrease the number of viable cells, whilst its effect on the proliferation was observed only as compared to the HEK293T cells expressing the catalytically inactive D370A-$_{fs}RF1/RF2_{PEG10}$ mutant.

The studied forms of ASPRV1 (GST-SASP14, GST-SASP28) were expressed in BL21(DE3) cells, followed by purification with affinity chromatography. Thereafter, SDS-PAGE- and by HPLC-based methods were applied for i*n vitro* characterisation. In case of GST-SASP14, the enzyme kinetic parameters (K_M, k_{cat}, k_{cat}/K_M), and the dependence of activity on pH optimum, ionic strength, and urea was determined by using an oligopeptide representing the P2-Leu-modified version of HIV-1 MA/CA cleavage site, furthermore, the amino acid preferences of S2 and S3 binding sites were also studied. We confirmed the inhibitory potential of indinavir, and proved the inhibitory effects of pepstatin A and acetyl-pepstatin. We proved that autoproteolysis of GST-SASP28 causes autoactivation of the enzyme, the self-processing was studied by autoproteolytic cleavage site mutant enzyme forms.

ACKNOWLEDGEMENTS

I would like to thank my supervisor *Prof Dr. József Tőzsér*, full professor, the head of the department for his professional directions and support he had given me during my work.

I would like to thank †*Dr. Péter Bagossi*, that I was able to start my Ph.D. work in the Laboratory of Retroviral Biochemistry with his initial supervision.

I would like to thank *Prof. Dr. László Fésüs* academic, full professor, 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. János András Mótyán and Dr. Krisztina Matúz,* that they helped me throughout my work with their friendship, professional and personal support!

I would like to thank *Dr. Ferenc Tóth*, *Dr. Mohamed Mahdi*, *Dr. Zsófia Szojka*, *Lívia Gazda*, *Márió Miczi*, *Norbert Kassay*, *Szilvia Janics-Pető*, *Vanda Toldi*, their friendly and professional help, a cheerful and good atmosphere which also made the research work easier. My special thanks goes to *Katalin Nagy and Dr. Tibor Nagy* for their professional help. I would like to thank for *Balázs Kunkli, Boglárka Molnár, Dr. András Szabó, Nataly Morales Granda, Noémi Miltner, Tamás Linkner, Viktor Ambrus,* colleagues for that, I could count on them any time.

I would like to acknowledge the assistance of the *employees of Department Biochemistry and Molecular Biology* and all those who contributed to the preparation of my Ph.D. theses in any form.

I would like to thank my *family* for their support without whom all these could not have been achieved.

My Ph.D. work was financed in part by the GINOP-2.3.2-15-2016-00044 "PHARMPROT teaming" project, by the Higher Education Institutional Excellence Programme (NKFIH-1150-6/2019) of the Ministry of Innovation and Technology in Hungary, within the framework of the Biotechnology thematic programme of the University of Debrecen. This work was also financed by the TÁMOP 4.2.2.A-11/1/KONV-2012-0023 "VÉD-ELEM" project and TÁMOP-4.2.2.D-15/1/KONV-2015-0016 project, and by the Hungarian Science and Research Fund (K-101591, NKFI-125238).

UNIVERSITY of DEBRECEN

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/72/2021.PL
Subject: PhD Publication List

Candidate: Mária Golda
Doctoral School: Doctoral School of Molecular Cellular and Immune Biology

List of publications related to the dissertation

1. **Golda, M.**, Mótyán, J. A., Nagy, K., Matúz, K., Nagy, T., Tőzsér, J.: Biochemical characterization of human retroviral-like aspartic protease 1 (ASPRV1).
   Biomolecules. 10 (7), 1-25, 2020.
   IF: 4.082 (2019)
2. **Golda, M.**, Mótyán, J. A., Mahdi, M., Tőzsér, J.: Functional Study of the Retrotransposon-Derived Human PEG10 Protease.
   Int. J. Mol. Sci. 21 (7), 1-22, 2020.
   DOI: http://dx.doi.org/10.3390/ijms21072424
   IF: 4.556 (2019)

List of other publications

3. Kassay, N., Mótyán, J. A., Matúz, K., **Golda, M.**, Tőzsér, J.: Biochemical Characterization, Specificity and Inhibition Studies of HTLV-1, HTLV-2, and HTLV-3 Proteases.
   Life (Basel). 11 (2), 1-21, 2021.
   DOI: http://dx.doi.org/10.3390/life11020127
   IF: 2.991 (2019)
4. Mahdi, M., Mótyán, J. A., Szojka, Z., **Golda, M.**, Miczi, M., Tőzsér, J.: Analysis of the efficacy of HIV protease inhibitors against SARS-CoV-2·s main protease.
   Virol. J. 17 (1), 1-8, 2020.
   DOI: http://dx.doi.org/10.1186/s12985-020-01457-0
   IF: 2.579 (2019)

UNIVERSITY of DEBRECEN

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

5. Miczi, M., **Golda, M.**, Kunkli, B., Nagy, T., Tőzsér, J., Mótyán, J. A.: Identification of Host Cellular Protein Substrates of SARS-COV-2 Main Protease.
Int. J. Mol. Sci. 21 (24), 1-19, 2020.
DOI: http://dx.doi.org/10.3390/ijms21249523
IF: 4.556 (2019)

Total IF of journals (all publications): 18,764
Total IF of journals (publications related to the dissertation): 8,638

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.

01 March, 2021