To my dearest Grandparents
for their incredible strength…

THE DEVELOPMENT AND APPLICATIONS OF A RECOMBINANT FLUORESCENT FUSION PROTEIN-BASED *IN VITRO* PROTEASE ASSAY PLATFORM FOR KINETIC AND SPECIFICITY STUDIES

by Beáta Bozóki

Supervisor: Dr. *József Tőzsér*

UNIVERSITY OF DEBRECEN

DOCTORAL SCHOOL OF MOLECULAR CELL AND IMMUNE BIOLOGY

DEBRECEN, 2020

Contents

1. List of Abbreviations

AIDS	acquired immune deficiency syndrome
ANOVA	analysis of variance
CA	capsid protein
CFP/ ECFP	cyan fluorescent protein / enhanced cyan fluorescent protein
CV %	coefficient of variance
E. coli	*Escherichia coli*
EDTA	ethylenediaminetetraacetic acid
ES	enzyme-substrate complex
FET	fluorescent energy transfer
FLT	fluorescent lifetime
FP	fluorescent protein
FRET	fluorescent resonance energy transfer
GFP / EGFP	green fluorescent protein / enhanced green fluorescent protein
GST	glutathione S-transferase
H-bond	hydrogen-bond
His_6	hexahistidine tag
HIV-1	human immunodeficiency virus type-1
HTS	high-throughput screening
IMAC	immobilized metal affinity chromatography
IPTG	isopropyl β-d-1-thiogalactopyranoside
LB	Luria-Bertani
MA	matrix protein
MBP	maltose binding protein
MPC	magnetic particle concentrator
Ni-NTA	nickel-nitrilotriacetic acid
nsP	non-structural protein
ORF	open reading frame
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PMSF	phenylmethylsulphonyl fluoride

PR	protease
RFU	relative fluorescent intensity
RP-HPLC	reverse-phase high performance liquid chromatography
SAMB	substrate-attached magnetic beads
SAM-MTase	S-adenosyl-L-methionine-dependent RNA methyltransferase
SD	standard deviation
SDS	sodium-dodecyl-sulfate
SFV	semliki forest virus
SINV	sindbis virus
SUMO	small ubiquitin-related modifier
TCEP	tris (2-carboxyethyl) phosphine hydrochloride
TEV	tobacco etch virus
TFA	trifluoroacetic acid
VEEV	venezuelan equine encephalitis virus
YFP / mEYFP	yellow fluorescent protein / monomeric enhanced yellow fluorescent protein
β-ME	beta-mercaptoethanol

2. Theoretical background

2.1. Terminology of proteolytic enzymes

Bergmann and Ross (**1936**) applied the word 'peptidase' as a universal term for peptide bond hydrolase. Crystallographic studies indicated that in many of the cases, the active site of a peptidase is placed in the groove on the surface of the enzyme molecule and formed by adjacent structural domains. EC classification is the internationally accepted classification and nomenclature system of IUBMB (International Union of Biochemistry and Molecular Biology). In this system, the hydrolases (EC 3) acting on peptide bonds (EC 3.4) are subdivided divided into 14 sub-subclasses (EC 3.4.14). Major groups are illustrated on **Figure 1**. Beside EC classification, Rawlings and Barrett has started to develop a new type of classification and nomenclature system, called the *MEROPS*, that relies on the structural comparison and evolutionary relationship of the included peptidases (**Barrett, 2001; Rawlings et al., 2012**).

A model was introduced by Berger and Schechter (**1970**) that considers the catalytic site of the enzyme to be flanked on one or both sides by specificity 'subsites', each of them being able to accommodate the side chain of a single amino acid residue. The subsites are numbered from the catalytic site, $S1...S_n$ towards the N-terminus of the substrate, and $S1'...S_n'$ towards the C-terminus. The embedded amino acid residues of the substrate are marked $P1...Pn$, and $P1'...Pn'$, according to the followings:

Substrate: – P3 – P2 – P1 **+** P1'– P2'– P3'–

Enzyme: – S3 – S2 – S1 * S1'– S2'– S3'–
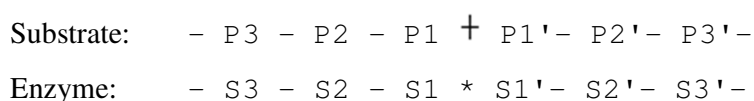
The catalytic site of the protease is marked by *, while **+** is the symbol of the scissile bond (the peptide bond to be cleaved) of the substrate. Proteases can be classified according to the chemical nature of their catalytic site. According to the Hartley (**1960**) and the proceeding discoveries (**Barrett 1980, 1986; Seemuller 1995; Fujinaga et al., 2004; Tajima et al., 2010**) up to date seven distinct groups of peptidases have been described: serine, cysteine, aspartic, metallo, threonine, glutamic and asparagine types of peptidases. Due to their relevance in this present study, cysteine proteases and aspartic proteases are discussed in detail.
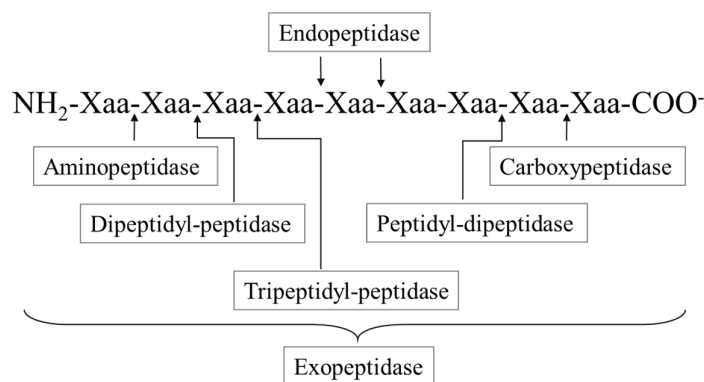
Figure 1. **Major groups of peptidases according to EC classification.** Arrows represent the typical cleavage positions of the different types of peptidases on a schematic polypeptide.

Proteases are essential components of all living organisms from viruses to humans. Proteases account for 2 % of the human genome and 1-5 % of the genome of infectious organisms (**Puente et al., 2003**). The complete set of proteases in a given organism is called the 'degradome' (**López-Otín and Overall, 2002**) and plays a role in almost all biological pathways (**Turk, 2006**; **López-Otín and Bond, 2008**), thus, alterations in the proteolytic system accounts for several pathological conditions, for example cancer, neurodegenerative disorders, and inflammatory and cardiovascular diseases. Proteases are key enzymes in the life-cycles of numerous infectious microorganisms and viruses, therefore, became potential therapeutics, drug targets and diagnostic and prognostic biomarkers in the pharmaceutical industry. Proteolytic enzymes take ~ 5-10 % of all pharmaceutical targets that are pursued for drug development (**Drag and Salvesen, 2010**).

As the genomic sequences of many model organisms has been solved, huge amount of information has become available regarding the different degradomes, the size of which is continuously growing as new proteolytic enzymes and catalytic mechanisms are described (**Lopez-Otín and Bond, 2008**). Consistent with this, there is a great demand for the development of both *in vitro* and *in vivo* technologies aiming either to profile the so called 'orphan proteases', the function and specificity of which are unexplored, or to further investigate the already described ones *e.g.* by compound profiling. A system-wide approach that includes the methods developed to study proteolytic enzymes, their substrates and their inhibitors is called 'degradomics' (**Lopez-Otin and Overall, 2002**). Degradomics involves the combination of biochemical studies, genetic tactics, cell-based assays, and proteomic methods that are essential in 'de-orphaning' the newly discovered proteases and to define the regulatory and functional interactions between the members of the different proteolytic systems (**Lopez-**

Otín and Bond, 2008). Assays aiming to characterize the proteolytic activity of the different macromolecules can be performed *in vitro* using purified proteins/peptides or may be performed *in vivo* using either prokaryotic or eukaryotic cell-based system. In this current study only *in vitro* protease assay systems are utilized.

2.2. *In vitro* protease assays

Quantification of proteolytic activity depends on the substrate type, the activity rate of the enzyme to be investigated, and on the sensitivity and precision requirements of the utilized detection system. Generally, *in vitro* protease assays detect protease activity on either whole proteins or short synthetic peptide molecules. Traditionally, whole proteins such as gelatin, casein, hemoglobin, *etc*. or their chromogenic derivates, like azocasein, azoalbumin or azogelatin are commonly used substrates mainly for screening general endopeptidase activity from crude materials. In some rare cases, using the native whole protein substrate is especially required *e.g.* collagen for assaying collagenase, however, their general use is limited in many cases. The purification of whole native proteins is often labor-intensive and difficult, their water-solubility is limited, furthermore, upon cleavage by the protease of interest the separation of the cleavage fragments from the intact substrate is needed to be performed. The main disadvantage of this substrate type is that due to the multiply and/or subsequently occurring proteolytic events, the real cleavage site sequence will rarely be defined, no relevant enzyme kinetic data can be obtained for a single cleavage site and the results obtained for proteases from different sources are not comparable. Moreover, due to the innumerable ionizable groups in the intact substrate, the changes in the velocity with the reaction conditions may reflect the susceptibility of the substrate for the proteolysis rather than the ability of the enzyme to process the substrate (**Gautam et al, 2001**). In cases, where a short peptide sequence in its denatured form carries the adequate information for the catalytic site of the enzyme, synthetic peptide sequences overcome the limitations of native whole protein substrates and offer a diverse and versatile tool for assaying protease activity. On the other hand, recent revolution of recombinant protein engineering technology has introduced new possibilities in the field of *in vitro* protease assays as well. The utilization of recombinant proteins as substrates can be a flexible alternative to both native whole proteins and synthetic oligopeptides.

2.2.1. Protease assays using synthetic peptide substrates

Due to their versatile nature, synthetic peptide substrate-based methods are the most popular protease assay platforms. Synthetic peptide sequences are usually modified by different reporter groups including fluorogenic, chromogenic, or luminogenic moieties. Class-specific substrates can be developed for the differentiation of endopeptidases or exopeptidases (**Gautam et al, 2001**). Several oligopeptide-based assay formats have already been developed, but the time- and cost efficient high-throughput screening (HTS)-compatible platforms are the most preferred. *In vitro* protease assays using synthetic peptide substrates can be classified into (i) homogenous, (ii) separation-based and (iii) heterogeneous assay formats (**Zhang et al., 2012; Ong and Yang, 2017**). In case of homogenous and separation-based assays both the substrate(s) and the protease are in an aqueous phase. In the former format the substrate turnover can be detected as a signal change directly from the reaction mixture without further sample processing, while in case of separation-based formats products and/or the remaining substrates are isolated from the reaction mixture after the termination of the reaction at a pre-determined linear range, where the amount of the generated products give a robust signal. In contrast, in heterogeneous assay formats, the substrates are immobilized on solid surfaces, while the protease of interest is in the aqueous phase.

2.2.1.1. Homogenous assays

The vast majority of homogenous proteolytic assays in drug discovery are based on fluorescence assay technology (**Woelcke and Hassiepen, 2009**), discussed in **Section 2.2.3**. Principally, its highly sensitive nature allows detection in a dense, low volume and HTS-compatible form (**Zhang et al., 2012**). Most commonly applied oligopeptide-based fluorescent formats use (i) chemically quenched dyes or (ii) are based on fluorescent energy transfer (FET) (**Figure 2**). In the former one typically a fluorescent dye comprises a reactive amine group and is covalently attached to the C-terminal end of a peptide substrate through an amide bond. The fluorescence of the dye in its peptide-attached form is quenched, however when the cleavage occurs, the dye is released, and its fluorescence is significantly increased due to the reconstitution of the fluorophore system. Notably this approach cannot be applied for proteases having strict specificity requirement for P1' site, e.g. for carboxypeptidases. In those cases, a similar solution, based on colorimetric difference between substrate and product was developed (**Plummer et al., 1980**). In FET-based protease assays a so called 'donor' and an 'acceptor' fluorescent groups are attached to each end of an oligopeptide molecule comprising the

cleavage site extended on both the non-primed and the primed sides. If the two dyes are in close proximity an efficient energy transfer takes place from the donor to the acceptor molecule. As a consequence, the fluorescence of the donor is quenched, while the fluorescence of the acceptor is increased.

Besides the undoubtedly most popular fluorescence-based assay platforms, several oligopeptide substrate-based homogenous protease assay methods have been developed by the application of other types of readout for example (i) spectrophotometric methods that monitor the difference in molar absorptivity between the substrates and products (**Gautam et al, 2001**), (ii) assay using radioactive peptide substrates offering great sensitivity and reliability even at low protease concentration and (iii) the novel bioluminescence-based techniques using aminoluciferin-conjugated peptides. In this latter case the bioluminescent signal is generated by the interaction of firefly luciferase and aminoluciferin formed upon cleavage of the peptide substrate by the protease.
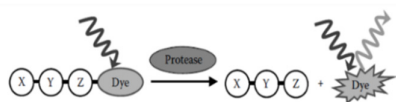
Figure 2. **Most popular oligopeptide-based fluorescent substrate formats.** (i) Assay using substrates with chemically quenched dyes can be optimized to A) fluorescent intensity or B) fluorescent polarization readout, while (ii) typical fluorescent energy transfer (FET)-based assay formats can be optimized to C) Fluorescent intensity or to D) fluorescent lifetime readouts **(Woelcke and Hassiepen, 2009).**

2.2.1.2. Separation- based assays

Typically, this assay format is used if the substrate cannot be labeled, or the labeled substrate is not suitable for specific studies. The separation of the substrate and the product is often performed by liquid chromatography and analyzed independently based on their retention times on a column, or based on their molecular weights by mass spectrometry (**Zhang et al., 2012**). This assay type combined with reversed-phase high-performance liquid chromatography

(RP-HPLC) is also commonly applied in our laboratory for different lentiviral (**Tőzsér et al., 1991**) and retroviral (**Fehér et al., 2006**) proteases. A great advantage of these platforms is that they allow kinetic reading of the enzymatic reactions, provide high sensitivity, can analyze both peptidyl or protein substrates, and the separated reaction fractions are possible to be collected for further analyses like mass spectrometry. Unfortunately, the use of HPLC-based approaches is also limited by the relatively high instrumentation cost, by the high vulnerability of columns that must be frequently changed, and by the fact that only one sample can be analyzed at a time and turn-around times for the injections may be long (**Gautam et al, 2001**).

2.2.1.3. Heterogeneous assays

Due to the limited number of probes, homogenous assays have limited multiplexed sensing capability, while heterogeneous platforms allow the simultaneous and multiplexed detection of proteases presented in a single sample. These assay types generally use large amount of substrate at different locations on a solid surface, which must be carefully chosen to be compatible with the signal readout to be detected. For example, colorimetric detection needs optically transparent surface, while in case of electric signals conductive surfaces are required. According to **Ong and Yang (2017)**, heterogeneous assay formats can be classified into (i) electrochemical assays, (ii) surface plasmon resonance assays, (iii) surface enhanced Raman scattering assays, (iv) liquid chromatography-based assays and (iv) other enzyme-linked assays.

2.2.2. Protease assays using recombinant fusion protein substrates

Recombinant protein engineering technology has introduced new possibilities in the field of protease assays by the utilization of recombinant proteins as substrates. These substrates are designed specifically to the experimental aims. They usually contain only a single cleavable site to eliminate the limitations hold for native protein substrates due to the multiple cleavage sites present in their sequences. Microbial systems offer cost-efficient, convenient and powerful tools for recombinant protein production (**Ferrer-Miralles et al., 2009**) although protein insolubility, conformational, structural and stability issues, low purification yields and host cell toxicity are potential challenges that are to be overcome (**Young et al., 2012**). In order to solve the listed difficulties, several different fusion tags have been developed, that can be applied to improve the characteristics of recombinant protein substrates (**Table 1**). Generally, a single fusion tag does not provide overall solution to all issues, therefore, dual- or multiple-tagging of the protein of interest is widely applied to satisfy the different needs.
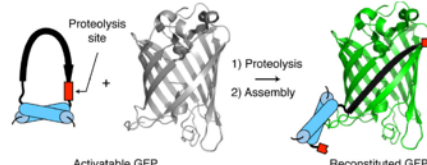
Table 1.   **Commonly applied fusion tags for different experimental purposes.**

Aim	Fusion tags
protein detection, characterization, and purification by western blot, immunoprecipitation, immunofluorescence and affinity chromatography	c-myc, hemagglutinin antigen (HA), FLAG epitopes, 1D4 epitope, polyArg, polyHis and streptavidin binding sites
fluorescent detection and evaluation of the expression	GFP and its variants
enhancing protein expression and/or solubility	glutathione S-transferase (GST), maltose-binding protein (MBP), thioredoxin A (TrxA), small ubiquitin-related modifier (SUMO), N-utilization substance A (NusA), Protein disulfide isomerase I (DsbA) and Mistic
driving expression to inclusion bodies if the expressed protein is sensitive to intracellular degradation or toxic to the host cell.	ketosteroid isomerase (KSI), and TrpΔLE

Despite of the several advantages, fusion tags can occasionally affect the structure or function of the protein of interest for example by inhibiting target protein activity, impacting protein crystallization and by triggering undesirable immune response for proteins in human therapy. Specific proteases such as enterokinase (**Choi et al., 2001**), factor Xa (**Smith and Johnson, 1988**) SUMO protease (**Malakhov et al., 2004**), tobacco etch virus (TEV) protease (**Parks et al., 1994**), thrombin (**Sticha et al., 1997; Vothknecht et al., 1996**), and 3C protease (**Cordingley et al., 1990**) have been effectively used to remove the different carrier proteins or fusion tags after expression and purification. For this, specific cleavage sequences of the different site-specific proteases are included as linkers between the native protein and the fusion tag sequence. Likewise, similarly-structured fusion proteins can be utilized as substrates to assay specific protease activity.

Similar to the oligopeptide-based methods, the vast majority of *in vitro* recombinant protein-based systems rely on fluorescent intensity readout (**Section 2.2.3**). In a few methods a fluorescent moiety is attached to the recombinant substrate molecule either chemically or *via* a fluorescently labelled antibody, however, in most of the cases the source of the detectable fluorescence is a genetically encoded fluorescent protein. By the recent developments in fluorescent protein engineering, wide variety of the fluorescent proteins have become available with several different advantages and properties to best suit the experimental demands (**Shaner et al., 2005**). A few examples for the different fluorescent intensity-based assay platforms are listed in **Table 2**. Beside fluorescent intensity-based formats, assays using recombinant fusion protein substrates based on different readouts have also been developed. Some examples for these platforms are shown in **Table 3**.

Table 2. **Examples for fluorescent intensity-based *in vitro* protease assays using recombinant protein substrates.**

Format	Assay Principle	Substrate type	Examined protease	Reference
Separation-based	Uncleaved substrates and N-terminal products were separated by Ni-NTA coated magentic beads after termination of the reaction by heat-shock.	His_6 – cleavage site – EGFP/GFPuv	thrombin	**Pehrson et al. (1999)**
Heterogenous	Ni-NTA magnetic bead attached substrates were used in the reaction, which was terminated by the removal of the fluorescent-cleavage product containing supernatant.		PSA, thrombin	**Patel et al. (2001)**
			Bacillus lentus alkaline protease	**Chaparro-Riggers et al. (2005)**
Separation-based	Uncleaved substrates were precipitated upon heat-denaturation, while the released GFP remained soluble and after centrifugation its fluorescence was measured	Tus – cleavage site – GFP	trypsin, caspase-3 neutrophil elastase	**Askin at el. (2011)**
Homogenous	Intramolecular FRET pair was separated upon proteolytic cleavage, thereby the FRET signal between the FPs decreases.	BFP – cleavage site – YFP	thrombin, SUMO PR	**Zhang (2004) Shen et al. (2006) Martin et al. (2007) Liu et al. (2012)**
Homogenous		rsGFP – cleavage site – DsRed- Strep Tag II	TEV PR	**Kohl et al. (2002)**
Homogenous	Time resolved- FRET was performed using substrates with chemically attached terbium and the encoded YFP	Terbium ------ full-length ubiquitin– YFP	DUB	**Horton et al. (2007)**
Homogenous	mCherry acted as an acceptor for the emitting fluorescence of the immobilized QD donors	mCherry – cleavage site – His_6 --- [CdSe-ZnS DHLA-QD surface]	Caspase-3	**Boeneman et al. (2009)**
Heterogenous	The energy was transferred upon the catalytic activity of a luciferase donor to a fluorescent acceptor molecule via BRET	CBG – cleavage site – tdTomato	Differen caspases	**Gammon et al. (2009)**
Homogenous	The activity of firefly luciferase (Luc) resulted in yellow-green bioluminescence, that excited mKate via BRET and the fluorescent light emitted by mKate was transferred to the AF680 by FRET (Sequential BRET-FRET = SRET)	His_6 – mKate – cleavage site – Luc | Alexa Fluor 680	Factor Xa, thrombin, caspase-3	**Branchini et al. (2011)**
Homogenous	A GFP molecule was fragmented a into GFP1-10 and GFP11. The GFP11 fragment was distorted and self-assembly of GFP was prevented. However, when proteolytic cleavage occured, GFP11 regained its extended structure allowing the assembly of the two fragments into a functional GFP.		thrombin, caspase-3, Factor Xa, enteropeptidase and HIV-1 PR	**Callahan et al. (2010)**

EGFP =enhanced green fluorescent protein, GFPuv=UV light optimized GFP, BFP=blue emitting fluorescent protein, YFP=Yellow emitting fluorescent protein, rsGFP=red-shifted GFP, FRET=fluorescent resonant energy transfer, BRET=bioluminescent resonant energy transfer, QD=quantum dots, PSA=prostate-specific antigen SUMO=Ubl-specific protease 1, DUB=Deubiquitinating enzyme, HIV-1 PR=Human immunodeficiency virus type 1 protease, TEV PR=Tobacco etch virus protease, CBG=D-luciferin-based click beetle green luciferase

Table 3. **Recombinant substrate-based assay formats using miscellaneous detection principles**. DAL=diaminopropionate ammonia-lyase, QD=quantum dot, AUNP=gold nanoparticle, MBP=maltose binding protein, Luc=luciferase

Format	Principle	Substrate type	Examined protease	Reference
Homogenous	The tetraCys motif allows the site specific labeling with a bis-arsenical fluorophore, and enables detection and quantification via fluorescent intensity, and also allows monitoring of small product release via fluorescence anisotropy detection.	His_6 - CCPGCC – cleavage site – protein of interest S-tag – His_6 – MBP – TEV PR site – – His_6 – CCPGCC – protein of interest	thrombin, Factor Xa, enterokinase, TEV PR	**Blommel and Fox (2005)**
Homogenous	The DAL activity is inhibited by the fusion of the different affinity tags, but upon the removal of the tags by the protease, the activity of DAL is increased and become detectable using a colorimetric assay	His_6MBP – cleavage site –DAL His_6GST – cleavage site –DAL His_6 – cleavage site –DAL	TEV PR, factor Xa, thrombin and DnaB intein	**Zhou et al. (2014)**
Heterogenous	The bioluminescence protein acts as a BRET donor, while QD or AuNP serves as a quenching acceptor. Upon cleavage by the protease, the bioluminescence emission of a mutant luciferase is recovered and can be effectively detected	Luc – cleavage site ---- Au	matrix metalloproteinases	**Kim et al. (2010)**
		Luc – cleavage site ---- QD		**Xia et al. (2008)**

2.2.3. Fluorescent readout

Due to its versatile application, high sensitivity and broad availability, vast majority of the *in vitro* proteolytic assays are based on fluorescence readout. The assay platform developed in this present study was also based on the detection of sample fluorescent signal, therefore, this type of readout is discussed in detail. Fluorescence emission of a sample can be characterized by different parameters including *intensity, polarization, emission spectrum,* and *lifetime*. Any of the listed parameters can be utilized as an assay signal if (i) it is measurable, (ii) its change is highly above the experimental noise, (iii) its change correlates with the biochemical reaction to be tested, and (iv) other uncontrollable processes do not perturb its detection (**Pritz et al., 2011**). The different types of fluorescent readouts are summarized in **Figure 2**.

2.2.3.1. Fluorescent intensity

This readout measures the total amount of the emitted light of the analyzed sample and it excellently fulfills the first three of the above listed requirements. In some cases, assay compounds can interfere with the detection, thereby causing perturbation of the signal. This is principally presented in assays that rely on excitation at relatively short wavelengths (λ_{ex}= ~350 nm) with detection of fluorescence in the blue spectral region (λ_{em}=450–495nm). This interference can often lead to false-positive and false-negative results, however, most of these phenomena, such as compound fluorescence or aggregation, can be avoided by the optimization of the assay in the higher spectral range or can be significantly reduced by using buffer-additives such as non-ionic detergents, respectively (**Thorne et al., 2010**).

2.3. Substrate discovery and structure-based inhibitor design

Proteolytic enzymes are regarded as especially important targets for drug discovery. The development of the therapeutics based on the inhibition of proteolytic activity demands for a lead candidate molecule that can serve as a good basis for the investigation. The lead is frequently derived from the structure of the substrate or a related target molecule. Many of the potential leads are peptide-like molecules with disadvantageous pharmaceutical properties, therefore, their investigation is highly challenging.

Different approaches has been regarded as invaluable tools in the discovery and design of new protease inhibitors for medicine development including the testing of natural product sources like cultures of various species of *Actinomyces* (**Umezawa et al., 1970**) or snake venom

(**Cushman et al., 1973**), HTS-based methods used for substrate discovery (*e.g.* combinatorial chemistry, bacteriophage display, *etc.*) and *in silico* approaches that combines the available structural information with well-developed computer algorithms.

For substrate discovery of the newly described protease species and/or of 'orphan proteases', large number of potential cleavage sequences are to be investigated. As it was described above, *in vitro* protease assays are generally based on the use short peptides or whole proteins (either native or recombinant). In those cases, where the cleavage of short peptide sequences reflects the cleavage properties sufficiently, the following standard approaches are applicable for substrate discovery and optimization. Substrate discovery can be performed by examining standard protein substrates, by testing commercially available substrates of other proteases or by preparing derivatives of known biological targets. The relative importance of each substrate residue can be investigated via alanine-scanning, or by preparing incremental changes at one or more positions within an already defined substrate. Substrate specificity of an enzyme can be efficiently scanned by screening of numerous enzyme and substrate variants, in order to gain deeper understanding regarding the structure and function relationship. Commonly applied method for the generation of sizeable libraries of mutant enzymes is Target Combinatorial Mutagenesis (TCM). Combinatorial or multiplex substrate libraries can either be generated synthetically, or by biological methods. Combinatorial chemistry-based methods for the synthetical generation of substrate libraries and for the screening of the protease-substrate interactions include analyses based on (i) HPLC-MS, (ii) Edman degradation, (iii) acyl transfer reactions, (iv) fluorescently labelled peptide libraries generated and immobilized on to separate solid phase entities (beads, resins or membranes) *e.g.* by 'split and mix' technology and or (v) by positional scanning – synthetic combinatorial libraries (PS-SCL). On the other hand, there are biology-based approaches for the generation of substrate libraries and for testing the specificity of the proteases. These genetic methods include substrate phage display, retroviral display, bacterial display, and yeast α-halo assays (**Richardson, 2002; Diamond, 2007**) Other novel platforms are also available for substrate generation *e.g.* the formation of proteome-derived peptide libraries (**Schilling and Overall, 2008**).

2.4. Aspartic proteases

Aspartic peptidases are named after the Asp residue that acts as ligand of the catalytic water molecule. In most cases a pair of aspartic residues work together to bind and activate the water molecule. All the enzymes involved in this group are endopeptidases and are classified into five clans according to MEROPS database 12.0. Clan AA, more precisely Family A1 is the most abundant group, and its members are found mainly in eukaryotes (**Rawlings and Bateman, 2009**). Most of them have acidic pH preference (*e.g.* pepsin, chymosin, cathepsin D and E), but some of them work at neutral pH (*e.g.* renin) as well. The structure of the molecule is built up by two homologous lobes; one of the lobes has been evolved from the other by gene-duplication. The substrate groove is located between the two lobes, and contains the active site formed by two, homologous Asp residues (each belongs to one lobe). In pepsin A, the active site residues, Asp32 and Asp215, are the part of a highly similar sequence motif of each lobe. The core motif is commonly referred to as the 'DTG motif' (Asp-Thr-Gly), but the full sequence is 'Xaa-Xaa-Asp-Xbb-Gly-Xbb', in which Xaa is a hydrophobic residue, while Xbb can be Ser/Thr. Although the lobes are structurally similar, in the members of Family A1 there is an extra β-hairpin loop, also called as the 'flap' on the N-terminal domain, that covers the active site, and contains Tyr75 and Thr77 (pepsin numbering) as part of the S1 subsite, and play important role in the substrate specificity of the enzyme (**Rawlings and Barrett, 2016**).

Family A2 is another prominent group of Clan AA including peptidases that play a central role in the life-cycle of several retroviruses and retrotransposons. The tertiary structure of the Family A2 proteases is highly similar to that of Family A1. Similar to the pepsin-like proteases, the catalytically active Asp residues are also part of the 'DTG motif', however the amino acid following the Gly residue of the motif is Ala, in contrast to Ser/Thr of most Family A1 enzymes. The substitution to Ala is claimed to be responsible for the switch of activity to neutral pH range from acidic pH preference. For example, renin, classified as a Family A1 protease has Ala in this given position and has a neutral pH preference compared to other acidophilic members of Family A1 bearing Ser/Thr in the same position (**Ido et al., 1991**). Due to its relevance to the current study, further characteristics of aspartic proteases are illustrated on the example of human immunodeficiency virus type 1 retropepsin.

2.4.1. Human immunodeficiency virus type 1 (HIV-1) retropepsin

Inevitably, human immunodeficiency virus (HIV), the causative agent in acquired immunodeficiency syndrome (AIDS) (**Gallo and Montagnier, 1988**) is one of the most well-

known and investigated member of retroviruses. According to the latest UNAIDS report (**2017**) in 2016 approximately 36.7 million people were living with HIV, about 1.8 million people were newly infected and about 1.0 million people died in AIDS-related diseases. The human immunodeficiency virus type-1 (HIV) is classified into the Genus Lentivirus of the Family *Retroviridae* (**Luciw, 1996**). Due to the epidemic significance of HIV, great efforts have been made for investigating the role of the different viral elements and for the development of effective antiretroviral inhibitors, targeting the critical steps in the life-cycle of the virus especially for HIV type-1 (HIV-1). Due to its central role in the maturation and assembly of the virus, HIV-1 retropepsin (EC 3.4.23.16), also called as HIV-1 PR, has become one of the most frequently investigated and targeted viral components.

2.4.1.1. Structure and specificity of HIV-1 PR

The HIV-1 PR is composed of 99 amino acids and its unbounded form it exhibits a 2-fold rotational (C_2) symmetry. It acts as a dimer of two identical subunits with only one active site, which lies across the dimer interface (**Figure 3**). The substrate groove is stabilized by a four-stranded mixed β-sheet of each subunit and by the '-Xaa-Xaa-Asp-Thr-Gly-' motif (where Xaa is a hydrophobic residue) of each monomer. The DTG motif contains the catalytically essential aspartyl residues, Asp-25 and Asp-25', which lie on the bottom of the cavity.
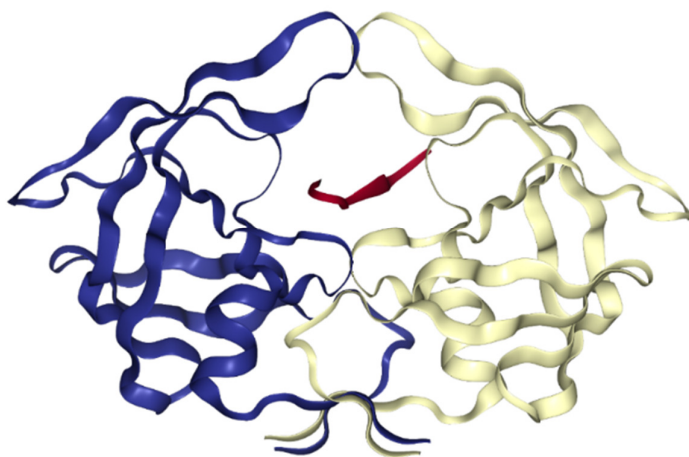
Figure 3.   Crystal structure of HIV-1 PR homodimer. Figure represents structure of MDR769 HIV-1 PR complexed with a QNYPIVQ hepta-peptide substrate. Subunits are colored by blue and wheat, while the peptide substrate by red. Figure was obtained from Protein Data Bank (PDB ID: 3OTS) (**Liu et al., 2011**).

Compared to the cellular aspartic proteases, the active site cleft of which accommodates substrate residues P5-P3', the cleft of retropepsin is slightly shorter and provides subsites only for residues P3-P3' (**Wlodawer and Erickson, 1993**). Each monomer carries an extended β-sheet region forming a glycine-rich loop (Lys[45]-Met-Ile-Gly-Gly-Ile-Gly-Gly-Phe-Ile-Lys[55]), known as the 'flap', that overhangs the substrate groove. The role of the flaps is to control access to the active site. In the ligand-bound form, the flaps are 'closed' thus pulled in toward the active site cavity. In contrast, the ligand-free form exhibits a 'semiopen' conformation, where the flaps are moved away from the DTG motif but still closed over the substrate groove and are in contact with each other (**Hornak et al., 2005**). NMR studies have suggested that the 'closed', the 'semiopen' and the 'fully open' conformations of the enzyme molecule are in dynamic equilibrium, with the 'semiopen' form dominating in the case of the unbounded protease (**Nicholson et al., 1995; Freedberg et al., 2002**). The PR-mediated proteolysis occurs at up to twelve different cleavage sites within the Gag and Gag-Pol polyproteins (**Pettit et al., 1993**) and in Nef (**Freund et al., 1994; Gaedigk-Nitschko et al., 1995**).

Cleavage site sequences are proposed to be divided into two types. In the type-1 cleavage sites, the cleavage occurs between an aromatic (Tyr/Phe) and a Pro residue like in the cases of MA/CA, p6/PR and PR/RT sites, with generally an Asn in the P2 and an Ile in the P2' positions. In contrast in the type-2 cleavage sites the hydrolysis occurs between two hydrophobic residues generally different from each other *e.g.* CA/p2 and NC/p6 (**Salvesen and Rawlings, 2013**). Tőzsér *et al.* (**1997**) have suggested that the described classification may not be generally valid, and it is rather the sequence context that predominantly defines the preference for a residue in a given position.

Crystallography studies suggest that the HIV-1 PR, similarly to other aspartic proteases, binds its substrates through interactions along the extended active site cleft forcing the cleavage site sequence into a β-strand conformation. Because of this extended interaction, oligopeptide sequences must consist of at least 6-8 amino acids to be efficiently processed. The cleavage site sequences do not share sequence homology, however, according to Prabu-Jeyabalan *et al.* (**2002**) natural substrates bounded to the active site cleft adopt a conserved consensus volume that serves as a recognition motif for the protease. This volume is also called as the 'substrate envelope', which at a certain degree also contributes to resistance development of the virus to protease inhibitors (PIs) described below.

2.4.1.2.   Inhibition and mechanism of resistance

Both Family A1 and A2 can be efficiently inhibited by the small molecule inhibitor pepstatin, however, the development of such PIs that selectively inhibits viral enzyme but not the related human enzymes have been required. Due to its central role in the maturation process of HIV-1, the PR has been a major target for developing inhibitors. HIV-1 PIs are essential part of the highly active antiretroviral therapy (HAART). Despite of the fact that several highly effective and selective inhibitors have been designed and proved to be successful in human trials, the emergence of viral forms that resist current inhibitor regimens has remained the most critical factor in the antiretroviral therapeutic failure. This rapid development of resistant viral mutants is the consequence of the heterogeneity and the rapid turnover of the virus coupled with the error-prone nature of the viral RT (**Kantor et al., 2001**) and the selective pressure forced by the therapeutic PIs. Although innumerable attempts have been made to design selective inhibitors with improved effectiveness against resistant viral forms, the development of new inhibitors are still on demand (**Yilmaz, 2017**).

2.5.   Cysteine Proteases

Cysteine proteases can be found in all living organisms (both in cellular and viral levels), and greatly differs from one another from evolutionary point of view. However, as a common catalytic feature, all of them utilize a histidine-activated cysteine residue as part of a catalytic dyad or triad (**Verma et al., 2016; Klein et al., 2018**). As the active site thiol group is very sensitive to oxidation, therefore, for optimal cleavage performance most of these enzymes prefer reducing reaction conditions and/or slightly acidic pH (4.0-6.5) (**Grzonka et al., 2001; Klein et al., 2018**). In order to prevent unwanted digestion, cysteine peptidases are synthetized as zymogens, the structure of which involves a regulatory prodomain and a catalytic mature domain (**Verma et al., 2016**).

The most abundant and investigated groups of cysteine peptidases are Clan CA and Clan CD. Although there are several important members of Clan CA, papain (C1A) isolated from the fruit *Carica papaya* is the most studied. The structure of papain is composed of two main structural domains. The N-terminal domain is formed mainly by α-helices and a long helix runs through the middle of the molecule, while the C-terminal domain composed of a β-barrel (Barrett and Rawlings, 2001). The active site residues Cys25 and His159 are part of the opposite domains and are located in the cleft formed by the domains. The sulfur atom is in the plane of

the imidazole ring of His15, hydrogen-bonded to Asn175, that is analogous to the Asp residue in the catalytic triad of serine peptidases.

Due to their relevance in this current study, in the followings, the characteristics of two cysteine proteases, namely tobacco etch virus NIa protease (TEV PR) and venezuelan equine encephalitis virus nsP2 protease (VEEV nsP2pro) are described in detail. By the processing of the translated viral polyprotein, these proteases are considered as key elements in the life-cycle of positive strand (+) RNA viruses of the picornavirus-like and alphavirus-like superfamily, respectively (**Koonin et al., 2015**).

2.5.1. Tobacco etch virus NIa protease (TEV PR)

TEV is a member of *Potyviridae* family considered to be the most economically important class of plant viruses, as they infect numerous host plants causing enormous agricultural, economical, and biological losses worldwide (**Ivanov et al., 2014**). TEV is known for infecting various species of *Solanaceae* including cultivated tobacco (*Nicotiana tabacum*) (**Revers and Garcia, 2015**).

TEV has a 9.5 kilobase long (+) RNA genome. The translated polyprotein is subsequently processed by three proteases, including P1, HC-Pro and Nuclear Inclusion protein a (NIa) (**Adams et al., 2005; Carrington and Dougherty, 1987a, 1987b**). The NIa protease is also referred as TEV PR or TEVp (EC 3.4.22.44) (**Carrington and Dougherty, 1987b; Dougherty and Parks, 1991**). It is released from the polyprotein by proteolysis and is considered to be critically important for viral maturation, as it has the largest contribution to polyprotein processing (**Figure 4**).
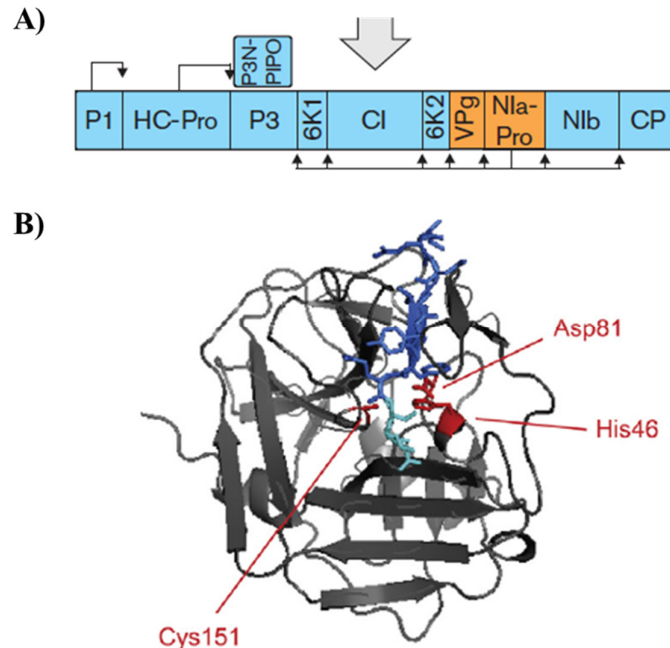
Figure 4. **Important elements of TEV proteome.** A) Structure of the viral polyprotein. Proteolytic cleavage sites in the viral polyprotein of P1, HC-Pro and NIa proteases are shown by arrows (**Ivanov et al., 2014**). B) The structure of bilobal TEV PR. The active site residues (His46, Asp81 and Cys151) are indicated in the substrate groove by the arrows (**Cesaratto et al., 2016**).

2.5.1.1. Structure and specificity of TEV PR

The primary structure of TEV PR is highly similar to that of picornavirus 3C protease and His46, Asp81 and Cys151 together are considered as the catalytic triad. Higher-order structure is confirmed to be built by two anti-parallel β-barrel motifs, which highly resembles structure of trypsin and chymotrypsin (**Nunn et al., 2005; Phan et al., 2002**). Studies of Bazan *et al.* (**1988**) have confirmed that these type of viral cysteine proteases are structurally and functionally homologous to the trypsin-like family of serine protease but utilize a cysteine thiol instead of a serine hydroxyl as the active-site nucleophile.

The specific recognition site of TEV PR can be defined as EXXYXQ↓S/G, where X denotes any type of amino acid. According to Phan et al (**2002**), residues in the position of P6, P4, P3, P2, P1, and P1' directly interact with the substrate binding subsites, and only P5 residue is exposed to the solvent. Consequently, at position of P5, almost any amino acid is tolerated without altering the catalytic efficiency of the enzyme (**Dougherty et al., 1989**), however, results of several mutagenesis studies indicate that TEV PR highly prefers substrates that contain Gln in P1, Phe in P2, Tyr in P3 and Leu in P4 (**Dougherty et al., 1989; Phan et al.,**

2002; Tőzsér et al., 2005; Yi et al., 2013). The optimal cleavage site has been defined as ENLYFQ↓S/G, which is actually the wild-type cleavage site between NIb and CP of TEV polyprotein (**Boulware et al., 2010; Kostallas et al., 2011**). Regarding P1' position, the proteolytic activity is retained at the presence of any amino acids, except Pro. This characteristic fundamentally distinguishes TEV PR from other highly specific viral proteases like rhinovirus 3C protease and Tobacco vein mottling virus (TVMV) NIa protease, that do not share this property (**Sun et al., 2010; Waugh, 2011**), which practically means that TEV PR can release almost every protein fused downstream to ENLYFQ↓S/G, leaving the desired amino acid at the N-terminus.

Additionally, TEV PR possesses a self-cleavage motif between residues 213–219 (GHKVFM↓S), which highly resembles the canonical cleavage site. It has been demonstrated that after self-cleavage (truncated at residue 218), the TEV PR retains its activity, however, its catalytic activity and affinity for the substrate is reduced, as cleaved C-terminal peptide (219–242 residues) remains tightly bound to the substrate pocket *via* residues 235–242 (**Nunn et al., 2005; Cesaratto et al., 2015**). It has been shown that by the substitution of Ser219 with Val improves flexibility of the catalytic core and facilitate the formation of a C-terminal helix that avoids self-cleavage and hence retains full proteolytic activity (**Kapust et al., 2001; Wang et al., 2013**).

2.5.1.2. *In vitro* biotechnological applications

TEV PR is one of the best studied and most extensively utilized proteases for biotechnological applications, due to its advantageous properties: i) relatively high *in vitro* activity rate, ii) tight specificity, iii) resistance against many protease inhibitors (*e.g.* PMSF, AEBSF, TLCK, Bestatin, pepstatin A, EDTA, and E-64) (**Dougherty et al., 1989**), iv) it can be successfully applied at wide range of enzyme/substrate concentrations without off-target cleavage activities, and v) it is active at several reaction conditions and cleavage buffers (**Sun et al., 2012, Parks et al., 1995**).

TEV PR is widely used for affinity tag removal, while different TEV PR variants were used efficiently for on-column cleavage on different type of resins (**Zhu et al., 2016**) as the immobilized protein was found to be active after storage or multiple applications (**Puhl et al., 2009, Miladi et al., 2012**).

2.5.2. Venezuelan equine encephalitis virus non-structural protein 2 protease

Venezuelan equine encephalitis virus (VEEV) is a member of *Alphavirus* genus of the *Togaviridae* family and causes human and livestock disease mainly in Central and South America. The virus infection occurs *via* mosquitos to birds, horses, rodents, humans and to other vertebrates. The virus is especially dangerous for equine, with an average fatality rate of 20-80 % (**Deardorff et al., 2009; Zacks and Paessler, 2010**). In humans, acute VEEV infections are resolved by innate and adaptive immune responses, however, about 14 % of the infected humans develop neurological symptoms and approximately 1 % of the infection causes lethal *encephalitis* (**Zacks and Paessler, 2010; Johnson and Martin, 1974**). The inhaled virus can enter the brain *via* the olfactory neurons and as its viral particles are unusually resistant to desiccation and can be stably freeze-dried and aerosolized, VEEV has been regarded as potential biological weapon, that have been reportedly developed by several nations including the US and former Soviet Union (**Hu et al., 2016; Bronze et al., 2002**).

Alphaviruses, including VEEV, are small spherical enveloped viruses, with ~70 nm diameter size. The viral single strand positive sense RNA genome is packaged in an icosahedral capsid, surrounded by a host cell-derived lipid envelope containing spikes formed by viral transmembrane glycoproteins (**Rupp et al., 2015**). The ~11 kilobase long viral mRNA consists of two cistrons: one which corresponds to two-third of the genome is translated into a single polyprotein nsP123 or nsP1234, comprising the non-structural proteins, while the second open reading frame (ORF) is coding for the structural proteins that form the viral particles translated in the late stage of the infection. Non-structural proteins play a pivotal role in the formation of the viral replication complex and are released from the nsP123 and nsP1234 polyprotein upon the proteolytic activity of nsP2 (EC 3.4.22.-) (**Figure 5**). Additionally, nsP2 protease is supposed to play an essential function in antagonizing the host interferon response *via* processing proteins related to the production of innate immune response by recognizing short stretches of homologous host-pathogen protein sequences (SSHHPS) (**Morazzani et al. 2019**). Mechanisms aiming to neutralize early events of host innate immune response are connected to the proteolytic activity of viral nsPs are proposed to be of importance in other Group IV viruses including *Picornaviridae*, *Flaviviridae* and *Coronaviridae.* Recently, coronaviruses have been exposed to extremely intensive characterization, since a number of novel pathogens contributing to recent global pandemia have been emerging from this family *e.g.* middle east respiratory syndrome- and severe acute respiratory syndrome-associated coronavirus (MERS-

and SARS-CoV, respectively), and SARS-CoV-2. *Coronaviridae* and *Togaviridae* show resemblance in their mechanism of action, and accordingly Compton *et al.* (**2017**) have proposed structural similarities between VEEV nsP2 and coronaviral papain-like proteases/deubiquitinases of the SARS- and MERS-CoV.
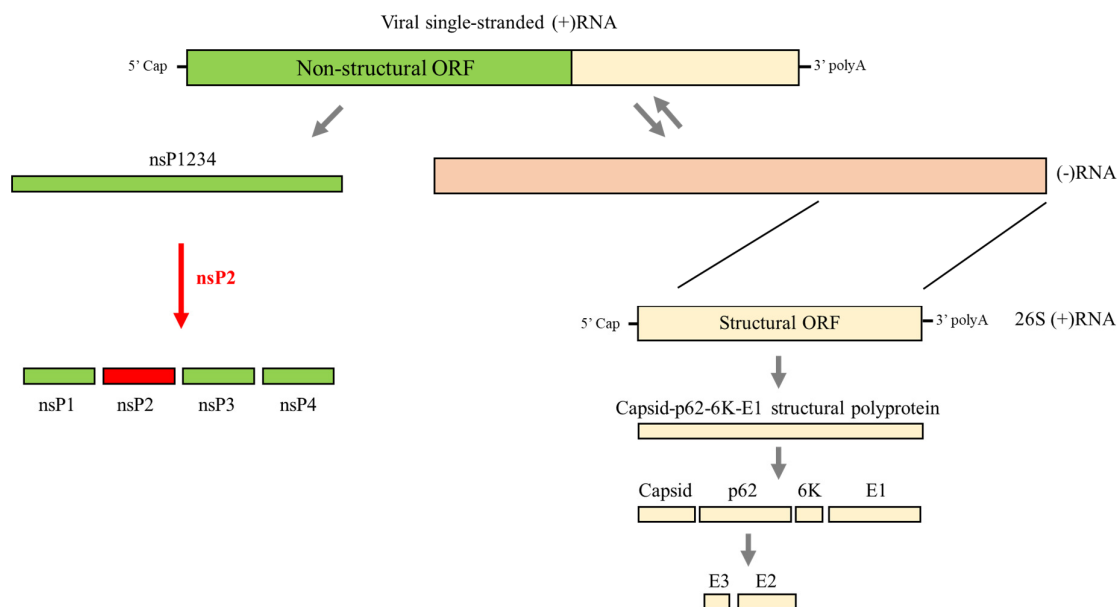
Figure 5. **The schematic illustration of the genome and proteome organization of alphaviruses.** Non-structural protein 2 (nsP2), marked in red is responsible for the proteolytic processing of the viral non-structural polyprotein (nsP1234) **(**Figure was prepared based on **Waarts et al., 2004).**

2.5.2.1. Structural characteristics

The non-structural protein 2 (nsP2) is considered as the central element of VEEV life-cycle and has multiple enzyme activities. Its N-terminal region from Gly1-Ile456 has been associated with ATP-ase and GTP-ase activity (**Rikkonen et al., 1994**), RNA helicase activity (**Gomez de Cedron et al., 1999**), and RNA 5'-triphosphatase activity (**Vasiljeva et al., 2000**), while the C-terminal region between Met457 and Cys794 is considered to regulate 26S subgenomic RNA synthesis (**Suopanki et al., 1998**), to downregulate minus-strand RNA synthesis late in infection (**Sawicki et al., 2006; Sawicki and Sawicki, 1993**), to target nsP2 for nuclear transport (**Peranen et al., 1990**), and last but not at least, according to its proteolytic activity, it processes the alphavirus nonstructural polyprotein replication complex (**Vasiljeva et al., 2001, 2003**). The C-terminal region of nsP2, frequently referred to as the nsP2pro, consists of main two structural regions: a papain-like cysteine protease domain linked to an S-

adenosyl-L-methionine-dependent RNA methyltransferase domain (SAM MTase), the function of which has not been clarified (**Figure 6**).
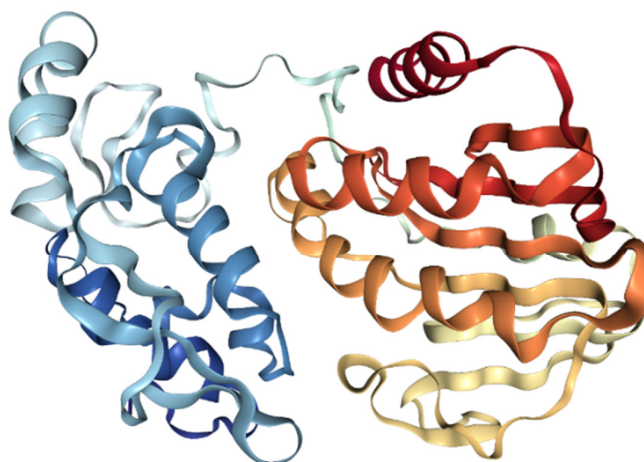
Figure 6. **The structure of VEEV nsP2pro.** Crystal structure of the papain-like cysteine protease domain of VEEV nsP2 protease is colored by blue, while the S-adenosyl-L-methionine-dependent RNA methyltransferase domain (SAM MTase) by orange **(Russo et al., 2006).** The figure was obtained from Protein Data Bank (PDB ID: 2HWK).

In the MEROPS database VEEV nsP2pro is the member of Clan CN. Its active site is formed by a catalytic dyad including residues Cys477 and His546, while the substrate binding cleft is considered to be positioned adjacent to the interface between the protease and SAM methyltransferase domains. Most of the residues that contribute to substrate binding are located in the N-terminal protease domain, however, Hu *et al.* (**2016**) have confirmed that at least three residues of the SAM MTase domain including Arg662, Lys705 and Lys706 are involved in the recognition of substrate and in the regulation of substrate binding cleft structure (**Russo et al., 2006**; **Hu et al., 2016**).

2.5.2.2. Specificity and structure of substrate binding pockets

In alphaviruses the nsP2pro processes the non-structural polyprotein at three different sites located between nsP1/nsP2; nsP2/nsP3 and nsP3/nsP4. All these sites share common sequential properties as it is depicted on **Figure 7**, which aligns the cleavage site sequences of 10 alphaviruses. This alignment reveals that the residue in the P2 position is Gly in all three sites, the residue in P1 position is always an amino acid with a small side chain. Additionally, the relatively low number of residues occurring in P3 position suggest the highly conserved

nature of this site too. In contrast, residues in P4 position shows greater variability, which proposes that the major recognition signal for cleavage may be displayed by residues P3-P2-P1. Alignment result indicate that residues in P1'-P4' positions show similar pattern only within the same cleavage sites, which suggest that these residues are likely to be determined by their function in the protein rather than by a requirement for cleavage site recognition.

| | nsP1/nsP2 | | | | ↓ | | | | | nsP2/nsP3 | | | | ↓ | | | | | nsP3/nsP4 | | | | ↓ | | | |
	P4	P3	P2	P1		P1'	P2'	P3'	P4'	P4	P3	P2	P1		P1'	P2'	P3'	P4'	P4	P3	P2	P1		P1'	P2'	P3'	P4'
SIN	D	I	G	A		A	L	V	E	G	V	G	A		A	P	S	Y	G	V	G	G		Y	I	F	S
AURA	D	A	G	A		A	L	V	E	G	S	G	A		A	P	S	Y	G	V	G	G		Y	I	F	S
WHA	D	I	G	A		A	L	V	E	G	V	G	A		A	P	S	Y	G	V	G	G		Y	I	F	S
VEE	E	A	G	A		G	S	V	E	E	A	G	C		A	P	S	Y	D	A	G	A		Y	I	F	S
EEE	E	A	G	A		G	S	V	E	E	A	G	R		A	P	A	Y	E	A	G	A		Y	I	F	S
WEE	E	A	G	A		G	S	V	E	E	A	G	R		A	P	A	Y	E	A	G	A		Y	I	F	S
ONN	R	A	G	A		G	I	V	E	R	A	G	C		A	P	S	Y	R	A	G	G		Y	I	F	S
SF	H	A	G	A		G	V	V	E	T	A	G	C		A	P	S	Y	R	A	G	A		Y	I	F	S
RR	R	A	G	A		G	V	V	E	T	A	G	C		A	P	S	Y	R	A	G	A		Y	I	F	S
MID	R	A	G	A		G	V	V	N	T	A	G	C		A	P	S	Y	R	A	G	A		Y	I	F	S

Figure 7.   The alignment of the non-structural protein-cleavage site sequences of 10 alphaviruses. Residues for which at least 7 out of the 10 viruses have identical amino acids are boxed (**Strauss and Strauss, 1994**).

Russo *et al.* (**2006**) have described the S1, S2, and S3 substrate binding pockets of VEEV nsP2pro as shallow depressions on the protein surface, lining a long, deep groove formed at the interface between the protease and the SAM MTase domains. According to their findings, S1 pocket is formed by Val476, Asn475, and Ala509 residues, that are thought either to be highly conserved or to have only conservative substitutions, while backbone amides from Cys477 and Val476 likely form the oxyanion hole. The S2 subsite is mainly defined by the highly conserved Trp547 that is located next to the catalytic His546 residue and, together with the fact that the corresponding P2 site contains a highly conserved Gly motif, this structure of VEEV nsP2pro highly resembles to the so called 'glycine specificity motif' of other cysteine proteases (**Golubtsov et al. 2006**). The S3 binding pocket is localized on the SAM MTase domain, is built by Ile698 and Met702 and is flanked by highly conserved Ala509 and His510.

The alphaviral nsP2 proteases play crucial role in the viral lifecycle and were found to contribute to the neutralization of early innate immune response of the host, hence they are validated targets of antiviral drug development (**Strauss et al., 1992; Reichert et al.,2009**). Furthermore, as these enzymes pose some degree of sequence specificity, they have been also

considered as potential biotechnological tools for the *in vitro* removal of affinity tags from recombinant proteins.

The available literature on the characterization of nsP2pro *in vivo* and *in vitro* enzyme activity and specificity has been performed mainly in the context of sindbis virus (SINV) and Semliki forest virus (SFV), however, large differences in the cleavage efficiencies of the three sites were observed across the different *in vitro* and *in vivo* studies using either artificial or native polyprotein as substrates. In case of VEEV nsP2pro only limited data available regarding its specificity, however, the crystal structure of both the free and the E-64d bound form has been solved (**Russo et al., 2006; Hu et al., 2016**).

Previously our laboratory investigated the activity of three alphaviral proteases for their utilization as tools for affinity tag removal in collaboration with Macromolecular Crystallography Laboratory (MCL), Center for Cancer Research, National Cancer Institute (NCI) (Frederick, MD, USA). The results of the study were published by Zhang *et al.* (**2009**). The nsP2pro domains of SIN, SFV and VEEV were generated by using bacterial expression system, and the activity of these proteases were investigated on different fusion protein substrates and on artificial oligopeptides substrates designed based on the P6-P6' residues of the three corresponding natural recognition sites. As the recognition site of the three enzyme is highly similar, beside investigating the activity of the proteases on their cognate substrates, their cross-reactivity was also tested. Kinetic parameters on oligopeptide substrates could be determined for only three enzyme-substrate pairs including VEEV nsP2pro on SFV p1/p2 substrate and SFV nsP2pro on its cognate SFV p1/p2 and p3/p4 cleavage site. The corresponding k_{cat}/K_M values were determined as 0.028 ± 0.005 mM^{-1} s^{-1}, 0.036 ± 0.008 mM^{-1} s^{-1} and 0.277 ± 0.050 mM^{-1} s^{-1}, respectively, and were considered as substantially lower than the catalytic efficiency of TEV and TVMV proteases. Furthermore, their ability to process fusion protein substrates was also relatively poor, therefore, it was concluded that despite of the fact that alphavirus proteases theoretically exhibit sufficient sequence specificity to be useful reagents for affinity tags removal, in practice their *in vitro* cleavage properties, at least in their investigated form, could not offer any advantage over the already successfully utilized potyviral proteases.

3. Scope of the study

Several excellent biological and chemically-based methods have been engineered to investigate the different proteolytic activities and to find and/or optimize substrates for the newly discovered proteases. However, many of them are relatively insensitive, expensive, time-consuming, and labor-intensive, some of them are applicable only for a particular group of proteases or provide incomplete coverage and no or only limited data on reaction kinetics (**Kostallas et al., 2011**). Accordingly, the development of HTS-compatible and quantitative assay formats offering new advantages and/or integrating the already existing ones of other methods is still on demand. Our aim was to optimize a quantitative, HTS-compatible, versatile fluorescence-based proteolytic assay platform using an interchangeable recombinant fusion protein substrate system, that can be applicable to both academic and industrial environments. The assay was aimed to be optimized by using HIV-1 and TEV PRs and to be applied for mapping the specificity and activity of newly designed, VEEV PR constructs, in which the sequence of C-terminal protease domain would be extended by (i) the whole N-terminal domain, further referred to as VEEV nsP2 and (ii) Ala436-Met457 region of VEEV nsP2, named as nsP2pro-2.

Aim 1: Development of recombinant fusion protein substrate system.
1) Generation of His_6-MBP-FP 'empty' expression vectors and plasmids coding for the naturally occurring or modified protease cleavage sites.
2) Optimization of the expression and purification of the recombinant protein substrates.

Aim 2: Development of a recombinant fusion protein substrate-based protease assay
1) Development, optimization and application of a nickel-nitrilotriacetic acid (Ni-NTA) magnetic bead-based assay system, by enzymatic reactions of HIV-1 and TEV PRs.
2) Optimization of a protocol for the in-gel renaturation of denatured protein substrates.
3) Providing written and video step-by-step protocols for the applications of the Ni-NTA magnetic bead-based protease assay

Aim 3: Investigation of VEEV PR specificity
1) Cloning, expression, and purification of i) VEEV nsP2 and (ii) VEEV nsP2pro-2 proteins, fused to His_6 and to His_6-MBP affinity tags.
2) Cloning, expression and purification of recombinant mEYFP-fused substrates comprising wild-type and modified alphaviral protease cleavage site sequences.
3) Enzymatic reactions for the investigation of VEEV PR specificity.

4. Materials and methods

All materials were purchased from Sigma-Aldrich, otherwise is indicated. A stock solution of HIV-1 PR for this work was purified by Ferenc Tóth as described by Bozóki *et al.* (**2018**), while stock solutions of TEV (S219V) PR were a kind gift of David S. Waugh (NCI-Frederick, USA) that have been prepared by the method of Kapust and its co-workers (**Kapust et al., 2001**).

4.1. Generation of the 'empty' expression vector for the substrates

pDON221 and pDEST-His$_6$-MBP vectors were kind gifts of dr. David S. Waugh (NCI-Frederick, USA), while plasmids coding for the different fluorescent proteins were purchased from Addgene with the exception of pmEYFP-N1 (**Table 4**). pmEYFP-N1 plasmid was produced in our laboratory from EYFP-N1, which was a kind gift of Prof. Thomas Jovin (Göttingen). The modification included the elimination of the EYFP dimerization surface by introducing A208K mutation into the protein structure (**Zacharias et al. 2002**)

pDest-His$_6$-MBP-FP plasmids listed in **Table 4** were prepared by Gateway Cloning Technology (Thermo Fischer Scientific, Invitrogen) based on Tropea *et al.* (**2007**). The linear DNA sequences to be transferred via pDON221 donor vector into pDEST-His$_6$-MBP (**Tropea et al., 2007**) were amplified by a two-step polymerase chain reaction (PCR).

Table 4. **List of the produced 'empty' expression vectors and the corresponding templates applied in the first PCR reaction.** ECFP=enhanced cyan fluorescent protein, mEGFP=monomeric enhanced green fluorescent protein, mEYFP=monomeric enhanced yellow fluorescent protein. wt=wild-type.

Generated empty expression vector	Applied Template
pDEST-His$_6$-MBP-ECFP	pECFP-N1
pDEST-His$_6$-MBP-mTurquiose2	pPalmitoyl-mTurquoise2
pDEST-His$_6$-MBP-mEGFP	pmEGFP-1
pDEST-His$_6$-MBP-GFP(wt)	pGFP-1
pDEST-His$_6$-MBP-mEYFP	pmEYFP-N1
pDEST-His$_6$-MBP-mApple	pmApple-N1
pDEST-His$_6$-MBP-mCherry	pFPV-mCherry

Gateway cloning steps were carried out according to the manufacturer's protocol. Briefly, in the first step, the coding sequences of the desired fluorescent proteins were amplified by a PCR reaction (100 µL final volume) using 300-300 ng N1 and C primers (5'-GAGAACCTGTACTTCCAGGGTGGTGCATCTACACGCGGTTTAATTAATCTTCTGCT

AGCATGGTGAGCAAGGGCG-3') and C primers (5'-GGGGACACACTTTGTACAAG-AAAGCTGGGTTATTACTTGTACAGCTCGTCCAT-3'), and 1800 ng template plasmid (**Table 4**). Reaction mixtures contained *Pfu* DNA polymerase in *Pfu* buffer + $MgSO_4$ (10x) (Thermo Fischer Scientific, Fermentas), 10 mM dNTP mix (Biorad), and nuclease-free water was added to set final volume to 100 µL. The following PCR protocol was used: cycle1 (1x): 5 min at 94 °C; cycle2 (30x): 1 min at 94 °C, 1 min at 45 °C, and 3 min at 72 °C; cycle3 (1x): store at 4 °C. The PCR products were separated by gel electrophoresis using 1 % agarose gel. The linear PCR products (having ~800 base pair lengths) were cut out from the gel and purified by Qiagen DNA extraction kit (Qiagen). 25 ng of the purified DNA was used as a template in the second PCR reaction (100 µL final volume) containing 300 ng N2 (5'-GGGGACAAGTTTGTACAAAGCAGGCTCGGAGAACCTGTACTTCCAG-3') and C primers, and *Pfu* buffer + $MgSO_4$ (10x), *Pfu* DNA polymerase, 10 mM dNTP mix, and nuclease-free water was added to set final volume to 100 µL. The PCR protocol was the same as in the case of the first-step reaction. The PCR products were purified by Qiagen Nucleotide Removal Kit (Qiagen). 300 ng of the purified linear DNA sequence flanked by *attB* sites was used in a BP reaction (recombination between *att**B*** and *att**P*** sites) with pDON221. Thereafter, LR reaction (recombination between *att**L*** and *att**R*** sites) was performed for 4 hour using 300 ng of the purified entry clone and 800 ng pDEST-His$_6$-MBP destination vector. Gateway cloning steps and the following transformations were carried out according to the manufacturer's protocol.

4.2.   Generation of substrate coding dsDNA by random mutagenesis

pT7-Blue-3 plasmid (Novagen) was cleaved by *BamHI* and *NheI* (New England Biolabs, NEB). Cleavage products were separated by 1 % agarose gel electrophoresis, and the linear plasmids were cut out from the gel and purified by Qiagen DNA extraction kit (Qiagen). Oligonucleotide primers, coding wild-type cleavage site between the nsP1 and nsP2 of SFV (SFV-1, EYHAGA↓GVVETP) and flanked by *BamHI* and *NheI* ends, were inserted into linearized pT7-Blue-3 plasmid. For ligation, 40 ng linearized pT7- Blue-3 plasmid was incubated with 800 ng SFV-1 *BamHI* forward primer (5'-GATCCTTAATTAAAGAGTACCATGCTGGTGCTGGTGTGGTGGAGACACCGG-3') and 800 ng SFV-1 *BamHI* reverse primer (5'-CTAGCCGGTGTCTCCACCACACCAGCA-CCAGCATGGTACTCTTTAATTAAG-3') for 2 minutes at 65 °C, then for 2 minutes at 4 °C. Hereafter, *T4* DNA Ligase Reaction Buffer (10x) and *T4* DNA Ligase (NEB) were added,

followed by incubation overnight at 16 °C. The final volume of the ligation mixture was 20 µL. 100 µL of *E. coli* DH5α competent cells (NEB) were transformed by 5 µl of the ligation reaction mixture by heat shock (at 42 °C). The transformation reaction was spread on Luria-Bertani (LB) agar plates containing 100 µg/mL ampicillin. Grown colonies were cultured at 37 °C for overnight in LB medium containing 100 µg/mL ampicillin followed by plasmid preparation using Qiaprep Spin Mini Prep Kit (Qiagen). Successful ligation was assessed by cleaving the purified pT7-Blue-3-SFV-1 plasmid by *NheI* and *PacI* (NEB). Random mutagenesis of P1' residues in the wild-type SFV-1 was performed by QuickChange Lightning Multi-Site Directed Mutagenesis Kit (Stratagene). The reaction contained: 50 ng pT7-Blue3-SFV-1 plasmid, 100 ng SFV-1 P1' DEG oligonucleotide primer (5'-GAGTACCATGCTGGTGCTNNNGTGGTGGAGACACCGGCTAGC-3'), Quick Change Buffer (10x), 10 mM dNTP mix, 1 µL Quick Change enzyme and nuclease-free water to 25 µL. The following PCR protocol was used: cycle1 (1x): 2 min at 95 °C; cycle2 (30x): 20 sec at 95 °C, 30 sec at 55 °C, and 2 min at 65 °C; cycle3 (1x): 5 min at 65 °C; cycle4 (1x): store at 4 °C. After the reaction, methylated DNA was digested by *DpnI* for 5 min at 37°C, then the 45 µL XL10-Gold ultracompetent cells were transformed by 1.5 µL *DpnI*-treated PCR reaction according to the manufacturer's instruction. Plasmids were prepared from the cultures of grown colonies using Qiagen Plasmid Midi Kit (Qiagen). The mutagenized DNA sequences of SFV-1 cleavage sites were cut out of the purified plasmids using *NheI* and *PacI* (NEB) and separated on polyacrylamide gels containing 15 % urea. After excision from the gel, the short, mutagenized fragments were purified by Qiaex II Gel Extraction Kit (Qiagen).

4.3.   Generation of substrate coding-expression vectors by ligation

Circular pDest-His$_6$-MBP-FP plasmids were cleaved by *PacI* and *NheI* (NEB). Cleavage products were separated on 1 % agarose gel, and the linear plasmids were cut out from the gel and purified by Qiagen DNA extraction kit (Qiagen). Hereafter, the short dsDNAs coding for the cleavage sites of interest were ligated into the linearized plasmids:

i.   If the DNA to be inserted had been generated by random mutagenesis, 200 ng linearized pDest-His$_6$-MBP-FP plasmid was mixed with 100 ng mutagenized dsDNA fragment.

ii. In those cases, where chemically synthetized complement oligonucleotides primers were applied, annealing of the complement oligonucleotides coding for the cleavage sites to be inserted was required. Annealing was performed by mixing 200 ng forward and 200 ng reverse primers with 150 ng purified linearized pDEST-His$_6$-MBP-FP

plasmid. For the list of the chemically synthetized complement oligonucleotide primers corresponding to the generated cleavage site see **Table 5**.

In both cases the volume was adjusted to 16.5 µL with nuclease-free water and the reaction mixtures were incubated at 65 °C for 2 minutes then at 4 °C for 2 minutes. Thereafter, T4 DNA Ligase Reaction Buffer (10x) and T4 DNA Ligase (NEB) were added, and the reaction mixture was incubated at 25 °C for 1 hour followed by incubation at 16 °C overnight. Next day 100 µL of *E. coli* BL21(DE3) competent cells (Thermo Fischer Scientific, Invitrogen) were transformed by the ligation reaction mixture (15 µL) using heat shock, followed by plasmid preparation from the cultures of grown colonies. DNA sequences of the purified plasmids were verified by capillary DNA sequencing using pDEST-His$_6$-MBP-FP sequencing forward (5'-GATGAAGCCCTGAAAGACGCGCAG-3') and/or reverse primer (5'-GCAAGGCGATTAAGTTGGGTAACGC-3').

Table 5. **List of the chemically synthetized oligonucleotide primers.** The primers were applied for the generation of substrates for A) HIV-1 PR and B) VEEV PR. Positions that are different from the wild-type residues are marked in red. FWD=forward, REV=reverse.

A) HIV-1 PR:

Coded cleavage site sequence	Applied oligonucleotide primers
VSQNY↓PIVQ	FWD:5'-TAAAGTGAGCCAGAACTATCCGATTGTGCAGG-3'
	REV:5'–CTAGCCTGCACAATCGGATAGTTCTGGCTCACTTTAAT–3'
VSQ LY↓PIVQ	FWD:5'-TAAAGTGAGCCAGCTGTATCCGATTGTGCAGG-3'
	REV:5'–CTAGCCTGCACAATCGGATACAGCTGGCTCACTTTAAT–3'
KARVL↓AEAM	FWD:5'-TAAAAAAGCACGTGTGCTGGCAGAAGCAATGG–3'
	REV:5'–CTAGCCATTGCTTCTGCCAGCACACGTGCTTTTTTAAT–3'
DTGHSNQVSQNY↓PIVQNIQGQMVH	FWD:5'-TAAAGATACCGGCCATAGCAACCAGGTGAGCCAGAACTTC-CGATTGTGCAGAACATTCAGGGCCAGATGGTGCATG-3'
	REV:5'-CTAGCATGCACCATCTGGCCCTGAATGTTCTGCACAATCGG-ATAGTTCTGGCTCACCTGGTTGCTATGGCCGGTATCTTTAAT–3'
GVGGPGHKARVLAEAMSQVTNSAT	FWD:5'-TAAAGGCGTGGGCGGCCCGGGCCATAAAGCACGTGTGCTG-GCAGAAG-CAATGAGCCAGGTGACCAACAGCGCAACCG–3'
	REV:5'-CTAGCGGTTGCGCTGTTGGTCACCTGGCTCATTGCTTCTGCC-AGCACAC-GTGCTTTATGGCCCGGGCCGCCCACGCCTTTAAT–3'
GVGGPGH VSQNYPIVQN QVTNSAT	FWD:5'-TAAAGGCGTGGGCGGCCCGGGCCATGTGAGCCAGAACTATC-CGATTG-TGCAGAACCAGGTGACCAACAGCGCAACCG-3'
	REV:5'-CTAGCGGTTGCGCTGTTGGTCACCTGGTTCTGCACAATCGGA-TAGTTCTGGCTCACATGGCCCGGGCCGCCCACGCCTTTAAT–3'
DTGHSNQ KARVLAEAMS IQGQMVH	FWD:5'-TAAAGATACCGGCCATAGCAACCAGAAAGCACGTGTGCTGG-CAGAAGCAATGAGCATTCAGGGCCAGATGGTGCATG–3'
	REV:5'-CTAGCATGCACCATCTGGCCCTGAATGCTCATTGCTTCTG-CCAGCACACGTGCTTTCTGGTTGCTATGGCCGGTATCTTTAAT–3'

34

B) VEEV PR

Cleavage site sequence	Oligonucleotide primers
EYHAGA↓GVVETP	FWD:5'-TAAAGAATATCATGCTGGTGCTGGTGTTGTTGAAACACCGG-3'
	REV:5'–CTAGCCGGTGTTTCAACAACACCAGCACCAGCATGATATTCTTTAAT-3'
EYHAGA↓YVVETP	FWD:5'-TAAAGAATATCATGCTGGTGCTTATGTTGTTGAAACACCGG-3'
	REV:5'–CTAGCCGGTGTTTCAACAACATAAGCACCAGCATGATATTCTTTAAT-3'
EYHAGA↓IVVETP	FWD:5'-TAAAGAATATCATGCTGGTGCTATTGTTGTTGAAACACCGG-3'
	REV:5'–CTAGCCGGTGTTTCAACAACAATAGCACCAGCATGATATTCTTTAAT-3'
EYHAGA↓QVVETP	FWD:5'-TAAAGAATATCATGCTGGTGCTCAGGTTGTTGAAACACCGG-3'
	REV:5'–CTAGCCGGTGTTTCAACAACCTGAGCACCAGCATGATATTCTTTAAT-3'
EYHAGA↓NVVETP	FWD:5'-TAAAGAATATCATGCTGGTGCTAACGTTGTTGAAACACCGG-3'
	REV:5'–CTAGCCGGTGTTTCAACAACGTTAGCACCAGCATGATATTCTTTAAT-3'
EYHAGA↓DVVETP	FWD:5'-TAAAGAATATCATGCTGGTGCTGATGTTGTTGAAACACCGG-3'
	REV:5'–CTAGCCGGTGTTTCAACAACATCAGCACCAGCATGATATTCTTTAAT-3'
EYHAGA↓AVVETP	FWD:5'-TAAAGAATATCATGCTGGTGCTGCGGTTGTTGAAACACCGG-3'
	REV:5'–CTAGCCGGTGTTTCAACAACCGCAGCACCAGCATGATATTCTTTAAT-3'
EYHAGA↓MVVETP	FWD:5'-TAAAGAATATCATGCTGGTGCTATGGTTGTTGAAACACCGG-3'
	REV:5'–CTAGCCGGTGTTTCAACAACCATAGCACCAGCATGATATTCTTTAAT-3'
EYHAGA↓WVVETP	FWD:5'-TAAAGAATATCATGCTGGTGCTTGGGTTGTTGAAACACCGG-3'
	REV:5'–CTAGCCGGTGTTTCAACAACCCAAGCACCAGCATGATATTCTTTAAT-3'
EYHAGA↓FVVETP	FWD:5'-TAAAGAATATCATGCTGGTGCTTTTGTTGTTGAAACACCGG-3'
	REV:5'–CTAGCCGGTGTTTCAACAACAAAAGCACCAGCATGATATTCTTTAAT-3'
EQHAGA↓GVVETP	FWD:5'-TAAAGAACAGCATGCTGGTGCTGGTGTTGTTGAAACACCGG-3'
	REV:5'–CTAGCCGGTGTTTCAACAACACCAGCACCAGCATGCTGTTCTTTAAT-3'
EYEAGA↓GVVETP	FWD:5'-TAAAGAATATGAAGCTGGTGCTGGTGTTGTTGAAACACCGG-3'
	REV:5'–CTAGCCGGTGTTTCAACAACACCAGCACCAGCTTCATATTCTTTAAT-3'
EYTAGA↓GVVETP	FWD:5'-TAAAGAATATACCGCTGGTGCTGGTGTTGTTGAAACACCGG-3'
	REV:5'–CTAGCCGGTGTTTCAACAACACCAGCACCAGCGGTATATTCTTTAAT-3'
EYRAGA↓GVVETP	FWD:5'-TAAAGAATATCGCGCTGGTGCTGGTGTTGTTGAAACACCGG-3'
	REV:5'–CTAGCCGGTGTTTCAACAACACCAGCACCAGCGCGATATTCTTTAAT-3'
EYGAGA↓GVVETP	FWD:5'-TAAAGAATATGGTGCTGGTGCTGGTGTTGTTGAAACACCGG-3'
	REV:5'–CTAGCCGGTGTTTCAACAACACCAGCACCAGCACCATATTCTTTAAT-3'
EYHAAA↓GVVETP	FWD:5'-TAAAGAATATCATGCTGCTGCTGGTGTTGTTGAAACACCGG-3'
	REV:5'–CTAGCCGGTGTTTCAACAACACCAGCAGCAGCATGATATTCTTTAAT-3'
EYHAVA↓GVVETP	FWD:5'-TAAAGAATATCATGCTGTGGCTGGTGTTGTTGAAACACCGG-3'
	REV:5'–CTAGCCGGTGTTTCAACAACACCAGCCACAGCATGATATTCTTTAAT-3'
EYHAGG↓GVVETP	FWD:5'-TAAAGAATATCATGCTGGTGGTGGTGTTGTTGAAACACCGG-3'
	REV:5'–CTAGCCGGTGTTTCAACAACACCACCACCAGCATGATATTCTTTAAT-3'
EYHAGV↓GVVETP	FWD:5'-TAAAGAATATCATGCTGGTGTTGGTGTTGTTGAAACACCGG-3'
	REV:5'–CTAGCCGGTGTTTCAACAACACCAACACCAGCATGATATTCTTTAAT-3'
EYHAGA↓GPVETP	FWD:5'-TAAAGAATATCATGCTGGTGCTGGTCCGGTTGAAACACCGG-3'
	REV:5'–CTAGCCGGTGTTTCAACCGGACCAGCACCAGCATGATATTCTTTAAT-3'
EYHAGA↓GSVETP	FWD:5'-TAAAGAATATCATGCTGGTGCTGGTAGCGTTGAAACACCGG-3'
	REV:5'–CTAGCCGGTGTTTCAACCGTACCAGCACCAGCATGATATTCTTTAAT-3'
LGRAGA↓YIFSSD	FWD:5'- TAAACTGGGTCGTGCTGGTGCTTATATTTTTTCTTCTGATG-3'
	REV:5'- CTAGCATCAGAAGAAAAAATATAAGCACCAGCACGACCCAGTTTAAT-3'

4.4. Expression of HIV-1 PR substrates and cell lysis

Small-scale overnight starter cultures were initiated by adding 10 µL freezer stocks of *E. coli* BL21(DE3) cells, transformed previously by the expression plasmids, to 5 mL LB medium in a 50 mL centrifuge tube and incubated overnight at 37 °C. Next day 5 mL of the starter culture was added to 50 mL LB containing 100 µg/mL ampicillin in a 500 mL

Erlenmeyer flask. Cells were grown at 37 °C up to an absorbance of 0.6-0.8 at 600 nm. Protein expression was induced by the addition of 1 mM isopropyl β-d-1-thiogalactopyranoside (IPTG) followed by incubation for 3 hours at 37 °C. Cell suspensions were aliquoted into 50 mL centrifuge tubes (25 mL/tube) and were harvested by centrifugation at 4,000 g for 15 minutes at 4 °C (Jouan CR 412). Cell pellets were stored at least for 1 hour at -70 °C before lysis.

For cell lysis, pellets were thawed on ice for 15 minutes and suspended in 2 mL lysis buffer (50 mM sodium-acetate, 300 mM NaCl, 10 mM imidazole, 0.05 % Tween 20, pH 8.0) containing PMSF at a final concentration of 25 µg/mL. Lysozyme and DNase enzymes (NEB) were added in 1 mg/mL and 10 U/mL final concentrations, respectively. Cell suspensions were vortexed and occasionally mixed during incubation on ice for 10 minutes, then were aliquoted into microcentrifuge tubes and sonicated for 3 minutes. Tubes were centrifuged at 10,000 g for 15-20 minutes (Eppendorf 5415D) at room temperature, and the supernatants (cleared bacterial lysate) were collected.

4.5. Small-scale substrate expression for VEEV nsP2pro-2 specificity study

Starter cultures were initiated as described in **Section 0.** Next day, 2.5 mL of the starter culture was added to 15 mL LB containing 100 µg/mL ampicillin in a 50 mL centrifuge tube. Cells were grown at 37 °C up to an absorbance of 0.5–0.6 at 600 nm. Protein expression was induced by the addition of 1 mM IPTG followed by incubation for 4 hour at 37 °C. Cells were harvested by centrifugation at 4,000 g for 15 min at 4 °C (Jouan CR 412). Cell pellets were stored at −70 °C at least for 1 h. The crude cell lysate was prepared as described in **Section 0**, but the volume of the applied lysis buffer was reduced to 1 mL.

4.6. Substrate expression for VEEV nsP2pro-2 kinetic study

For kinetic measurements, substrates representing the wild-type, P4-Glu, P4-Thr, P4-Arg, P4-Gly, P1-Gly, P1'-Thr, and P2'-Ser SFV-1 cleavage site sequences were applied (the sequences of the cleavage sites are shown in **Table 9**). Expression of the substrates were performed as described in **Section 4.4**, except that the cells were grown at 37 °C up to an absorbance of 0.5–0.6 at 600 nm when induction by ITPG was performed. The induced culture was incubated for 2 hours at 37°C, when protein translation was arrested by addition of tetracycline in 200 µg/mL final concentration. Hereafter the culture was further incubated at 37 °C for 2 hours until harvest.

4.7. Purification of the substrates

Cleared bacterial lysates containing the desired recombinant fusion substrate were added to Ni-NTA magnetic agarose beads (purchased either from Qiagen or Cube Biotech). The suspensions were incubated for 20-60 min at room temperature while continuously rotating. Substrate-attached magnetic beads (SAMBs) were washed three times by 1 % Tween 20 (pH 7.0), three times by washing buffer (50 mM NaH_2PO_4, 300 mM NaCl, 5 mM imidazole, 0.05 % Tween 20, pH 7.0) and three times by cleavage buffer (50 mM NaH_2PO_4, 300 mM NaCl, 0.05 % Tween 20, pH 7.0) using Dynamag^{TM}-2 magnetic particle concentrator (MPC) (Thermo Fischer Scientific, Invitrogen). Hereafter, substrates were eluted from the beads using either elution buffer A (100 mM EDTA, 0.05 % Tween 20, pH 8.0) used in **Section 4.12.1 - 4.12.4**, or elution buffer B (50 mM NaH_2PO_4, 300 mM NaCl, 500 mM imidazole, 0.05 % Tween 20, pH 8.0) used in **Section 4.12.5 - 4.12.7**. Buffer was exchanged to cleavage buffer by using 10K Amicon tubes (Merck-Millipore). Total protein concentrations were determined by BCA protein assay (Thermo Fischer Scientific, Pierce) and/or by measuring absorbance at 280 nm by NanoDrop 2000 equipment (Thermo Fischer Scientific) using the theoretical extinction coefficients and molecular weight calculated based on the primary structure of the proteins by ProtParam tool of ExPASy (**Gasteiger et al., 2005**).

4.8. Calibration of the fluorescent substrates

The purified substrates dissolved in either elution buffer A (**Section 4.12.1 - 4.12.4**) or elution buffer B (**Section 4.12.5 - 4.12.7)** or cleavage buffer, were serially diluted and each dilution points were transferred into black half-area plates. Relative fluorescent intensities were measured using the parameters listed in **Table 6**. Blank-corrected relative fluorescent intensity values (RFU) were plotted against the substrate concentration (mM). Linear regression was performed, and the parameters of the fitted lines were determined by using Microsoft Excel 2010 (Microsoft).

Table 6. List of the applied excitation and emission wavelength for the detection of the different fluorescent proteins in case of the applied devices.

Fluorescent protein	Excitation (nm)	Emission (nm)	Fluorimeter
mApple	590/35	645/40	Biotek Synergy2
	560	595	Biotek Synergy H1
mTurquoise2	400/10	460/40	Biotek Synergy2
	434	474	Biotek Synergy H1
mEYFP	544/15	590/10	Victor^2 Wallac 1420
	510	540	Biotek Synergy H1
mCherry	585	615	Biotek Synergy H1

4.9. Cloning of VEEV nsP2 and VEEV nsP2pro-2 expression plasmids

Plasmid encoding the cDNA of nonstructural proteins of VEEV were a gift from Dr. Christine L. Pugh (United States Army Research Institute of Infectious Diseases). Plasmids coding for the different VEEV PR constructs were prepared by Gateway Cloning Technology (Thermo Fischer Scientific, Invitrogen) based on Tropea *et al.* (**2007**). The linear DNA sequences that were transferred via pDON221 donor vector into pDEST™ 17 (further referred to as pDEST-His$_6$) and pDEST-His$_6$-MBP (**Tropea et al., 2007**) had been previously amplified according to the following protocol: (i) The ORF of VEEV nsP2 was amplified using PE2685 and PE2686 primers for PCR according to the following reaction setups: 10 ng template, 5 nmol PE2685, 5 nmol PE2686, 10 µL Phusion™ Flash High –Fidelity PCR Master Mix (Finnzyme, NEB) and nuclease-free water up to 20 µL using the following PCR protocol: cycle1 (1x): 60 s at 98 °C; cycle2 (30x): 15 s at 98 °C, 75 s at 72 °C, and 60 s at 72 °C; cycle3 (1x): store at 4 °C. (ii) The ORF of VEEV nsP2pro-2 was amplified using PE2687 and PE2686 primers for PCR, according to the following reaction setups: 10 ng template, 5 nmol PE2685, 5 nmol PE2686, 10 µL Phusion™ Flash High –Fidelity PCR Master Mix (Finnzyme, NEB) and nuclease-free water up to 20 µL using the following PCR protocol: cycle1 (1x): 60 s at 98 °C; cycle2 (30x): 10 s at 98 °C, 30 s at 72 °C, and 60 s at 72 °C; cycle3 (1x): store at 4 °C. The PCR products were purified by MinElute (Qiagen) and 60-80 ng of the resulting PCR amplicons were subsequently used as the templates for another PCR with 5 nmol PE277, 5 nmol PE2686, 10 µL Phusion™ Flash High –Fidelity PCR Master Mix (Finnzyme, NEB) and nuclease-free water up to 20 µL using the following PCR protocol (i) for VEEV nsP2: cycle1 (1x): 60 s at 98 °C; cycle2 (30x): 15 s at 98 °C, 75 s at 72 °C, and 60 s at 72 °C; cycle3 (1x): store at 4 °C; (ii) for VEEV nsP2pro-2: cycle1 (1x): 60 s at 98 °C; cycle2 (30x): 10 s at 98 °C, 30 s at 72 °C, and 60 s at 72 °C; cycle3 (1x): store at 4 °C. The PCR products were separated by gel electrophoresis using 1 % agarose gel. The linear PCR products were excised out from the gel and purified by MinElute (Qiagen) to recover DNA from the bands. BP reactions were performed using 150 ng purified linear DNA sequence flanked by *attB* sites, 150 ng pDON221, 2 µL Gateway® BP Clonase™ II Enzyme Mix (Invitrogen, Thermo Ficher Scientific) and TE buffer (10 mM Tris, 1 mM EDTA, pH 8.0) up to 10 µL. The reaction was incubated at 25 °C for 1 hour. Hereafter 50 µL electro-competent *E. coli* DH5α cells were electroporated at 1.5 kV with 2 µL of the BP reaction. Subsequently, 450 µL super optimal broth (SOC) medium was added to the cells, incubated at 37 °C with 250 rpm shaking for 1 hour and plated to LB agar plates containing kanamycin for clone selection. Entry vectors of the selected clones were isolated by Qiagen

miniprep kit (Qiagen). The sequence the of entry clones was verified by capillary sequencing. The applied oligonucleotide primers applied are listed in **Table 7**.

Table 7. **Applied oligonucleotide primers for the generation and sequencing of expression plasmids coding for VEEV nsP2 and nsP2pro-2.** PE2285, PE2286, PE2287, and PE2288 primers were used to sequence the C-terminal, while PE2730 and PE2731 were applied to sequence the N-terminal domain of VEEV nsP2. Primer N2, N1, and C are referring to the nomenclature used by **Tropea et al., 2007**.

Code	Description	Sequence
PE2685	Primer N1 for VEEV nsP2	5'-GGCTCGGAGAACCTGTACTTCCAGGGCTCAGTGGA-GACACCTCGTG-3'
PE2687	Primer N1 for VEEV nsP2pro-2	5'-GGCTCGGAGAACCTGTACTTCCAGGCCAAGTACCC-TGGGAATTTCACT-3'
PE2686	Primer C	5'-GGGGACCACTTTGTACAAGAAAGCTGGGTTATTAT-GTATAAATGTTGGTCAAGGTTGATGAAAGC-3'
PE277	Primer N2	5'-GGGGACAAGTTTGTACAAAAAAGCAGGCTCGGAGA-ACCTGTACTTCCAG-3'
PE2285	FWD sequencing primer	5'- ATGCGTGAGGTTCTTTGGAC-3'
PE2286	FWD sequencing primer	5'-CAGCAGTGTGAAGACCATGC-3'
PE2287	REV sequencing primer	5'-AGGACAGTTCTGCCCTTCAA-3'
PE2288	REV sequencing primer	5'-TTGGTCAAGGTTGATGAAAGC-3'
PE2730	FWD sequencing primer	5'-CCTTCCATGAATTCGCCTAC-3'
PE2731	FWD sequencing primer	5'-AACGACGAATCCGAAAGAGA-3'

Thereafter, LR reaction was performed using 150 ng of the purified entry clones, 150 ng destination vector (pDEST-His$_6$-MBP or pDEST-His$_6$,), 2 µL Gateway™ LR Clonase™ II Enzyme Mix (Invitrogen, Thermo Fischer Scientific) and TE buffer up to 10 µL. The reaction was incubated at 25 °C for 1 hour. Afterwards, 50 µL electro-competent *E. coli* Rozetta cells were electroporated at 1.5 kV with 2 µL of the LR reaction. Subsequently 450 µL SOC medium was added to the cells, incubated at 37 °C with 250 rpm shaking for 1 hour and plated to LB agar plates containing ampicillin and chloramphenicol for clone selection. Expression vectors pBB2546, pBB2547, pBB2549, and pBB2550 from the selected and enriched clones were isolated by Qiagen Miniprep Kit (Qiagen) (**Table 8**).

Table 8. **The list of the generated expression vectors coding for the different VEEV nsP2pro-2 and nsP2 constructs.** The plasmid codes are shown based on the in-house nomenclature system of MCL (Frederick), and it represents the initials of the maker and the number of the so-far generated plasmid.

Plasmid code	Description
pBB2546	VEEV nsP2 in pDEST-His$_6$
pBB2547	VEEV nsP2pro-2 in pDEST-His$_6$
pBB2549	VEEV nsP2 in pDEST-His$_6$-MBP
pBB2550	VEEV nsP2pro-2 in pDEST-His$_6$-MBP

4.10. Expression of VEEV nsP2pro-2 and VEEV nsP2

Small-scale overnight starter cultures were initiated by adding 100 µL freezer stocks of pBB2546, pBB2547, pBB2549 and pBB2550 plasmids in *E. coli* Rozetta cells to 100 mL LB medium and incubated overnight at 37 °C. Next day 30 mL of the starter culture was added to 1000 mL LB medium containing 125 µg/mL ampicillin, 30 µg/mL chloramphenicol and 0.2 % glucose. Cells were incubated at 37 °C up to an absorbance of 0.5–0.6 at 600 nm. Hereafter expression was induced by the addition of 5 mL of 200 mM IPTG to the flasks and incubated for 4 hours at 30 °C with 250 rpm shaking. After the incubation, cells were centrifuged at 5000 rpm for 10 min using a tabletop centrifuge. After the removal of the medium the cell pellets were stored at -80 °C.

Additionally, to improve the solubility of VEEV nsP2, expression settings different from described above were also tried in case of pBB2549 plasmid in *E. coli* Rozetta. In this setup, the only difference was compared to the above described parameters that when the 1000 mL culture reached an absorbance of 0.5 at 600 nm, the temperature was shifted to 18 °C, and the culture was further incubated for 20 min. Hereafter expression was induced by 5 mL of 200 mM IPTG and culture was incubated overnight at 18 °C with 250 rpm shaking.

In order to compare the protein expression profile of the different constructs, cultures were sampled immediately before the addition of IPTG (t=0), and also right before harvesting the cells (t=4, or t=overnight). One mL from each sample was spun at 10,000 g for 5 min on a tabletop centrifuge, the supernatants were discarded, and the pellets were stored at -80 °C. Next day, after thawing the pellets were re-suspended in distilled water. For setting approximately equal cell concentration, the volume of the added water was set based on the absorbance values at 600 nm measured at the time of sampling. In order to assess the solubility profile of the expressed proteins, 10 mL of each cell suspensions right before harvesting the cells were removed and centrifuged at 5,000 g for 10 min in a tabletop centrifuge. The supernatants were discarded, and the pellets were stored at -80 °C. Next day, after thawing, cells were lysed using B-PER buffer (prepared by mixing 1.5 mL B-PER™ (Pierce™, Thermo Fischer Scientific) reagent with 3 µL lysozyme (2 mg/mL) and 3 µL DNase (NEB). For the lysis of the cells 4 µL B-PER buffer was used per 1 mg cell pellet and the suspension was incubated for 15 min at room temperature. Total protein fraction was analyzed from samples taken from the crude cell lysates, while samples for analyzing the soluble protein fraction was taken from the cleared lysate after the centrifugation of the crude suspension at 5,000 g for 10 min on a tabletop centrifuge.

Samples were prepared for analysis using 2x loading buffer containing β-mercaptoethanol (β-ME) as reducing agent and were heated at 90 °C for 5 min and were analyzed by sodium-dodecyl-sulfate polyacrylamide gel electrophoresis (SDS-PAGE).

4.11. Purification of VEEV nsP2pro-2

Bacterial cells (~8 g) containing the His$_6$-MBP-VEEV nsP2pro-2 were suspended in 50 mL of ice-cold buffer A (50 mM sodium phosphate, 150 mM NaCl, 25 mM imidazole, pH 7.5). The cells were lysed with an APV Gaulin Model G homogenizer (Invensys, Albertslund, Denmark) at 10,000 psi and centrifuged at 30,000 g for 30 min at 4 °C. The supernatant was filtered through a 0.22 μm polyethersulfone membrane, applied to a 15 mL (3 x 5 mL) HisTrap FF crude affinity column (GE Healthcare, Piscataway, NJ) equilibrated in buffer A, and then eluted with a linear gradient from 5-50 % buffer B (50 mM sodium phosphate, 150 mM NaCl, 500 mM imidazole, pH 7.5). Fractions of the eluate (10 mL each) containing the His$_6$-MBP-nsP2pro-2 were analyzed on reducing SDS-PAGE. Fractions containing high amount the fusion protein were pooled and concentrated using an Amicon YM30 membrane (Millipore, Billerica, MA), then diluted 6- fold with 50 mM sodium phosphate buffer (containing 150 mM NaCl, pH 7.5) to reduce the imidazole concentration to approximately 25 mM. The fusion protein was then digested with a 5 mg/mL stock solution of the His-tagged TEV PR (70:1 v/v) overnight at 4 °C. Next, the products of the digest were applied to a 20 mL (4 x 5 mL) HisTrap FF crude affinity column equilibrated with buffer A. Both flow-through, containing the nsP2pro protease, and eluate fractions were collected. Fractions of the flowthrough were analyzed on reducing SDS-PAGE and fractions containing high amount of VEEV nsP2pro-2 were pooled. The sample was concentrated to about 10 mg/mL using an Amicon YM30 membrane (Millipore) and applied to a HiPrep 26/60 Sephacryl S200 column (GE Healthcare) equilibrated with 50 mM Tris-HCl buffer (containing 150 mM NaCl, 2 mM Tris (2-carboxyethyl) phosphine hydrochloride (TCEP), pH 7.5). The peak fractions (5 mL each) containing the nsP2pro protease were analyzed on reducing SDS-PAGE, pooled and concentrated to 1-5 mg/mL. The concentration of the purified and pooled VEEV nsP2pro-2 was determined as 2.9 mg/mL by A=280 method. The theoretical extinction coefficient and molecular weight were calculated based on the primary structure of the proteins by ProtParam tool of ExPASy (**Gasteiger et al., 2005**). Aliquots were flash-frozen with liquid nitrogen and stored at -80 °C until further use.

4.12. Ni-NTA magnetic bead-based assays

A detailed, standardized written protocol and a visual demonstration have been prepared to support the execution of the different applications of the developed Ni-NTA magnetic agarose bead-based assay platform (**Bozóki et al., 2018**), the video protocol is supplemented to the thesis.

SAMBs were generated as it was described in **Section 4.7**. SAMBs were dissolved in cleavage buffer in order to generate the SAMB stock solution that is used for generation of the assay samples. The volume of the cleavage buffer was dependent on the individual experimental design and was calculated based on the density of magnetic bead, on the number and volume of samples to be assayed.

For time course, inhibitory and pH-dependence experiments, equal amounts of homogenous SAMB suspensions, while for substrate-dependent kinetic measurements, increasing amounts of the homogenous SAMB suspensions were measured into 2.0 mL Protein Lobind Micro-centrifuge tubes (Eppendorf). The tubes were applied to an MPC, the supernatant was removed, and the beads were suspended in equal volume of cleavage buffer. After the addition of the enzyme/enzyme buffer or elution buffer A (**Section 4.12.1 - 4.12.4**) or elution buffer B (**Section 4.12.5 - 4.12.7)**, the final volume of the reaction buffer was 70 µL. For determining the amount of the substrate on the bead surface, substrate control (C) samples were also prepared in the same way as the reaction samples but SAMBs were suspended in elution buffer A (**Section 4.12.1 - 4.12.4**) or elution buffer B (**Section 4.12.5 - 4.12.7**), instead of the cleavage buffer. Parallel to the reaction (R) samples, substrate blank (B) samples, where enzyme buffer was added to the SAMBs instead of the enzyme were prepared. Reactions were terminated by separating the magnetic beads on MPC and the fluorescence of the supernatants were measured as described in **Section 4.8**. The amount of C-terminal fluorescent cleavage product in the reaction samples were calculated by dividing the blank corrected RFU values by the slope of the cleavage-buffer-based calibration curve. The concentration of the eluted substrate (mM) in the supernatants of the C sample was calculated by dividing their corrected RFU values by the slope of the elution buffer-based calibration curve. For calculating the molar concentration of the substrates in each R sample, the substrate concentration (in mM) of the SAMB stock solution used for creating the assay samples could be determined according to the followings:

$$c_{SAMB} \ (mM) = \frac{c_C \ (mM) \ x \ V_r \ (\mu L)}{V_{SAMB}(\mu L)}$$

where, c_{SAMB} is the molar concentration of the SAMB stock solution; c_C is the molar concentration of the eluted substrate in the C sample; V_r is the volume of the reaction mixture after the addition of the reaction buffer and the enzyme buffer; and V_{SAMB} is the volume of the SAMB stock solution in the C sample. The SD of the duplicate values was calculated by Microsoft Excel (2010) according to the following equation:

$$SD = \sqrt{\frac{\sum (x-\bar{x})^2}{(n-1)}}$$

4.12.1. Time course kinetic studies of HIV-1 PR

Cleavage of His$_6$-MBP-VSQNY↓PIVQ-mTurquoise2 and His$_6$-MBP-VSQNY↓PIVQ-mApple substrates (at 1.6 µM and 2.4 µM final concentrations, respectively) by HIV-1 PR (at 36.4 nM final concentration) was followed by measuring fluorescence after 0, 20, 40, and 60 minutes of incubation at 37 °C, while continuously shaking at 600 rpm. Molar concentration of the C-terminal fluorescent cleavage product in the reaction samples were plotted against time (min). Linear regression was performed, and the parameters of the fitted lines were determined by GraphPad Prism version 5.00.

4.12.2. Inhibition of HIV-1 PR

His$_6$-MBP-VSQNY↓PIVQ-mTurquoise2 recombinant substrate (0.028 mM final concentration) was used to study inhibitory effect of amprenavir (ranging from 1 nM to 1 µM final total concentrations) on HIV-1 PR (at 36.4 nM final concentration). The assay samples were incubated for 60 minutes at 37 °C, while continuously shaking at 600 rpm. Initial velocity values (nMs^{-1}) were plotted against the logarithms of amprenavir concentrations (nM). Inhibitory effect at 50 % initial velocity value (IC$_{50}$) was determined by fitting five parameter logistic curve on the data using GraphPad Prism version 5.00 based on Richards equation, which is also referred to the following model:

$$LogX_b = LogEC50 + \left(\frac{1}{Hill\ slope}\right) * Log\left(\left(2^{\frac{1}{S}}\right) - 1\right)$$

$$Numerator = Top - Bottom$$

$$Denominator = (1 + 10^{(LogX_b-X)*Hill\ slope})\wedge S$$

$$Y = bottom + \left(\frac{Numerator}{Denominator}\right)$$

where Bottom and Top are the plateaus at each end of the curve, in the same units as Y. LogEC50 is the concentrations that give half-maximal effects, in the same units as X. Hill Slope is the unitless slope factor. S is the unitless symmetry parameter. LogXb, is called as the. inflection point.

In order to determine active site concentration linear regression was fitted on the velocity values plotted against the amprenavir concentration (nM) in the low nanomolar range, where the R^2 value of the fitted regression was >0.9. The active enzyme concentration (6.05 nM) was determined at y=0 of the fitted line and further used for assessing kinetic parameters and inhibitory constant. The inhibitory constant (K_i) was calculated from the IC_{50} value using the following equation:

$$K_i = \frac{IC_{50} - \frac{[E]}{2}}{1 + \frac{[S]}{K_m}}$$

where [E] and [S] are the concentrations of active enzyme and substrate, respectively, while K_M is the Michaelis constant (**Morrison, 1969**).

4.12.3. Substrate-dependent kinetic studies of HIV-1 PR

Cleavage reactions were performed by using 36.4 nM HIV-1 PR (final, total enzyme concentration) on increasing amount of His_6-MBP-VSQNY↓PIVQ-mTurquoise2 and His_6-MBP-VSQNY↓PIVQ-mApple substrates. The assay samples were incubated for 60 minutes at 37 °C while continuously shaking at 600 rpm. Initial velocity values (nMs^{-1}) were calculated by dividing the calculated C-terminal fluorescent cleavage product (nM) by the incubation time (s) and were plotted against the initial substrate concentration (mM). Kinetic parameters were determined at <20 % substrate turnover by Michaelis-Menten (**1913**) non-linear regression analysis using GraphPad Prism version 5.00 for Windows (GraphPad Software, La Jolla, California USA, www.graphpad.com) according to the following model:

$$Y = \frac{V_{max} * X}{K_M + X}$$

where V_{max} is the maximum enzyme velocity in the same units as Y, and K_M is the substrate concentration needed to achieve a half-maximum enzyme velocity also referred to as the Michaelis-Menten constant, in the same units as X. The 'error' of V_{max} and K_M, were calculated by the software. The value of k_{cat} was calculated as follows:

$$k_{cat} = \frac{V_{max}}{[E]}$$

where [E] is the concentration of the enzyme or (if applicable) active enzyme. The k_{cat}/K_M (Q) and its 'error' (q) was calculated according to Fenner (**1931**):

$$Q \pm q = \frac{A \pm a}{B \pm b} = \frac{A}{B} \pm \frac{1}{B^2}\sqrt{B^2 a^2 + A^2 b^2}$$

where A ± a represents k_{cat} and its error, while B ± b represents K_M and its error respectively.

4.12.4. Measurements with TEV PR

The substrate-dependent kinetic measurements were performed as described in **Section 4.12.3**, by incubating the reactions with TEV PR (45.7 nM final concentration) at 30 °C. To study dependence of enzyme activity on pH, TEV PR (91.4 nM final concentration) was incubated with His$_6$-MBP-VSQNY↓PIVQ-mTurquoise2 recombinant substrate (at 0.03 mM final concentration) in cleavage buffer. The pH of the cleavage buffer was set to 6.0, 6.5, 7.0, 7.5, 8.0, and 8.5. The obtained initial velocity values (nM^{-1}s^{-1}) were plotted against pH. Significance of the difference between the initial velocity values of the enzyme at pH 6.5-8.5 range and the highest one measured at pH 7.0 were determined by using GraphPad unpaired t-test (http://graphpad.com/quickcalcs/ttest2/). If p-value < 0.05, the difference was considered to be statistically significant.

4.12.5. Microplate-based specificity studies of VEEV nsP2pro-2

The specificity of VEEV nsP2pro-2 was studied by mEYFP-fused recombinant substrates coding either for the wild-type or for the modified SFV-1 cleavage site sequences listed in **Table 9**. For screening, a 96-well plate–based adaptation of the Ni-NTA magnetic bead-based assay platform was used, which is illustrated in **Figure 8**.

Table 9.   Cleavage site sequences examined in the VEEV nsP2pro-2 specificity studies. Residues that are different from the wild-type are marked in red.

Sequence	Variant	Sequence	Variant	Sequence	Variant
EYHAGA↓GVVETP	wild-type	EYHAGA↓YVVETP	P1'-Tyr	EQHAGA↓GVVETP	P5-Gln
EYHAGA↓TVVETP	P1'-Thr	EYHAGA↓IVVETP	P1'-Ile	EYEAGA↓GVVETP	P4-Glu
EYHAGA↓VVVETP	P1'-Val	EYHAGA↓QVVETP	P1'-Gln	EYTAGA↓GVVETP	P4-Thr
EYHAGA↓PVVETP	P1'-Pro	EYHAGA↓NVVETP	P1'-Asn	EYRAGA↓GVVETP	P4-Arg
EYHAGA↓HVVETP	P1'-His	EYHAGA↓DVVETP	P1'-Asp	EYGAGA↓GVVETP	P4-Gly
EYHAGA↓CVVETP	P1'-Cys	EYHAGA↓AVVETP	P1'-Ala	EYHAAA↓GVVETP	P2-Ala
EYHAGA↓SVVETP	P1'-Ser	EYHAGA↓MVVETP	P1'-Met	EYHAVA↓GVVETP	P2-Val
EYHAGA↓RVVETP	P1'-Arg	EYHAGA↓WVVETP	P1'-Trp	EYHAGG↓GVVETP	P1-Gly
EYHAGA↓LVVETP	P1'-Leu	EYHAGA↓FVVETP	P1'-Phe	EYHAGV↓GVVETP	P1-Val
EYHAGA↓KVVETP	P1'-Lys	EYHAGA↓FVVETP	P1'-Phe	EYHAGA↓GPVETP	P2'-Pro
EYHAGA↓EVVETP	P1'-Glu			EYHAGA↓GSVETP	P2'-Ser

For screening of substrates, VEEV nsP2pro-2 (2.1-6.0 µM final concentration) was incubated with His$_6$-MBP-mEYFP substrates (1-5 µM final concentration) at 30 °C while continuously shaking at 600 rpm. Incubation time was set to be 40 minutes in case of P5, P4, P1, and P2' variants, while 20-22 hours for P1'-modified substrates.

The final volume of the assay samples was 70 µL per well. The pH of the applied buffers was 7.5. During the experiments flat-bottom 96 well plates (Greiner BioONE or Qiagen) were applied in combination with 96-Well Magnet Type A magnetic particle concentrator (Qiagen). While incubation and washing of the beads, plates were shaken by a digital shaker (IKA MS3). The detection of the fluorescence, and the determination of the concentration of the products and substrates were carried out as it is described in **Section 4.8** and **Section 4.12**., respectively. Cleavage preferences of VEEV nsP2pro-2 on the different substrates was compared based on the substrate conversion (%) calculated according to the followings:

$$substrate\ conversion\ (\%) = \frac{\text{fluorescent product conc. (mM)}}{\text{initial substrate conc. (mM)}} * 100$$

Significance of the differences between the conversion rates of the P5, P4, P2, P1, and P2'varinats and the wild-type substrate were analyzed (i) by GraphPad unpaired t-test (http://graphpad.com/quickcalcs/ttest2/) and also (ii) by one-way analysis of variances (ANOVA) using Microsoft Excel 2010 (Microsoft). The difference was considered to be statistically significant if (i) p-value < 0.05 and (ii) $F < F_{critical}$.

Figure 8. **Schematic representation of the 96-well plate adapted workflow of Ni-NTA magnetic bead-based assay platform.** All the indicated volumes are for a single well. Three different assay sample types are illustrated as C=substrate control, B=substrate blank, and R=reaction.
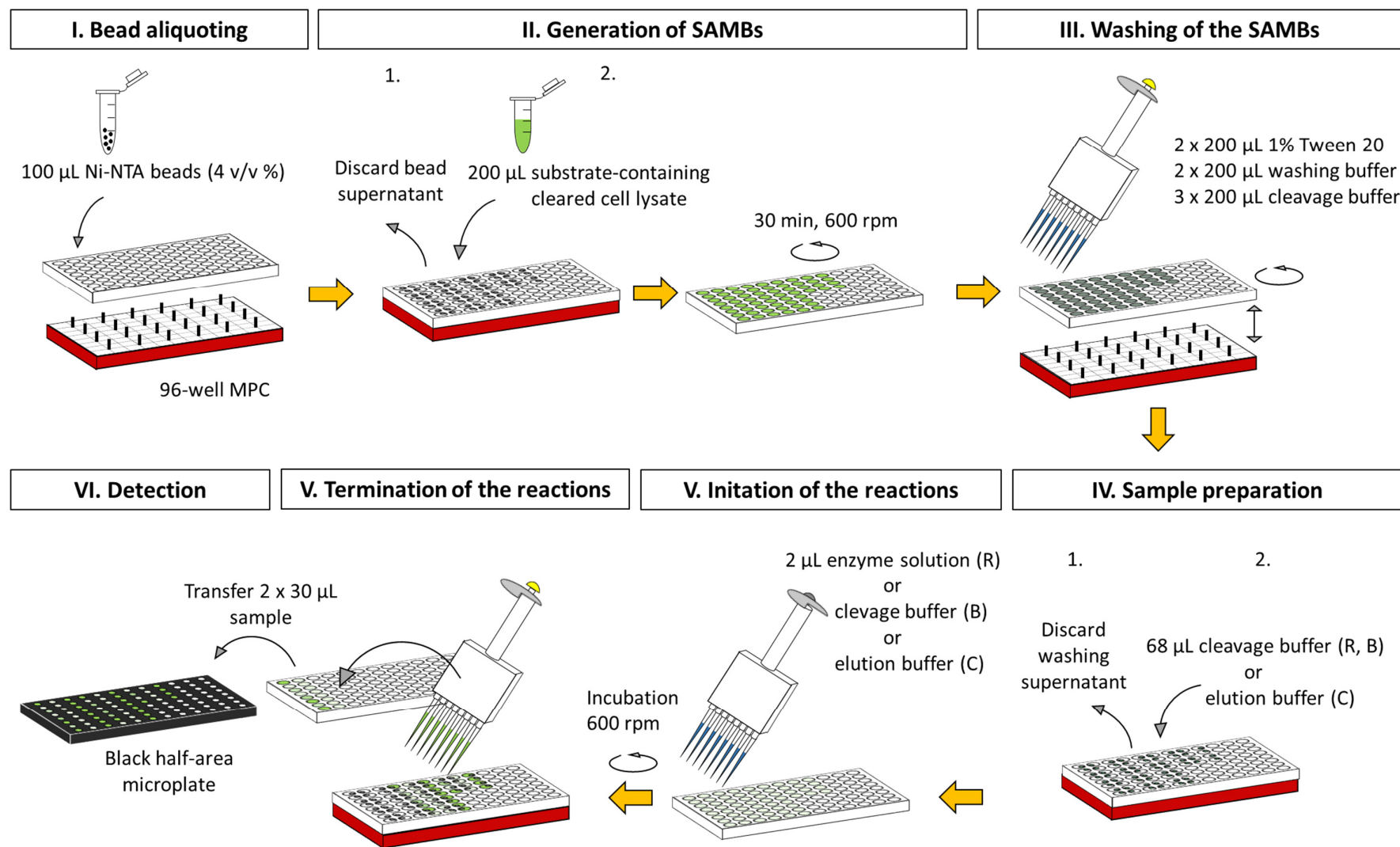
4.12.6. Time course kinetic studies of VEEV nsP2pro-2

Cleavage of the wild-type His$_6$-MBP-SFV-1-mEYFP substrate (at 0.02 mM final concentration) by VEEV nsP2pro-2 (at 3.0 µM final total concentration) was followed by measuring fluorescence after for 0, 5, 10, 15, 20, 30, 50, 70, and 90 minutes at 30 °C, while continuously shaking at 600 rpm. Molar concentration of the C-terminal fluorescent cleavage product in the reaction samples were plotted against time (min). Linear regression was performed, and the parameters of the fitted lines were determined by GraphPad Prism version 5.00.

4.12.7. Substrate-dependent kinetic studies of VEEV nsP2pro-2

Kinetic measurements were performed on substrates listed in **Table 10**. The applied enzyme concentration was different in each case as listed in **Table 10**. The samples were incubated for 10 minutes at 30 °C, while continuously shaking at 600 rpm. Evaluation of the enzyme kinetic parameters was performed as described in **Section 4.12.3**. Due to the lack of any tight-binding selective inhibitor for alphavirus proteases, the active enzyme concentration could not be determined, therefore, when calculating the catalytic constants 100 % activity was assumed for the VEEV PR nsP2pro-2. Significance of the differences between the k_{cat}/K_M values of the tested variants and the wild-type substrate were determined by GraphPad unpaired t-test (http://graphpad.com/quickcalcs/ttest2/). If p-value < 0.05, the difference was considered to be statistically significant.

Table 10. **The list of the substrates on which VEEV nsP2pro-2 kinetic studies were performed.** The corresponding substrate and enzyme concentrations are indicated.

Substrate	Substrate concentration (µM)	Applied VEEV nsP2pro-2 concentration (µM)
His$_6$-MBP-SFV-1(wt)-mEYFP	3.15 – 63.3	3.94
His$_6$-MBP-SFV-1 P4-Glu-mEYFP	0.71 – 28.5	1.46
His$_6$-MBP-SFV-1 P4-Thr-mEYFP	0.76 – 30.5	1.88
His$_6$-MBP-SFV-1 P4-Arg-mEYFP	0.84 – 21.1	3.75
His$_6$-MBP-SFV-1 P4-Gly-mEYFP	0.71 – 14.2	5.21
His$_6$-MBP-SFV-1 P1-Gly-mEYFP	0.79 – 15.8	5.21
His$_6$-MBP-SFV-1 P1'-Thr-mEYFP	1.06 – 21.1	7.29
His$_6$-MBP-SFV-1 P2'-Ser-mEYFP	0.76 – 19.4	4.11

4.13. Regeneration of the Ni-NTA magnetic beads

Ni-NTA magnetic beads were used repeatedly during the experiments; the beads were regenerated before each use. Beads were washed by 1.8 mL buffers (**Table 11**). Each washing step was repeated five times in the order indicated in **Table 11.**

Table 11.   List of applied buffers during the regeneration of Ni-NTA magnetic beads. For long-term storage the beads were suspended into Regeneration buffer E at 2-8 °C.

Bead type	Name of the buffer	Order of buffers in washing steps	Buffer compositions
Qiagen	Regeneration buffer A	1	0.05 % Tween 20, 0.5 M NaOH
	Regeneration buffer B	2, 4, 6	0.05 % Tween 20
	Regeneration buffer C	3	0.05 % Tween 20, 100 mM EDTA, pH 8.0
	Regeneration buffer D	5	0.05 % Tween 20, 100 mM NiSO$_4$, pH 8.0
	Regeneration buffer E	7	0.5 % Tween 20, 30 % ethanol, pH 7.0
PureCube	Regeneration buffer F	1	100 mM EDTA, 0.05 % Tween 20 pH 7.0
	Regeneration buffer B	2, 4, 6, 8	0.05 % Tween 20
	Regeneration buffer E	3	0.5 M NaOH, 2 M NaCl 0.05 % Tween 20
	Regeneration buffer D	5	0.05 % Tween 20, 100 mM NiSO$_4$, pH 8.0
	Regeneration buffer G	7	100 mM acetic acid, 150 mM NaCl
	Regeneration buffer H	9	20 mM Tris-base, pH 8.0

4.14. In-solution digestion of the recombinant substrates

In-solution digestion reactions were initialized by the addition of 2-5 µL purified HIV-1 PR (4.1 µM) or 5 µL TEV PR (21 µM) or 5 µL VEEV nsP2pro-2 (73 µM) to their corresponding purified recombinant fluorescent substrates (0.01-0.03 mM) dissolved in cleavage buffer. The final reaction volume was set to 70 µL. The reaction mixtures were incubated for 1-22 hours at 37 °C in case of HIV-1 PR or at 30 °C in case of TEV PR and VEEV nsP2pro-2 and were stopped by the addition of 6X denaturing or non-denaturing sample loading buffer. Samples were analyzed by PAGE.

4.15. Kinetic measurements on HIV-1 PR by RP-HPLC

Enzyme reactions were initiated by the mixing of 5 μL purified HIV-1 PR (dialyzed against 2X cleavage puffer, 430 nM final concentration), 10 μL 2X cleavage buffer, and 5 μL synthetic oligopeptide substrate (in 0.47-2.35 mM final concentrations) representing the naturally occurring MA/CA cleavage site of HIV-1 PR (VSQNY↓PIVQ). The mixtures were incubated at 37 °C for 10 minutes and stopped by the addition of 9 volumes of 1 % trifluoroacetic acid (TFA). The samples were injected onto Nova-Pak C18 reversed-phase chromatography column (Waters Associates, Inc.) using an automatic injector. Separation of substrates and cleavage products was performed by using acetonitrile gradient (0 to 100 %) in water, in the presence of 0.05 % TFA. Peptides were detected at 206 nm, followed by integration of the peak areas. Reactions were monitored at < 20 % substrate hydrolysis.

The molar concentration of the generated cleavage products (nM) was assessed according to the following:

$$[P] = \frac{(P1 + P2) * [S]}{S_{blank}}$$

where, P1, P2 are integration values of area under curve of the first and the second cleavage products. [S] is the final, initial molar concentration (nM) of the oligopeptide substrate in the reaction, while S_{blank} stands for the integration values of area under curve of the substrate in the blank, where no enzyme was added.

Initial velocity values (nM s^{-1}) calculated by dividing the generated cleavage products by incubation time (s), where plotted against the applied substrate concentration. Kinetic parameters were determined as described in **Section 4.12.3**. Active site titration of the enzyme was performed at the same reaction setting as above at 0.47 mM final oligopeptide substrate concentration but using 2X cleavage buffer contained protease inhibitor amprenavir in final concentration ranging from 1 nM to 80 nM. Active enzyme concentration (22.05 nM) has been determined as described in **Section 4.12.2**.

4.16. PAGE and in-gel renaturation of the fluorescent substrates and products

Purified recombinant protein substrates and cleavage products were analyzed by either native or SDS-PAGE using 14 % separating and 4 % stacking gel. Preparation of the samples for the PAGE analysis were performed based on two different protocols: (i) non-denatured samples: 6X loading buffer containing no reducing agents (300 mM Tris, 20 % glycerol, 0.05 % bromophenol blue, pH 6.8) was added to samples, which were not heat-treated before the electrophoresis (**Gross et al., 2000**); (ii) denatured samples: 6X loading buffer supplemented with reducing agents - 12 % SDS and 100 mM β-ME - was added to the samples, followed by heat denaturation at 95 °C for 10 minutes.

In case of denaturing PAGE, the electrophoresis was followed by rinsing the gels in distilled water for 30 minutes at room temperature in order to wash the SDS out from the gel and thereby to renature the separated proteins. In case of non-denaturing PAGE analysis, the separated fluorescent proteins were in their native state, therefore the washing step could be omitted. The fluorescent proteins in the unstained gels were visualized by using Dark Reader Blue transilluminator (Labgene Scientific) and by using UV imaging function of an AlphaImager gel documentation system (ProteinSimple). For staining the gels, PageBlue Protein Staining Solution (Thermo Fischer Scientific) was applied, and then band intensities were determined by Dr. János András Mótyán using ImageJ 1.43 software (**Schneider et al., 2012**) and by GelAnalyzer 2010a program (www.gelanalyzer.com).

5. Results

5.1. Expression vector system

We have prepared the pDEST-His$_6$-MBP-FP plasmids using Gateway cloning technology based on the description of Tropea et al (**2007**). N1 and C primers were used to amplify the desired FPs. Due to their common origin, the terminal sequences of the different FPs are often identical, therefore, the designed N1 and C primers are suitable for the amplification of those FPs whose terminal nucleotide sequence is identical to that of EGFP. Additionally, N1 primer was designed to code for a 'cloning cassette' positioned between the TEV PR cleavage site, originally incorporated into the cloning platform of Tropea *et al.* (**2007**), and the FP to be amplified.

Empty pDEST-His$_6$-MBP-FP plasmids carry the coding sequence of a His$_6$ affinity tag and an MBP fusion protein, followed by a TEV PR cleavage site (ENLYFQ↓G), the 'cloning cassette', and a C-terminal fluorescent protein (FP). The 'cloning cassette' contains cleavage sites of *PacI* and *NheI* restriction endonucleases next to each other and allows the insertion of the nucleotide sequence of the proteolytic cleavage site of interest by ligation. The structure of the 'empty' expression vector is shown in **Figure 9**.
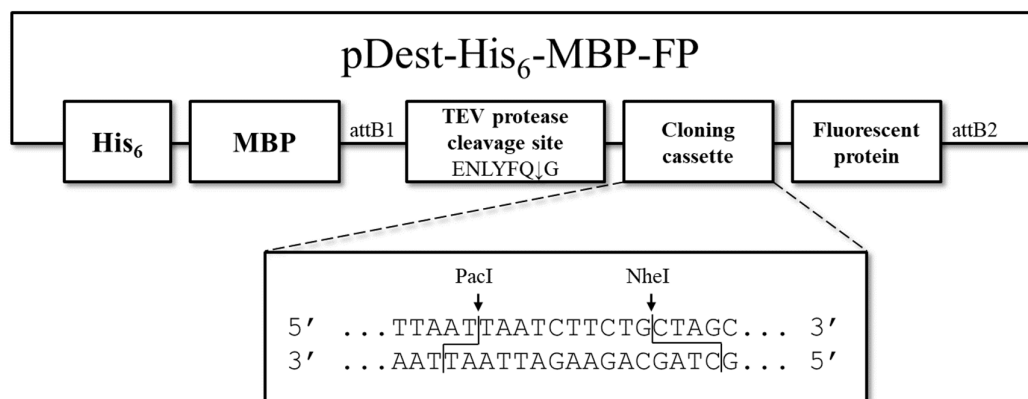
Figure 9. **Schematic illustration of pDEST-His$_6$-MBP-FP vector.** Nucleotide sequence of the 'cloning cassette' is zoomed, cleavage sites of *PacI* and *NheI* restriction enzymes are indicated. The cleavage site sequence of TEV PR is also shown.

The empty vectors linearized by *PacI* and *NheI* restriction enzymes have sticky ends and can be re-circularized by ligation after the addition of a linear dsDNA sequence flanked by *PacI* and *NheI* sticky ends. The dsDNA sequences coding for the cleavage site sequence of interest can be generated either by (i) random mutagenesis of a wild-type cleavage site sequence (**Figure 10**) or (ii) by the hybridization of complementary, *E. coli*-optimized oligonucleotide primers (**Figure 11**).
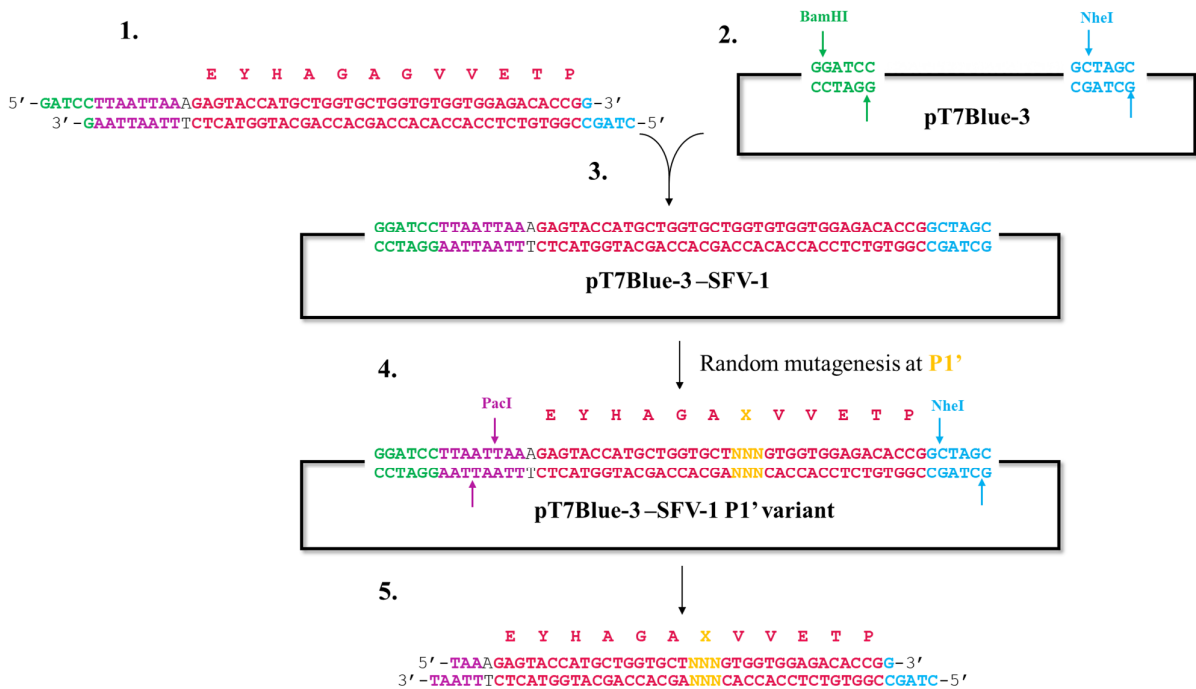
Figure 10. **Generation dsDNA inserts coding for SFV-1 P1' variants by random mutagenesis.** The procedure consists of five main steps: 1. Annealing of complementary, *E. coli*-optimized oligonucleotide primers coding for SFV-1 flanked by *BamHI* and *NheI* sticky ends. 2. Linearization of pT7-Blue-3 vector by *BamHI* and *NheI*. 3. Generation of pT7-Blue-3-SFV-1 by ligating the dsDNA generated in Step 1. into linearized pT7-Blue-3 vector. 4. Generation of vectors coding for several different SFV-1 P1'variant by random mutagenesis using a degenerated oligonucleotide primer at the codon corresponding to P1' residue. 5. Recovery of dsDNA coding from vectors coding for different SFV-1 P1' variants after digestion by *PacI* and *NheI* restriction enzymes.
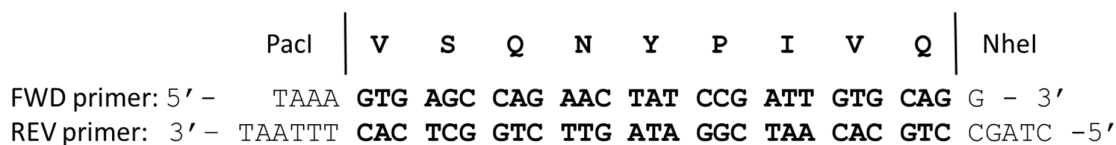
Figure 11. **Structure of a short dsDNA fragment coding for a protease cleavage site.** The annealing of the complementary oligonucleotide primers is illustrated on the example of HIV-1 matrix/capsid (MA/CA) cleavage site. Sticky ends, corresponding to that of *PacI* and *NheI* restriction endonucleases are represented at the termini of the short dsDNA.

Notably, the coding sequences of the fluorescent proteins are not in the ORF without the insertion of the properly designed linear dsDNA fragments, therefore, the fluorescent proteins are translated only after a successful ligation.

After preparation of the empty expression plasmids coding for different C-terminal FPs (including ECFP, mTurquoise2, EGFP, mEYFP, mApple, and mCherry), the dsDNAs coding for the

cleavage site sequences were cloned into different empty vectors. **Table 12** summarizes the vector library produced by the herein described cloning system. Some of the listed constructs are made for different studies, that are not considered to be part of this dissertation.

Table 12. **List of the generated fluorescent substrate-coding expression plasmids produced by using the different pDEST-His₆-MBP-FP 'empty' vectors.** Proteases that are specific to the generated substrates are indicated in the column 'Enzyme specificity'. Positions different from the wild-type amino acid sequences are marked in red. SFV-1 and SFV-3 are stands for the wild-type cleavage site between the nsP1/nsP2 and nsP3/nsP4 of semliki forest virus (SFV) respectively. Mutants produced by random mutagenesis are marked with asterisk, otherwise were produced by using complementary oligonucleotide primers.

Inserted cleavage site sequence	Name	C-terminal FP	Enzyme specificity
LGRAGA↓YIFSSD	SFV-3	mTurquoise2	VEEV PR, SFV PR
		ECFP	
		mEYFP	
		mCherry	
EYHAGA↓GVVETP	SFV-1	mTurquoise2	VEEV PR, SFV PR
		ECFP	
		mEYFP	
		mCherry	
EYHAGA↓TVVETP *	SFV-1 P1' Thr	mTurquoise2	VEEV PR
		mEYFP	
EYHAGA↓VVVETP *	SFV-1 P1'-Val	mEYFP	
EYHAGA↓PVVETP *	SFV-1 P1'-Pro	mEYFP	
EYHAGA↓HVVETP *	SFV-1 P1'-His	mEYFP	
EYHAGA↓CVVETP *	SFV-1 P1'-Cys	mEYFP	
EYHAGA↓SVVETP *	SFV-1 P1'-Ser	mEYFP	
EYHAGA↓RVVETP *	SFV-1 P1'-Arg	mEYFP	
EYHAGA↓LVVETP *	SFV-1 P1'-Leu	mEYFP	
EYHAGA↓KVVETP *	SFV-1 P1'-Lys	mEYFP	
EYHAGA↓EVVETP *	SFV-1 P1'-Glu	mEYFP	
EYHAGA↓YVVETP	SFV-1 P1'-Tyr	mEYFP	
EYHAGA↓IVVETP	SFV-1 P1'-Ile	mEYFP	
EYHAGA↓QVVETP	SFV-1 P1'-Gln	mEYFP	
EYHAGA↓NVVETP	SFV-1 P1'-Asn	mEYFP	
EYHAGA↓DVVETP	SFV-1 P1'-Asp	mEYFP	
EYHAGA↓AVVETP	SFV-1 P1'-Ala	mEYFP	
EYHAGA↓MVVETP	SFV-1 P1'-Met	mEYFP	
EYHAGA↓WVVETP	SFV-1 P1'-Trp	mEYFP	
EYHAGA↓FVVETP	SFV-1 P1'-Phe	mEYFP	

Inserted cleavage site sequence	Name	C-terminal FP	Enzyme specificity
E**Q**HAGA↓GVVETP	SFV-1 P5-Gln	mEYFP	
EY**E**AGA↓GVVETP	SFV-1 P4-Glu	mEYFP	
EY**T**AGA↓GVVETP	SFV-1 P4-Thr	mEYFP	
EY**R**AGA↓GVVETP	SFV-1 P4-Arg	mEYFP	
EY**G**AGA↓GVVETP	SFV-1 P4-Gly	mEYFP	
EYHA**A**A↓GVVETP	SFV-1 P2-Ala	mEYFP	
EYHA**V**A↓GVVETP	SFV-1 P2-Val	mEYFP	
EYHAG**G**↓GVVETP	SFV-1 P1-Gly	mEYFP	
EYHAG**V**↓GVVETP	SFV-1 P1-Val	mEYFP	
EYHAGA↓G**P**VETP	SFV-1 P2'-Pro	mEYFP	
EYHAGA↓G**S**VETP	SFV-1 P2'-Ser	mEYFP	
VSQNY↓PIVQ	MA/CA -Wild-type cleavage site of HIV-1 - (P5-P4')	mTurquoise2	HIV-1 PR
		mEYFP	
		mApple	
		mCherry	
VSQ**L**Y↓PIVQ	MA/CA P2 Leu (P5-P4')	mTurquoise2	HIV-1 PR
		mEYFP	
		mApple	
		mCherry	
KARVL↓AEAM	CA/p2 - Wild-type cleavage site of HIV-1 (P5-P4')	mTurquoise2	HIV-1 PR
		mEYFP	
		mApple	
		mCherry	
DTGHSNQVSQNY↓PIVQNIQGQMVH	MA/CA (P12-P12')	mTurquoise2	HIV-1 PR
		mEYFP	
		mApple	
		mCherry	
GVGGPGHKARVL↓AEAMSQVTNSAT	CA/p2 (P12-P12')	mTurquoise2	HIV-1 PR
		mEYFP	
		mApple	
		mCherry	
GVGGPGHVSQNY↓PIVQN**Q**VTNSAT	CA/p2 (P12-P6), MA/CA (P5-P5'), CA/p2 (P6'-P12')	mTurquoise2	HIV-1 PR
		mEYFP	
		mApple	
		mCherry	
DTGHSNQKARVL↓AEAMS**IQGQMVH**	MA/CA (P12-P6), CA/p2 (P5-P5'), MA/CA (P6'-P12')	mTurquoise2	HIV-1 PR
		mEYFP	
		mApple	
		mCherry	

5.2. Recombinant fluorescent fusion protein substrates

Recombinant fusion protein substrates contain an N-terminal hexahistidine affinity tag (His_6) fused to a maltose binding protein (MBP), followed by cleavage sites of TEV and the inserted cleavage site sequence of the protease of interest, while on the C-terminal end a fluorescent protein variant is fused to the substrate construct. The role of the His_6 affinity tag is to enable the recombinant proteins to be immobilized on metal-chelate surfaces, and also to facilitate the purification of the proteins by immobilized metal affinity chromatography (IMAC). The N-terminal MBP fusion protein enhances the water-solubility of the recombinant protein substrate and improves its folding (**Kapust et al., 1999; Fox et al., 2002, 2003**), while TEV protease cleavage site serves as an internal control cleavage site. The recombinant substrates were successfully expressed in *E. coli* BL21(DE3) strain in a final volume of 50 mL. The small-scale expressions, with 15 mL final volume were also optimized for the substrates used in the 96-well-based specificity screening studies of VEEV PR nsP2pro-2. As some FPs produced by *E. coli* cells may have a longer maturation time (**Balleza et al., 2018; Hebisch et al., 2013**); in order to increase the fluorescent yield of the substrate solution, the protein translation can be optionally arrested by the tetracycline treatment, as it was applied in case of mEYFP-fused substrates expressed for VEEV nsP2pro-2 kinetic measurements.

Cells used for the expression were disrupted and lysed under native conditions, and the cleared bacterial cell lysates containing the soluble substrates were used directly for sample preparation in a Ni-NTA magnetic agarose bead-based protease assay system (described in **Section 5.2.1**) or for substrate purification. Purified substrates can be further used for calibration, in-solution digestion and/or for PAGE analysis. The illustration of the versatile utilization of the expressed substrates is summarized in **Figure 12**.
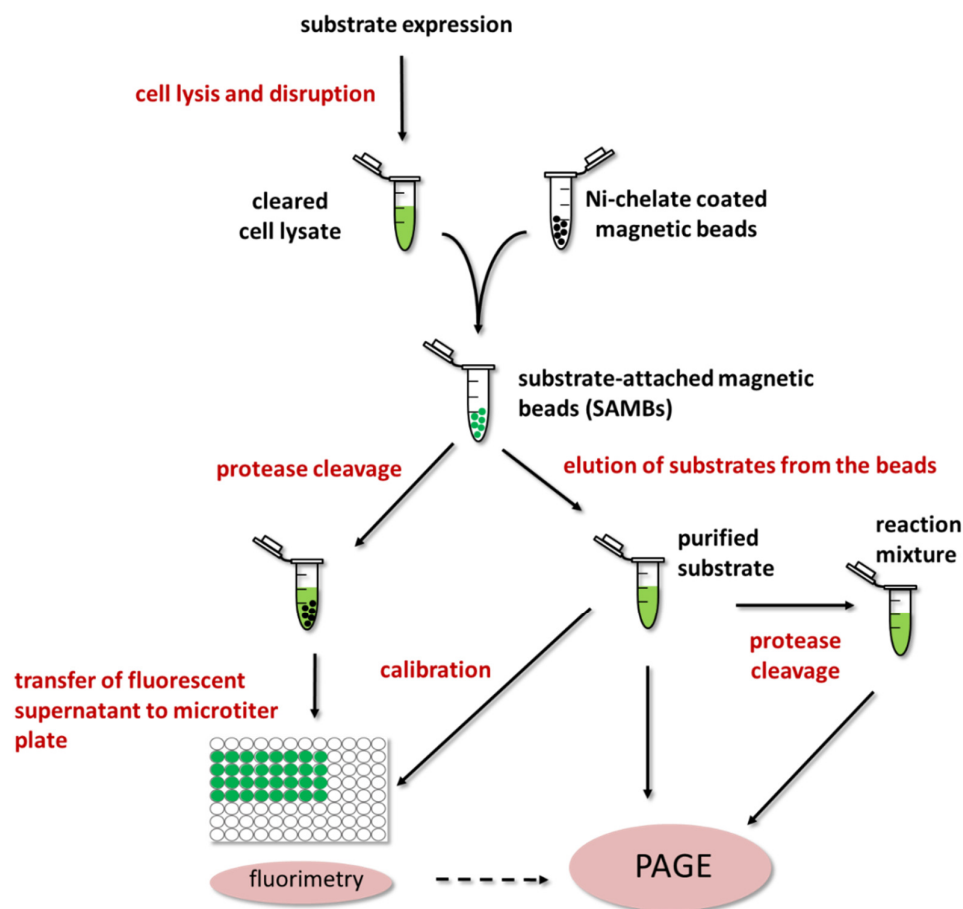
Figure 12. **A possible application of the recombinant fluorescent protein substrates.** Overexpressed substrates in the cleared bacterial cell lysates can be used for the generation of substrate attached magnetic beads (SAMBs). SAMBs can be further utilized in an on-bead protease assay or for substrate purification. Purified substrates can be further used for calibration, in-solution digestion. Intact substrates and cleavage products can be detected by fluorimetry and/or by PAGE analysis.

5.2.1. Ni-NTA magnetic-bead-based protease assay

Recombinant fluorescent substrates can be attached to Ni-chelate coated surface *via* their His_6 tag. The substrates attached to the beads are accessible for cleavage by the protease of interest. After processing, the N-terminal cleavage products (and the uncleaved substrates) remain attached to the beads, while the fluorescent C-terminal cleavage products are released into the supernatant and after separation can be detected with high sensitivity (**Figure 13**).
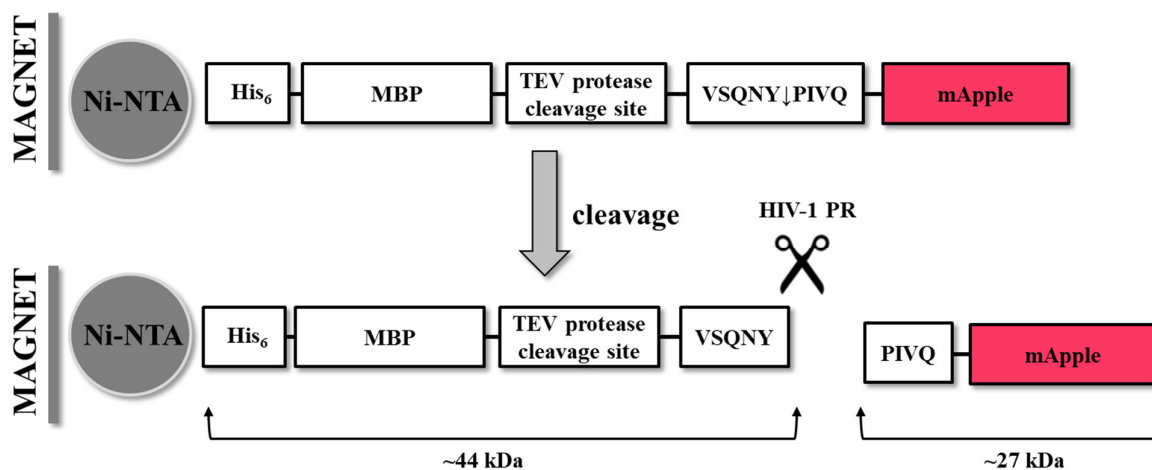
Figure 13.   Principle of the fluorescent protease assay. The assay is exemplified by the cleavage of His$_6$-MBP-VSQNY↓PIVQ-mApple substrate by HIV-1 PR**.**

5.2.1.1.   Assay workflow

Generally, the workflow of the developed protease assay includes six steps. (i) The procedure starts with the preparation of the assay samples. Firstly, the recombinant substrates, either purified or in crude form of the cleared cell lysate, are incubated with the affinity magnetic beads, and the substrate-attached magnetic beads (SAMBs) are formed. After washing steps, the SAMBs are aliquoted for the preparation of assay samples. (ii) The preparation of the reaction mixtures is followed by reaction's initialization by the addition of the protease. Upon cleavage, the proteolytic fragments are released into the supernatant. (iii) The reactions are terminated by the separation of the magnetic beads from the reaction mixture containing the fluorescent cleavage products and the enzyme. (iv) The supernatants are applied to the wells of a microtiter plate and the fluorescence is determined by fluorimetry. (v) Calibration curves are generated using purified fluorescent substrates solved in each assay buffers. (vi) The concentration of C-terminal fluorescent cleavage products and also that of the applied substrate in the assay samples are determined based on fluorescent intensity of the different assay sample types (see **Section 5.2.1.2**) by using the slope of the calibration curves (see **Section 5.2.1.3**).

5.2.1.2.   Assay sample types

There are three different sample types applied in the developed protease assay that are essential for the reliable evaluation of the assay results. All of the three sample types are generated from the same SAMB stock solution, are handled parallel to each other, and are exposed to same experimental conditions.

i. *Reaction (R) sample* contains both the SAMBs and the enzyme mixed in the reaction buffer and is used for assessing proteolytic cleavage properties.

ii. *Substrate blank (B) sample* contains only the SAMBs in cleavage buffer and applied to assess spontaneous substrate dissociation during the reaction.

iii. *Substrate control (C) sample* contains only the SAMBs in elution buffer and is used for determining the initial substrate concentration in the reaction by eluting and determining the amount of substrate attached to the Ni-NTA bead surface.

5.2.1.3. Substrate Calibration

Calibration procedures were performed to determine the quantities of fluorescent substrates and cleavage products. Substrates were serially diluted followed by measurement of fluorescence intensities, both in reaction and elution buffer. Based on our experiences, relative fluorescence intensities are directly proportional to the substrate concentration in the range applied in our experiments. **Figure 14** represents an example for the calibration curves of two different substrates at reaction buffer and elution buffer conditions. Linear correlation of the blank-corrected fluorescence intensity values to the corresponding concentrations is a prerequisite for evaluation, at least in the concentration range of the assay. R^2 value indicates a good correlation between the concentration of fluorescent protein and fluorescence only if a sufficient number of data points have been used to cover the entire concentration range. As experimental errors can highly affect the reliability of the calibration; therefore, a graphical evaluation of the regression lines may be also necessary.

To test whether fluorescence intensities of the substrates depend on the inserted cleavage site, calibrations were performed by using a series of mTurquoise2-, mEYFP and mApple-fused substrates, both in elution and reaction buffers. We found that coefficient of variance (CV %) values of the slopes were <15 % in all cases (**Table 13**), which implied that the fluorescence values of the substrates are independent on the length and sequence of the inserted cleavage site. Furthermore, the slope values, calculated based on the substrate calibration curves, can potentially be used for the determination of substrate concentration, if identical conditions were applied.
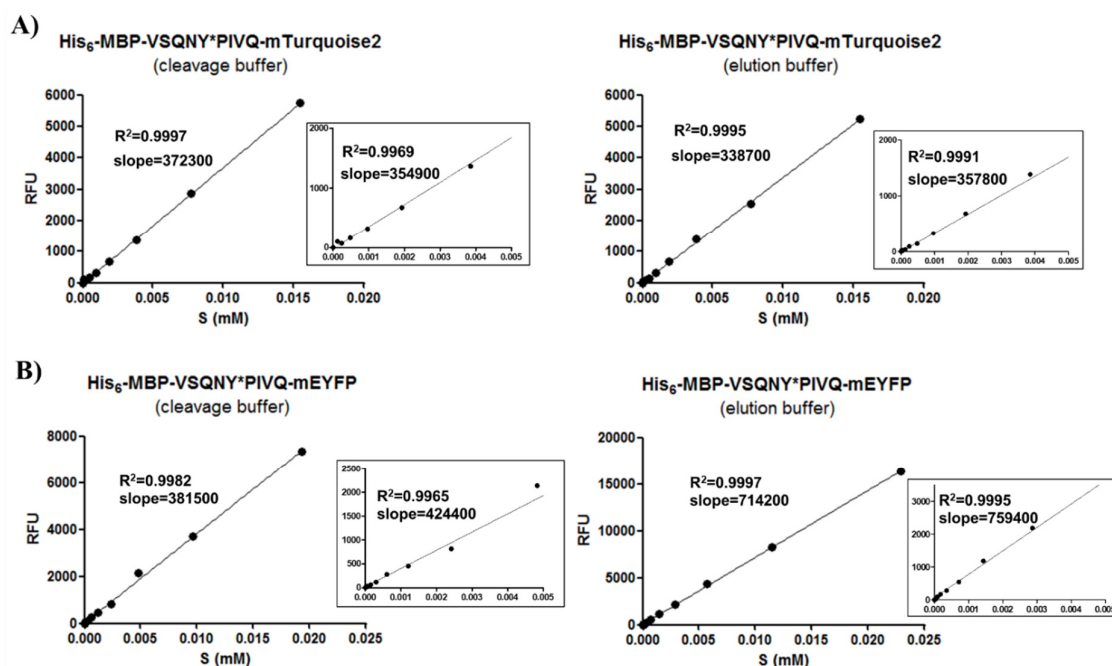
Figure 14. Representative substrate calibration curves. Calibration was performed by using A) His₆-MBP-VSQNY↓PIVQ-mTurquoise2 and B) His₆-MBP-VSQNY↓PIVQ-mEYFP substrates. Zoom-in figures are added to represent the linear regression of data points in the 0-0.005 mM substrate concentration range. Left and right panels show calibrations in cleavage and elution buffers, respectively.

Table 13. Coefficient of variance % of calibration curve slopes. Table shows slope values determined based on the calibration curves of substrates (containing the same C-terminal FP: mTurquoise2, mApple, or mEYFP) dissolved either in cleavage buffer or elution buffer.

Inserted Cleavage site sequence	mTurquoise2-fused		mApple-fused	
	elution	cleavage	elution	cleavage
VSQNY↓PIVQ	989896.51	1075347.41	2050938.07	1582899.26
DTGHSNQVSQNY↓PIVQNIQGQMVH	1217967.10	1193298.12	2670067.63	2183369.95
GVGGPGHVSQNY↓PIVQNQVTNSAT	1119401.30	1269630.77	2525333.42	1955601.48
KARVL↓AEAM	1276280.50	1261207.99	2908597.42	2122337.49
GVGGPGHKARVL↓AEAMSQVTNSAT	1262153.05	1242445.03	2244804.44	1944158.31
DTGHSNQKARVL↓AEAMSIQGQMVH	1169880.95	1247326.55	2316533.48	2091690.05
CV %	**6.04**	**9.11**	**10.92**	**12.68**

Inserted Cleavage site sequence	mEYFP-fused	
	elution	cleavage
EYEAGA↓GVVETP	445786.02	1123862.81
EYTAGA↓GVVETP	502034.16	1156585.29
EYGAGA↓GVVETP	553751.89	1097815.33
EYHAGG↓GVVETP	582854.24	1097131.78
EYHAGA↓TVVETP	505728.64	1165983.44
EYHAGA↓GSVETP	649229.03	1245777.81
CV %	**13.21**	**4.87**

5.2.2. PAGE analysis and in-gel renaturation

The substrates and cleavage products - both in the case of the Ni-NTA magnetic bead-based assay and in-solution digestion - can be analyzed by native or reducing PAGE.

In our experiments, cleavage reaction of HIV-1, TEV, and VEEV nsP2pro-2 proteases were also analyzed by PAGE. Samples were prepared at either denaturing or non-denaturing conditions. During non-denaturing sample preparation, the applied loading buffer contained no reducing agents and the samples were not heat-treated before the electrophoresis. In contrast, denaturing sample preparation was performed using loading buffer containing SDS and β-ME and the samples were heat-treated prior to the analysis. After the electrophoresis, the fluorescence of the protein bands in the case of non-denatured samples was readily detected in the gel under visible light (even by naked eye), by using a dark-reader blue light transilluminator or UV imaging using a gel documentation system (**Figure 15** and **Figure 16**). In case of denatured samples, where SDS distorts protein structure, the fluorescence of the analyzed proteins was not detected by any of the above listed tools. In order to regenerate protein structure and fluorescence of the separated fluorescent proteins, the gels were washed with distilled water in order to remove SDS. We found that removal of the SDS made the detection of fluorescence possible even by naked eye or by using either blue light transilluminator or UV imaging (**Figure 17 A and B**).
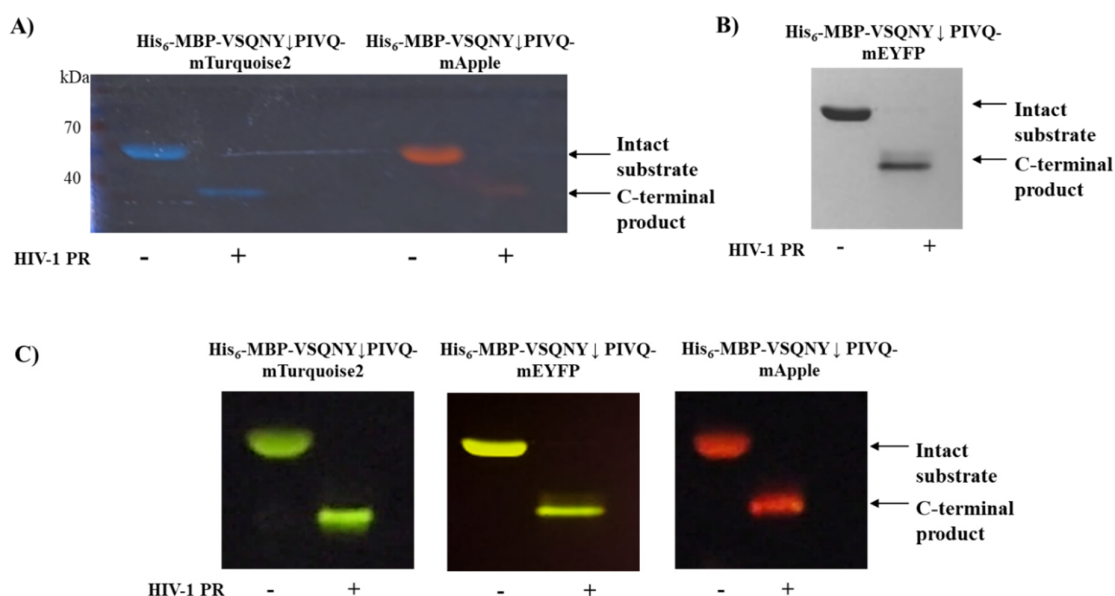
Figure 15.   Detection of intact substrates and cleavage products after in-solution digestion with HIV-1 PR. Fluorescent bands can be detected in the gel A) under visible light, with a black paper sheet behind the gel, B) by using UV transillumination and c) B) using blue light transilluminator. The samples were prepared at non-denaturing conditions and were analyzed on SDS-PAGE.
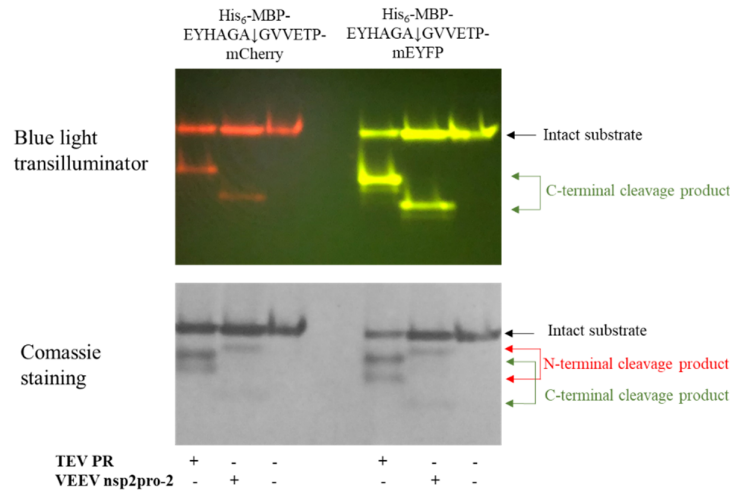
Figure 16. **Native PAGE analysis of purified intact substrates and their cleavage products after the in-solution digestion by TEV PR and VEEV nsP2pro-2.** The samples for the PAGE analysis were prepared at non-denaturing conditions. N-terminal cleavage products are visible only in the Coomassie-stained gel.
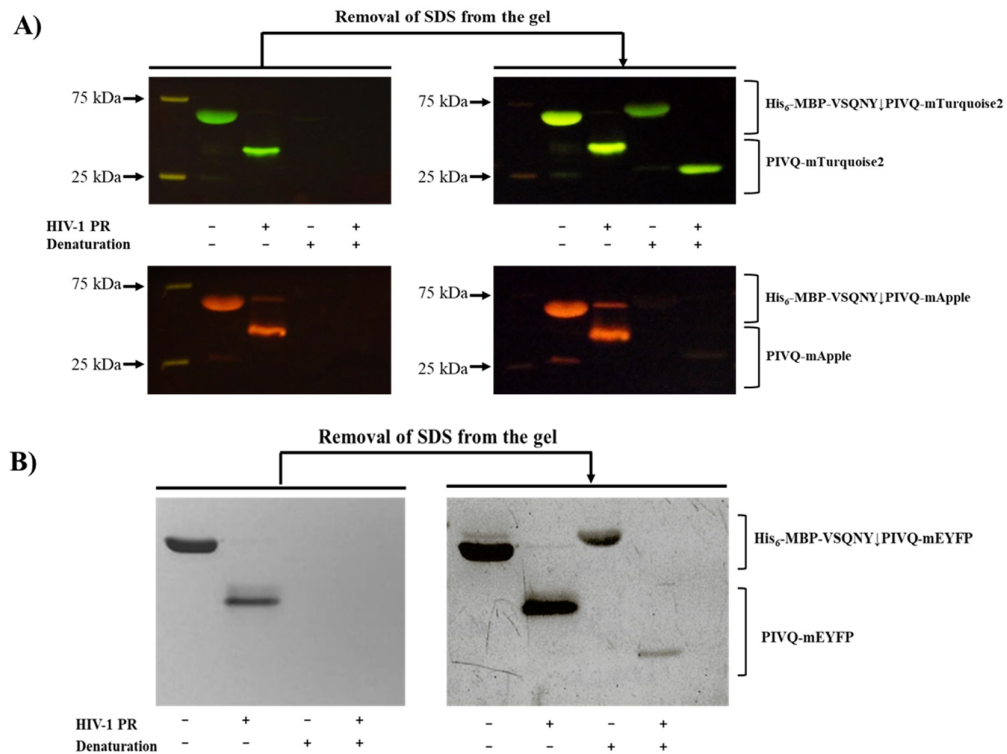
Figure 17. **Effect of renaturation on the in-gel detection of fluorescent proteins.** Fluorescent detection of the intact purified substrates and their cleavage products produced by in-solution digestion before and after the removal of SDS from the polyacrylamide gel. The detection was performed A) by using blue light transilluminator or B) by using UV light.

There were differences observed in the renaturation abilities of mApple and mTurquoise2. mTurquoise2 showed substantially better renaturation ability compared to mApple, when illuminated by blue or UV light (**Figure 18**). Based on the densitometry, band intensity of the renatured mTurquoise2 was at least 50 % compared to the non-denatured. In contrast with this, mApple showed only ~10 % fluorescence intensity after renaturation, however, it was still detectable in the gel after UV illumination.
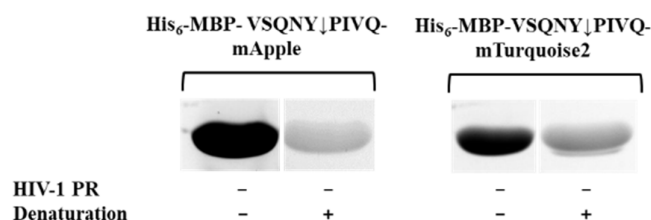
Figure 18. **Bands of the purified intact recombinant substrates visualized under UV light after in-gel renaturation.** mTurquoise2 shows higher degree of renaturation ability compared to mApple**.**

We have compared how the N- and C-terminal proteolytic cleavage products are separated on SDS-PAGE when prepared at non-denaturing or denaturing conditions. The cleavage products were generated by the in-solution digestion of substrates with HIV-1 PR, followed by electrophoresis. We have observed that the cleavage products were not separated fully from each other in the case of non-denaturing sample preparation, while can be well differentiated in the case where non-denaturing conditions were applied (**Figure 19**).
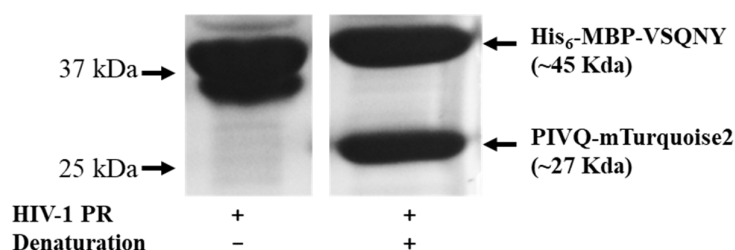
Figure 19. **Denatured and non-denatured cleavage products released from His6-MBP-VSQNY↓PIVQ-mTurquoise2 after HIV-1 PR digestion.** The protein bands were visualized by Coomassie staining after SDS-PAGE analysis.

Notably, SDS presented in the gel may cause the partial denaturation of the native protein, but still, in the samples prepared at non-denatured conditions, the native forms are more abundant (**Figure 20**).
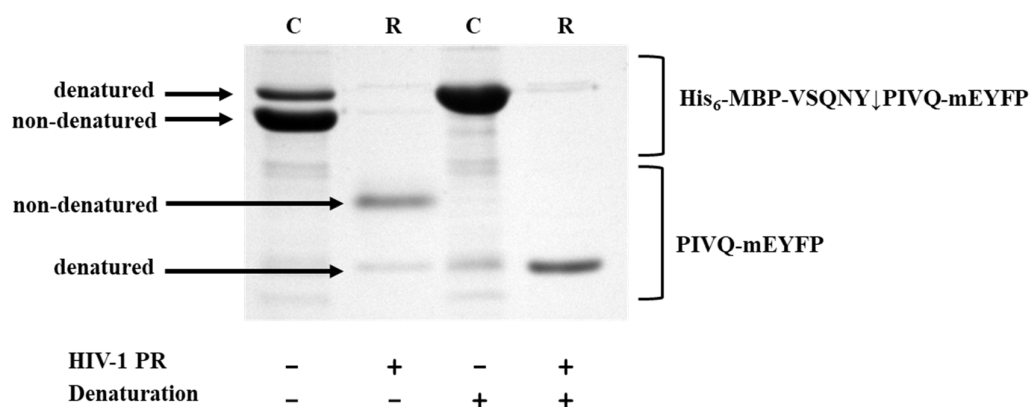
Figure 20. **Analysis of intact His$_6$-MBP-VSQNY↓PIVQ-mEYFP and its C-terminal cleavage products after on-bead digestion with HIV-1 PR.** Intact substrate were sampled from substrate control samples (C), while N-terminal cleavage fragment were sampled from the reaction samples (R) of Ni-NTA magnetic bead-based protease assay. The protein bands were visualized by Coomassie staining after SDS-PAGE analysis**.**

We have investigated whether the cleavage products of the same purified substrate can be separated efficiently after digestion either by HIV-1 or TEV PRs. As there is only a 15 amino acid residue difference in the lengths of the corresponding cleavage products, they have very similar molecular weights with about a ~1.5 kDa difference, which was clearly visible on Coomassie-stained gel ([Figure 21](#)).
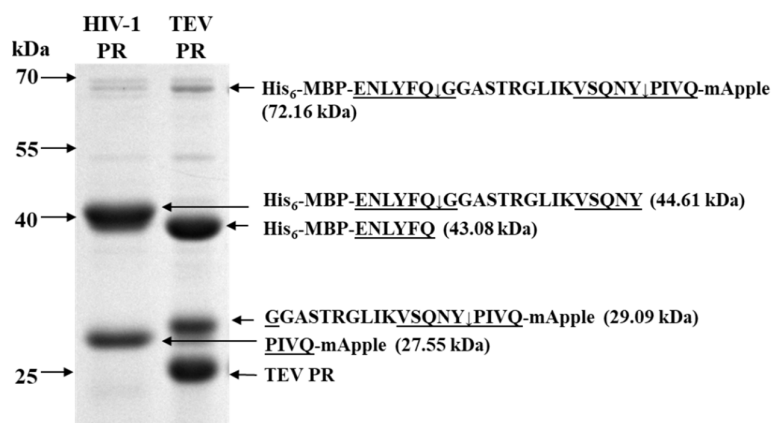
Figure 21. **Analysis of the N- and C-terminal cleavage products released from His$_6$-MBP-VSQNY↓PIVQ-mApple after in-solution cleavage by HIV-1 PR or TEV PR.** The protein bands were visualized by Coomassie staining after SDS-PAGE analysis**.**

We found that the cleavage products - showing only relatively slight differences in the molecular weights - can be separated by SDS-PAGE using 14 % polyacrylamide gel by good resolution, and this method may be sensitive enough for the detection of possible cleavages at alternative sites. Using the TEV PR control cleavage site, the cleavage positions in the recombinant

substrates can be determined by a MALDI TOF-MS-based protocol, which has already been applied successfully in studies on Ty1 PR (**Gazda et al., 2020**).

5.2.3. Standardized manual and video protocol

As the assay platform using His_6-MBP-FP recombinant substrates is a versatile tool for performing several types of proteolytic studies, the standardized instructions of the assay workflow were generated to support the design of proper experimental setup of future users. To generate reliable and reproducible kinetic results, sample handling and calibration procedures of the magnetic bead-based assay were carefully optimized during assay development. These fine-tuned procedures are recommended to be learnt and practiced prior to the application of the system. In order to support the fast acquirement of this knowledge and minimize initial assay failure rate, a tutorial video protocol has been prepared. The recorded assay illustrates the procedures described in **Section 4.3, 4.4, 4.7, 4.8, 4.12, 4.16**. The video protocol is supplemented to this work on the attached pen-drive or can be found here: https://www.jove.com/video/58824/use-recombinant-fusion-proteins-fluorescent-protease-assay-platform.

5.3. Kinetic and inhibition studies of HIV-1 PR

For the demonstration, that the developed assay system is suitable for the determination of enzyme kinetic parameters, we performed both time- and substrate-dependent kinetic measurements by HIV-1 PR, using mTurquoise2- and mApple-fused substrates.

Kinetic time course studies revealed that the formation of the fluorescent cleavage products is satisfactorily linear in the examined timeframe at below the lowest substrate concentration of the substrate-dependent studies (**Figure 22**).
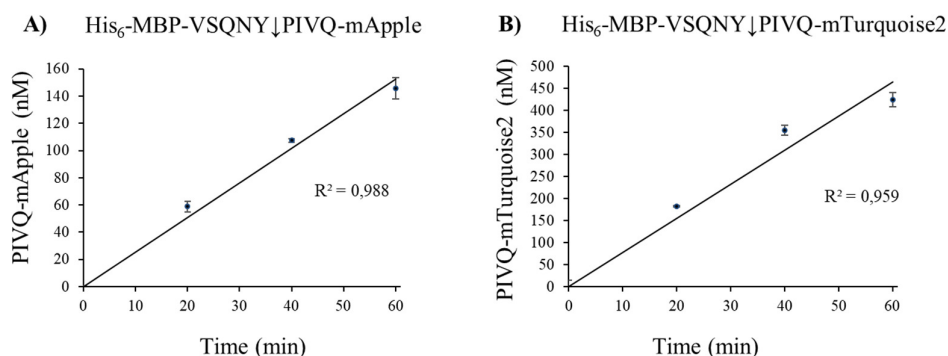
Figure 22. **Time-dependent formation of C-terminal fluorescent cleavage products.** For cleavage by HIV-1 PR, His_6-MBP-VSQNY↓PIVQ-mApple (A) and His_6-MBP-VSQNY↓PIVQ-mTurquoise2 (B) substrates were applied**.** Error bars represent SD, n=2.

For the characterization and comparison of the proteolytic reactions, kinetic parameters were calculated based on the data of substrate-dependent kinetic measurements. Results demonstrate that k_{cat}/K_M values determined for HIV-1 PR using mApple and mTurquoise2-fused substrates closely resemble each other, while the individual k_{cat} and K_M values differed substantially (**Table 14**). Representative curves of the non-linear regression analysis are presented in **Figure 23**.

A) His$_6$-MBP-VSQNY↓PIVQ-mApple

B) His$_6$-MBP-VSQNY↓PIVQ-mTurquoise2
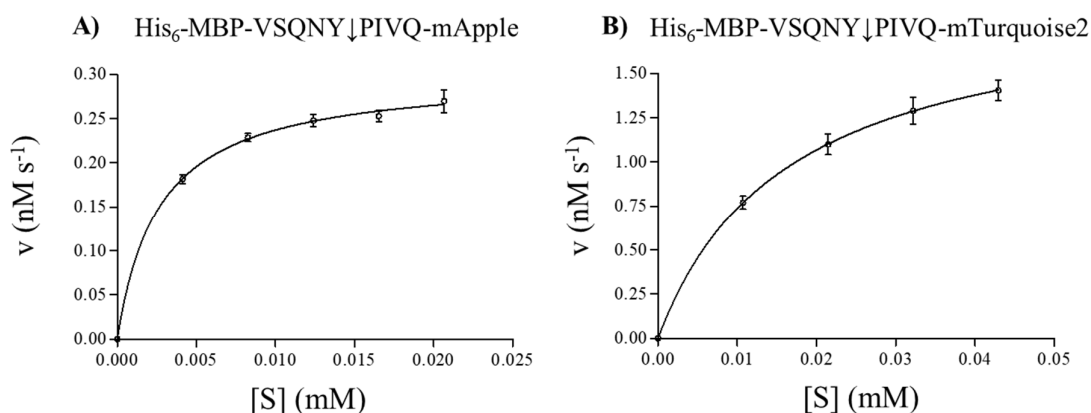
Figure 23. **Non-linear regression analysis of substrate-dependent kinetic curves.** Graphs represent the initial velocity rates of A) PIVQ-mApple and B) PIVQ-mTurquoise2 product formation upon HIV-1 PR treatment, in the function of initial substrate concentrations. Error bars represent SD, n=2.

We intended to compare the kinetic parameters of HIV-1 PR obtained by the magnetic bead-based protease assay with the values determined by using synthetic oligopeptide substrates.

Comparison revealed that k_{cat}/K_M values determined by the different methods closely resemble each other (**Table 14**). Notably, the individual kinetic parameters were substantially higher on synthetic oligopeptide substrate, which difference may be interpreted by the principal conformational differences between the applied substrate types.

Table 14. **Comparison of the kinetic parameters determined by different substrates for HIV-1 PR.**

Substrate	k_{cat} (s^{-1})	K_M (mM)	k_{cat}/K_M (mM^{-1} s^{-1})
His$_6$-MBP-VSQNY↓PIVQ-mTurquoise2	0.321±0.001	0.0163 ± 0.0001	19.69 ± 0.14
His$_6$-MBP-VSQNY↓PIVQ-mApple	0.050± 0.0008	0.0027 ± 0.0002	18.52 ± 1.40
VSQNY↓PIVQ (oligopeptide)	32.21 ± 5.34	2.02 ± 0.51	15.94 ± 4.82

Previously, the specificity constant for HIV-1 PR on VSQNY↓PIVQ oligopeptide substrate have been determined as 45.3 mM^{-1} s^{-1} by our laboratory at highly different buffer conditions including pH 5.6 and high ionic strength (2 M NaCl) (**Tőzsér et al., 1991**), while in our current

experiments the pH and the ionic strength was set to 7.0 and 300 mM NaCl, respectively, in order to optimize the performance of Ni-NTA magnetic beads in the assay. For the determination of the kinetic parameters, the active enzyme concentration of HIV-1 PR has been determined using amprenavir. The effect of amprenavir was tested on the cleavage of His_6-MBP-VSQNY↓PIVQ-mTurquoise2 recombinant fusion protein substrate and VSQNY↓PIVQ oligopeptide substrate, as well. In the RP-HPLC-based assay the folding efficiency of the enzyme was determined as 5.13 %, while in the Ni-NTA bead-based assay it was calculated as 16.6 %. Furthermore, based on the dose-response curve of the Ni-NTA magnetic based inhibitory study (**Figure 24**), the IC_{50} (21.43 nM) and K_i values (6.75 nM) were also determined.
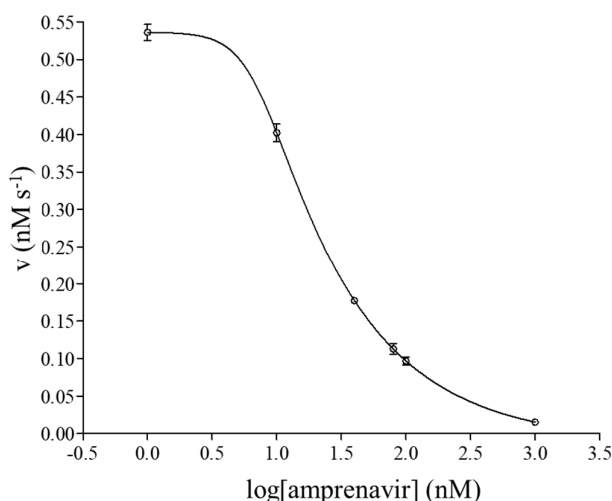
Figure 24. **Inhibition of HIV-1 PR by amprenavir.** Dose-response curve shows effect of amprenavir on the cleavage of His_6-MBP-VSQNY↓PIVQ-mTurquoise2 by HIV-1 PR. Error bars represent SD, n=2.

5.4. Kinetic and pH dependence studies of TEV PR

Kinetic measurements of TEV PR were performed on His_6-MBP-VSQNY↓PIVQ-mTurquoise2 and -mApple substrates. In accordance with the result of the kinetic measurement of HIV-1 PR, the specificity constants determined for TEV PR by different substrates were highly similar (**Table 15**) and were comparable with previously published value (**Kapust et al., 2002**). While k_{cat}/K_M values resembled that of determined by Kapust *et al.* (**2002**), individual kinetic parameters differ substantially, which can be explained by the conformational differences between the applied substrate types, similarly to case of the HIV-1 PR kinetic results. Furthermore, differences of the reaction conditions may also contribute to the difference in the kinetic parameters, as oligopeptide-based assay was performed at higher ionic strength (400 mM NaCl) and in the presence of reducing

agents (*e.g.* dithiothreitol), compared to Ni-NTA-magnetic bead-based assay, where 300 mM NaCl was applied in the absence of reducing agents.

Table 15. **Comparison of the kinetic parameters determined by different substrates for TEV PR.** * Value determined previously by Kapust *et al.* (**2002**).

Substrate	k_{cat} (s^{-1})	K_M (mM)	k_{cat}/K_M (mM^{-1} s^{-1})
His$_6$-MBP-VSQNY↓PIVQ-mTurquoise2	0.0050 ± 0.0001	0.0024 ± 0.0003	2.08 ± 0.26
His$_6$-MBP-VSQNY↓PIVQ-mApple	0.0038 ± 0.0001	0.0021 ± 0.0004	1.81 ± 0.35
TENLYFQ↓GGTRR (oligopeptide) *	0.27 ± 0.03	0.087 ± 0.017	3.08 ± 0.67

Experiments using TEV PR have given excellent opportunity to demonstrate the suitability of the Ni-NTA-agarose bead-based assay platform for investigating the performance of both the system and TEV PR activity at different pH conditions. The activity of TEV PR was followed at pH between 6.0-8.5 (**Figure 25**), and in accordance with the results of Parks et al., (**1995**), no significant differences were observed in the activities at pH 6.5- 8.5. Measured data at pH 6.0 however, was excluded from the evaluation, since the standard deviation (SD) between the parallel measurements was especially high. At pH 6.0 the interaction between Ni-NTA agarose beads and the His$_6$-tagged substrate weakens substantially and high rate of spontaneous intact substrate dissociation from the bead surface occurs as it was also depicted in the manufacturer's instructions of Ni-NTA magnetic agarose beads.
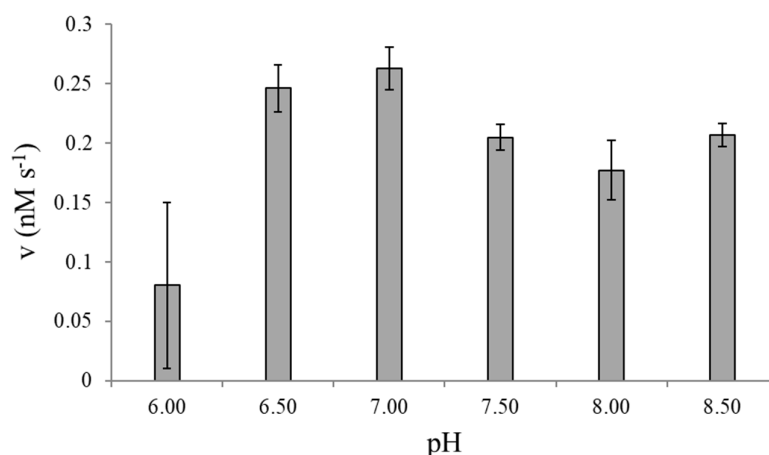
Figure 25. **Dependence of TEV PR activity on pH.** His$_6$-MBP-VSQNY↓PIVQ-mTurquoise2 substrate was used for activity measurements. Error bars represent SD, n=2.

The rate of spontaneous substrate dissociation from the surface of the Ni-NTA magnetic beads have been measured between pH 6.0-8.5 using His$_6$-MBP-VSQNY*PIVQ-mTurquoise2 substrate

(0.033 mM). The results indicate that >15 % of the substrates dissociate from the beads at pH 6.0 ([Figure 26](#)).
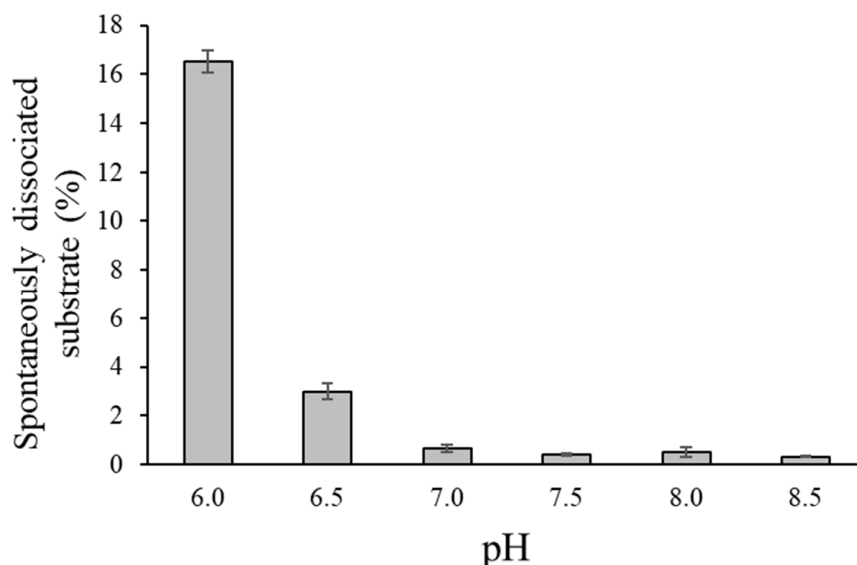
Figure 26. **Spontaneous dissociation of intact substrate from the magnetic bead surface at different pH.** His$_6$-MBP-VSQNY↓PIVQ-mTurquoise2 was applied in the experiment. Error bars represent SD, n=2.

5.5. VEEV PR constructs

The presented results have not been published yet; they are part of a manuscript that is under preparation for submission.

5.5.1. Expression vectors encoding different VEEV PR construct

A VEEV nsP2pro construct (further referred to as VEEV nsP2pro-1) was previously expressed to study the protease activity and showed unexpectedly low activity *in vitro* (**Zhang et al., 2009**). Our aim was to express new recombinant VEEV nsP2 constructs different from VEEV nsP2pro-1 and characterize their catalytic activity *in vitro* on recombinant fusion protein substrates. In this work, expression plasmids coding for two new VEEV PR constructs were prepared in an effort to generate a more active protease. (i) One, representing the almost full-length VEEV nsP2 protein (Gly1-Thr785), as according to Das *et al.* (**2014**), Chikungunya virus nsP2 activity was supposed to be dependent on contributions from both the N- and C-terminal domains and this suggests the possibility of an interaction or some kind of functional crosstalk between them. (ii) The other construct called VEEV nsP2pro-2, is a N-terminally slightly elongated version of the VEEV nsP2pro-1 by the Ala436-Met457 region. The rationale behind VEEV nsP2pro-2 design was that there is a

strong sequence conservation in the midst of the segment between residue 436-457 (marked with yellow on **Figure 27**), and it is possible, that this segment is structurally intertwined with the protease domain. Upstream to this segment, 357-436 region is the C-terminus of the most highly conserved region of the protein (corresponding to a RecA-like nucleotide-binding domain but also annotated as a helicase superfamily type I domain).

Beside engineering of the N-terminus, we decided to end the C-terminus of both constructs at Thr785, because that is the end of the density in the current structure. The schematic structure of the different VEEV PR constructs are illustrated on **Figure 28**. The ORFs of both VEEV nsP2 and VEEV nsP2pro-2 have been successfully cloned into both pDEST-His$_6$ and pDEST-His$_6$-MBP destination vectors according to Tropea *et al.* (**2007**).

```
VEEV         GSVETPRGLIKVTSYAGEDKIGSYAVLSPQAVLKSEKLSCIHPLAEQVIVITHSGRKGRYAVEPYHGKVVVPEGHAIPVQDFQALSESATIVYNEREFVNRYLHHIATHGGALNTDEEYY  120
Sindbis      ALVETPRGHVRIIPQANDRMIGQYIVVSPNSVLKNAKLAPAHPLADQVKIITHSGRSGRYAVEPYDAKVLMPAGGAVPWPEFLALSESATLVYNEREFVNRKLYHIAMHGPAKNTEEEQY  120
Chikungunya  GIIETPRGAIKVTAQPTDHVVGEYLVLSPQTVLRSQKLSLIHALAEQVKTCTHSGRVGRYAVEAYDGRVLVPSGYAISPEDFQSLSESATMVYNEREFVNRKLHHIAMHGPALNTDEESY  120
Sagiyama     GVVETPRNALKVTPQAHDHLIGSYLILSPQTVLKSEKLAPIHPLAEQVTVMTHSGRSGRYPVDKYDGRVLIPTGAAIPVSEFQALSESATMVYNEREFINRKLHHIALYGPALNTDEESY  120
Semliki      GVVETPRSALKVTAQPNDVLLGNYVVLSPQTVLKSSKLAPVHPLAEQVKIITHNGRAGRYQVDGYDGRVLLPCGSAIPVPEFQALSESATMVYNEREFVNRKLYHIAVHGPSLNTDEENY  120
             . :****. ::      :   :*.* ::**::**:. **:  * **:**     **.** *** *: *..:*:* * *:    :* :******:*******:** *:*** :* : **:** *

VEEV         KTVKPSEHDGEYLYDIDRKQCVKKELVTGLGLTGELVDPPFHEFAYESLRTRPAAPYQVPTIGVYGVPGSGKSGIIKSAVTKKDLVVSAKKENCAEIIRDVKKMKGLDVNARTVDSVLLN  240
Sindbis      KVTKAELAETEYVFDVDKKRCVKKEEASGLVLSGELTNPPYHELALEGLKTRPAVPYKVETIGVIGTPGSGKSAIIKSTVTARDLVTSGKKENCREIEADVLRLRGMQITSKTVDSVMLN  240
Chikungunya  ELVRAERTEHEYVDVDQRRCCKKEEAAGLVLVGDLINPPYHEFAYEGLKIRPACPYKIAVIGVFGVPGSGKSAIIKNLVTRQDLVTSGKKENCQEITTDVMRQRGLEISARTVDSLLLN  240
Sagiyama     EKVRAERAETEYVFDVDKKACIKKEEASGLVLTGDLINPPFHEFAYEGLKIRPAAPYHTTIIGVFGVPGSGKSAIIKNMVTTRDLVASGKKENCQEIMNDVKRQRGLDVTARTVDSILLN  240
Semliki      EKVRAERTDAEYVFDVDKKCCIKREEASGLVLVGELTNPPFHEFAYEGLKIRPSAPYKTTVVGVFGVPGSGKSAIIKSLVTKHDLVTSGKKENCQEIVNDVKKHRGLDIQAKTVDSILLN  240
             : .:. .   : **::*:*:: * *:* .:** * *:* :**:**:* *.*: **: **:   :** *.******.***. ** .***.*.***** ** ** : :*::: ::****::**

VEEV         GCKHPVETLYIDEAFACHAGTLRALIAIIRPK-KAVLCGDPKQCGFFNMMCLKVHFNH---EICTQVFHKSISRRCTKSVTSVVSTLFYDKKMRTTNPKETKIVIDTTGSTKPKQDDLIL  356
Sindbis      GCHKAVEVLYVDEAFACHAGALLALIAIVRPRKKVVLCGDPMQCGFFNMMQLKVHFNHPEKDICTKTFYKYISRRCTQPVTAIVSTLHYDGKMKTTNPCKKNIEIDITGATKPKPGDIIL  360
Chikungunya  GCNRPVDVLYVDEAFACHSGTLLALIALVRPRQKVVLCGDPKQCGFFNMMQMKVNYNH---NICTQVYHKSISRRCTLPVTAIVSSLHYEGKMRTTNEYNKPIVVDTTGSTKPDPGDLVL  357
Sagiyama     GCKRGVENLYVDEAFACHSGTLLALIALVRPSGKVVLCGDPKQCGFFNLMQLKVHYNH---NICTRVLHKSISRRCTLPVTAIVSTLHYQGKMRTTNPIQIDTTGSSKPASGDIVL  357
Semliki      GCRRAVDILYVDEAFACHSGTLLALIALVKPRSKVVLCGDPKQCGFFNMMQLKVNFNH---NICTEVCHKSISRRCTRPVTAIVSTLHYGGKMRTTNPCNKPIIIDTTGQTKPKPGDIVL  357
             **.: *: **:*******:*:* ****::::*  *.****** ******:* :**:**   :***.. :* ****** **::**:*.*  ***** :. * :* ** :**  .*:::*

VEEV         TCFRGWVKQLQIDYKGNEIMTAAASQGLTRKGVYAVRYKVNENPLYAPTSEHVNVLLTRTEDRIVWKTLAGDPWIKTLTAKYPGNFTATIEEWQAEHDAIMRHILERPDPTDVFQNKANV  476
Sindbis      TCFRGWVKQLQIDYPGHEVMTAAASQGLTRKGVYAVRQKVNENPLYAITSEHVNVLLTRTEDRLVWKTLQGDPWIKQPTNIPKGNFQATIEDWEAEHKGIIAAINSPTPRANPFSCKTNV  480
Chikungunya  TCFRGWVKQLQIDYRGYEVMTAAASQGLTRKGVYAVRQKVNENPLYASTSEHVNVLLTRTEGKLVWKTLSGDPWIKTLQNPPKGNFKATIKEWEVEHASIMAGICSHQMTFDTFQNKANV  477
Sagiyama     TCFRGWVKQLQIDYRGHEVMTAAASQGLTRKGVYAVRQKVNENPLYSPLSEHVNVLLTRTENRLVWKTLSGDPWIKVLTNVPRGDFSATLEEWQEEHDGIMRVLNERPAEVDPFQNKAKV  477
Semliki      TCFRGWVKQLQLDYRGHEVMTAAASQGLTRKGVYAVRQKVNENPLYAPASEHVNVLLTRTEDRLVWKTLAGDPWIKVLSNIPQGNFTATLEEWQEEHDKIMKVIEGPAAPVDAFQNKANV  477
             ***********:**.* * *:*************************** .:****** ******    *:* **::*: ** *: : :   : *. *:*

VEEV         CWAKALVPVLKTAGIDMTTEQWNTVD-YFETDKAHSAEIVLNQLCVRFFGLDLDSGLFSAPTVPLSIR-------NNHWDNSPSPNMYGLNKEVVRQLSRRYPQLPRAVATGRVYDMNTG  588
Sindbis      CWAKALEPILATAGIVLTGCQWSELFPQFADDKPHSAIYALDVICIKFFGMDLTSGLFSKQSIPLTYHPADSARPVAHWDNSPGTRKYGYDHAIAAELSRRFPVFQLA-GKGTQLDLQTG  599
Chikungunya  CWAKSLVPILETAGIKLNDRQWSQIIQAFKEDKAYSPEVALNEICTRMYGVDLDSGLFSKPLVSVYYA------DNHWDNRPGGKMFGFNPEAASILERKYPFTKGKWNINKQICVTTR  590
Sagiyama     CWAKCLVQVLETAGIRMTADEWNTIL-AFREDRAYSPEVALNEICTRYYGVDLDSGLFSAQSVSLFYE-------NNHWDNRPGGRMYGFNHEVARKYAARFPFLRGNMNSGLQLNVPER  589
Semliki      CWAKSLVPVLDTAGIRLTAEEWSTIITAFKEDRAYSPEVALNEICTKYYGVDLDSGLFSAPKVSLYYE-------NNHWDNRPGGRMYGFNAATAARLEARHTFLKGQWHTGKQAVIAER  590
             ****.*  :* **** :.  :*. .:    *: :*  .*: :* : :*:** *****   ::     **** *. ..:* :   .    :.        .   :

VEEV         TLRNYDPRINLVPVNRRLPHALVLHHNEHPQSDFSSFVSKLKGRTVLVV-GEKLSVPGKMVDWLSDRP--EATFRARLDLGIPGDVPKYDIIFVNVRTPYKYHHYQQCEDHAIKLSMLTK  705
Sindbis      RTRVISAQHNLVPVNRRLPHALVPEYKEKQPGPVKKFLNQFKHHSVLVVSEEKIEAPRKRIEWIAPIGIAGADKNYNLAFGFPPQ-ARYDLVFINIGTKYRNHHFQQCEDHAATLKTLSR  718
Chikungunya  RIEDFNPTTNIIPANRRLPHSLVAEHRPVKGERMEWLVNKINGHHVLLVSGYNLVLPTKRVTWVAPLGVRGADYTYNLELGLPATLGRYDLVVINIHTPFRIHHYQQCVDHAMKLQMLGG  710
Sagiyama     KLQPFSAECNIVPSNRRLPHALVTSYQQCRGERVEWLLKKIPGHQMLLVSEYNLAIPHKRVFWIAPPRVSGADRYDLDLGLPMDAGRYDLVFVNIHTEYRQHHYQQCVDHSMRLQMLGG  709
Semliki      KIQPLSVLDNVIPNRRLPHALVAEYKTVKGSRVEWLVNKVRGYHVLLVSEYNLALPRRRVTWLSPLNVTGADRCYDLSLGLPADAGRYDLVFVNIHTEFRIHHYQQCVDHAMKLQMLGG  710
             .   .     *::* **.***:*   :.       ..  ::.:    :*:*   ::   * : :*::    *      * *:*:*   :**::.:*: * ::  **:*** **:   *. *

VEEV         KACLHLNPGGTCVSIGYGYADRASESIIGAIARQFKFSRVCKPKSSLEETEVLFVFIGYDRKA-RTHNPYKLSSTLTNIYTGSRLHEAGC              794
Sindbis      SALNCLNPGGTLVVKSYGYADRNSEDVVTALARKFVRVSAARPDCVSSNTEMYLIFRQLDNSRTRQFTPHHLNCVISSVYEGTRDGVGA-             807
Chikungunya  DSLRLLKPGGSLLIRAYGYADRTSERVICVLGRKFRSSRALKPPCVTSNTEMFFLFSNFDNGR-RNFTTHVMNNQLNAAFVGQAT-RAGC             798
Sagiyama     DSLHLLRPGGSLLMRAYGYADRVSEMVVTALARKFSAFRVLRPACVTSNTEVFLLFSNFDNGR-RAVTLHQANQKLSSMYACNGLHTAGC             798
Semliki      DALRLLKPGGNLLMRAYGYADKISEAVVSSLSRKFSSARVLRPDCVTSNTEVFLLFSNFDNGK-RPSTLHQMNTKLSAVYAGEAMHTAGC             799
             .:.  *.***. :  .*****: ::   :.*:*    . :*  . .:**::** ::*   *.   *      . .: .    :.    :.   ..
```

Figure 27.   Alignment of alphavirus nsP2 sequences. Sequence in bold represents the VEEV nsP2pro-2, while the underlined sequence indicates VEEV nsP2pro-1. The segment formed by residue 436-457 is marked with yellow. The primary sequences of VEEV nsP2 (NCBI Ref. Seq.: NP_740697.1), SINV (NCBI Ref. Seq: NP_740671.1), chikungunya virus (GenBank: ADZ47896.1), sagiyama virus (GenBank: BAA92845.1) and SFV (NCBI Reference Sequence: NP_740666.1) were aligned using Multiple Sequence Alignment Tool Clustal Omega (EMBL-EBI, Hinxton). Highly conserved regions are marked with red.
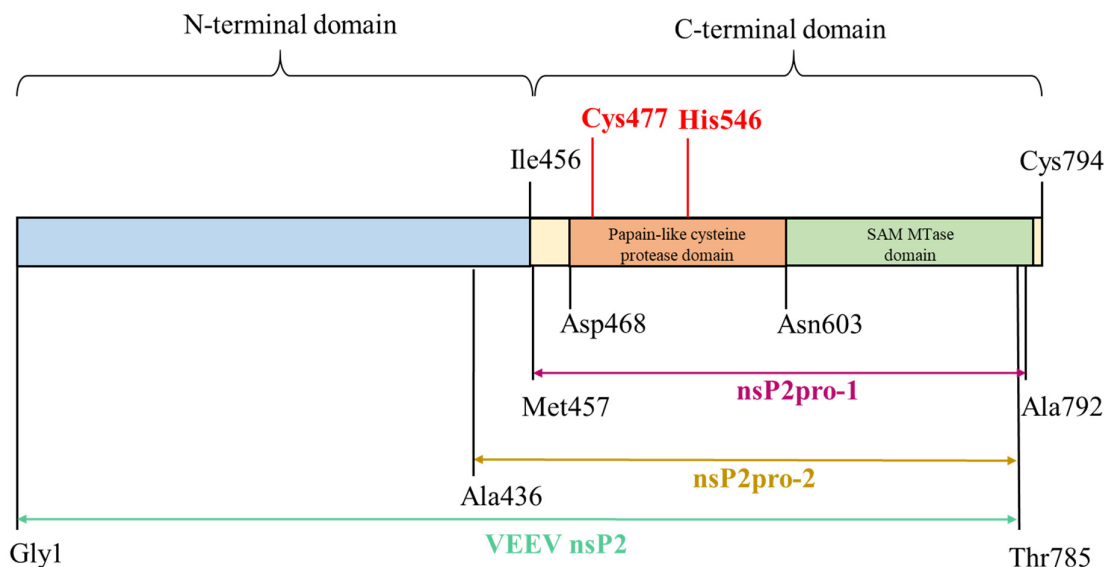
Figure 28. **Schematic structure of the full-length VEEV nsP2 and the corresponding enzyme construct.** VEEV nsP2pro-1 (purple) was applied in the previous studies (**Zhang _et al._, 2009; Hu _et al._, 2016**). Two different forms were designed, in which nsP2pro-1 is extended by (i) the whole N-terminal domain of nsP2 (VEEV nsP2, olive) and (ii) with residues 436-457 (VEEV nsP2pro-2, yellow). Catalytic residues of the protease are shown by red. SAM MTase=S-adenosyl-L-methionine-dependent RNA methyltransferase domain.

5.5.2. Expression and purification of VEEV PR constructs

The His_6- and His_6-MBP-fused VEEV PR constructs were expressed in _E. coli_ Rozetta cells. Protein expression profile of the cells right before (t=0 h) and after induction by IPTG (t=4 h), as well as the solubility profile of total protein content have been compared between the different constructs after bacterial cell lysis (**Figure 29**). In case of VEEV nsP2pro-2, both the His_6- and the His_6-MBP fusion forms were overexpressed by the IPTG-induced bacterial cells after 4 hours of incubation at 30 °C, however only the His_6-MBP-fusioned form was confirmed to be soluble after the clarification of the crude bacterial cells lysate (**Figure 29 B and D**).

In case of full length VEEV nsP2, the expression of the His_6- and His_6-MBP-attached forms were barely or slightly detectable and none of the constructs were found to be pronouncedly expressed over the host cell proteins using the same incubation conditions as in the cases of the VEEV nsP2pro-2 constructs. Furthermore, none of the VEEV nsP2 forms were detected in the cleared bacterial cell lysate (**Figure 29 A and C**). We would have liked to see if the expression and solubility of the His_6-MBP-attached form could be improved by an overnight incubation after IPTG induction at 18 °C, but no improvement has been achieved (**Figure 29 E**).
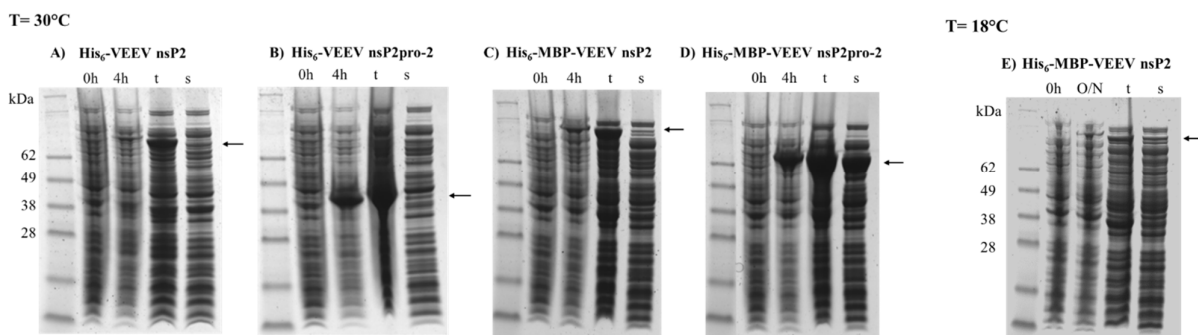
Figure 29. **Expression and solubility profile of VEEV nsP2 and nsP2pro-2 fusion construct.** Samples corresponds to the expression profile before (0h) and after 4 hours (4h) or overnight (O/N) incubation of the non-permeabilized IPTG induced cells, and the total (t) and soluble (s) protein fractions after cell permeabilization in case of A) His_6-VEEV nsP2, B) His_6-VEEV nsP2pro-2 C) His_6-MBP-VEEV nsP2 and D) His_6-MBP- VEEV nsP2pro-2 expressed at 30 °C and E) His_6-MBP-VEEV nsP2 expressed at 18 °C in *E. coli* Rozetta cells. Arrows show the expected molecular weight of the fusion proteins: His_6-VEEV nsP2 (91.7 kDa), His_6-VEEV nsP2pro-2 (43.3 kDa), His_6-MBP-VEEV nsP2 (131.3 kDa) and His_6-MBP-VEEV nsP2pro-2 (83.0 kDa).

According to the expression and solubility profile of the different VEEV PR constructs, we have decided to proceed with the purification of the His_6-MBP-fused form of VEEV nsP2pro-2 only. The VEEV nsP2pro-2 was successfully purified in a three-step downstream procedure starting with a Ni-NTA affinity chromatography, during which His_6-MBP-VEEV nsP2pro-2 has been purified from the crude cell lysate (**Figure 30**).
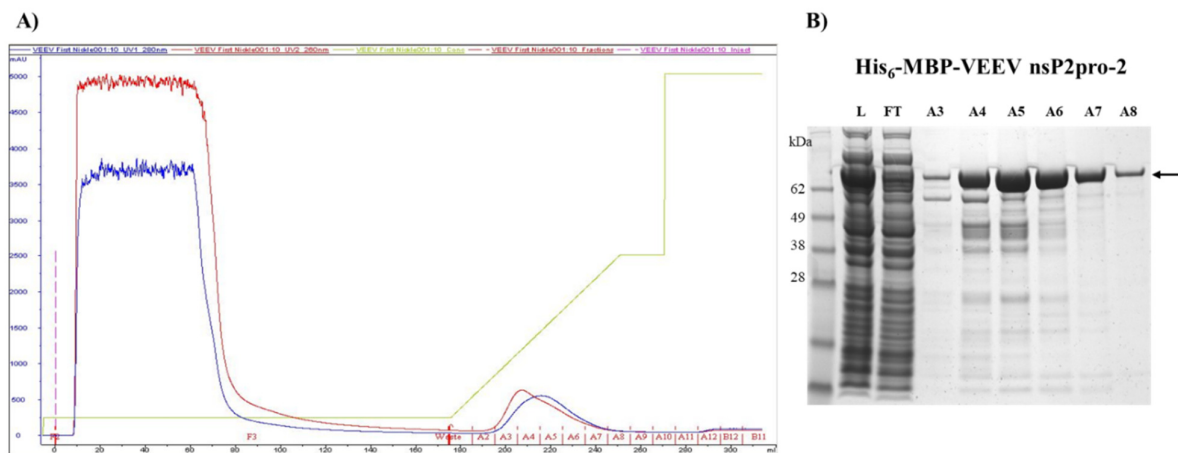
Figure 30. **Purification of His_6-MBP-VEEV nsP2pro-2 by 1st Ni-chelate affinity chromatography.** A) Cleared and filtered bacterial cell lysates were loaded on HisTrap FF crude affinity column equilibrated in buffer A, and then eluted with a linear gradient from 5-50 % buffer B. B) The load (L), the fractions of the flowthrough (FT) and the eluates (A4-A8) containing the His_6-MBP-nsP2pro-2 were analyzed on reducing SDS-PAGE. Fraction A4-A8 were pooled together to be further purified. Arrow shows His6-MBP-VEEV nsP2pro-2.

Next, the buffer-exchanged eluate was processed by TEV PR to remove the His$_6$-MBP- dual affinity tag from VEEV nsP2pro-2 (**Figure 31 A**).Cleavage reaction was used as a load in a second Ni-NTA affinity chromatography, during which the N-terminal His$_6$-MBP cleavage products were attached to the Ni-NTA column, and the C-terminal cleavage product VEEV nsP2pro-2 has been recovered from the flow-through (**Figure 31 B and C**). Finally, the concentrated flow-through was further purified by size-exclusion chromatography to get rid of the remaining impurities (**Figure 32**). The final product was judged to be at least 95 % pure by SDS-PAGE.
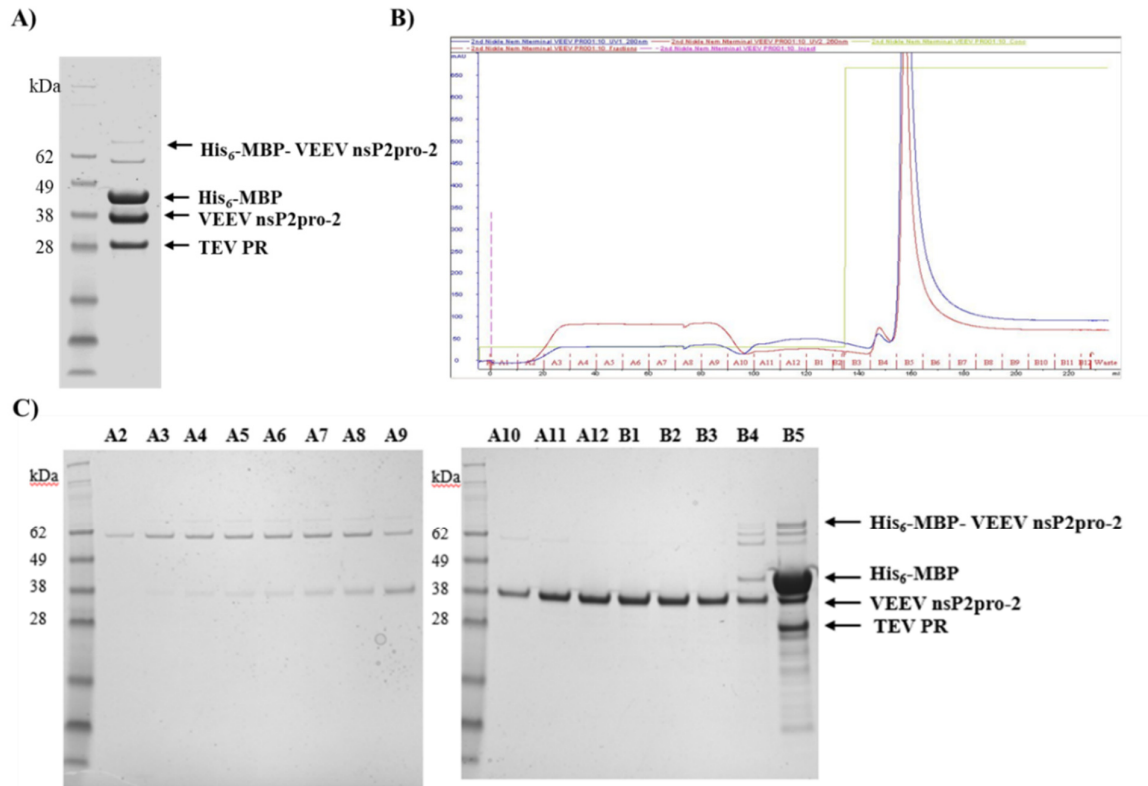
Figure 31.   Purification of VEEV nsP2pro-2 by the 2nd Ni-chelate affinity chromatography. A) His$_6$-MBP-VEEV nsP2pro-2 was digested by His-tagged TEV PR (5 mg/mL; 70:1 v/v) overnight at 4 °C, and the cleavage products were analyzed by reducing SDS-PAGE. B) The digestion mixture was loaded on a HisTrap FF crude affinity column equilibrated with buffer A. Flow-through fractions, containing the nsP2pro protease (A2-B3), and eluate fractions (B4-B12) were collected and C) were analyzed on reducing SDS-PAGE. Fractions A10-B4 were pooled and applied for the subsequent size-exclusion chromatography.
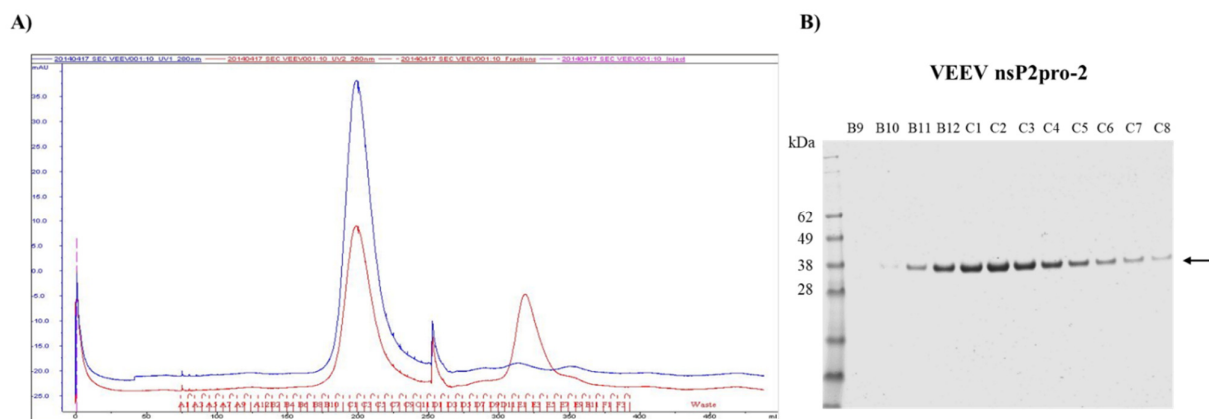
Figure 32. Purification of VEEV nsP2pro-2 by size-exclusion chromatography. A) The concentrated eluate of the 2nd Ni-chelate affinity chromatography was loaded onto a HiPrep 26/60 Sephacryl S200 column equilibrated with 50 mM Tris- HCl buffer (containing 150 mM NaCl, 2 mM TCEP, pH 7.5). B) The peak fractions (B9-C8) containing the protease were analyzed on reducing SDS-PAGE. Fractions B10-C8 were pooled and concentrated to 2.9 mg/mL final concentration. Arrow shows VEEV nsP2pro-2.

5.6. Specificity study of VEEV nsP2pro-2

In the literature, most of the previously performed mutagenesis studies were focusing on testing SINV and SFV protease activity prominently on their cognate substrates mutated at P5-P1' (**Zhang et al, 2009**). In a previous study of our laboratory (**Zhang et al, 2009**) VEEV nsP2pro-1 showed activity *in vitro* on oligopeptide substrate modelling P6-P6' positions of the natural cleavage site of SFV nsP1/nsP2 (SFV-1, EYHAGA↓GVVETP). Our current aim was to investigate and characterize VEEV nsP2pro-2 activity on His$_6$-MBP-mEYFP fusion protein substrates comprising cleavage sites representing either the wild-type or modified SFV-1 sequence. SFV-1 variants were designed by the substitution of residues at P5, P4, P2, P1, P1', and P2' sites.

Due to its closeness to the site of the cleavage and as it represents the very first residue of the generated C-terminal product, the importance of the Gly residue in P1' position was investigated extensively by testing all the possible amino acid substitution in this position. Furthermore, other SFV-1 variants have also been generated according to the primary sequence of other related alphavirus strains. These variants include the following modifications: P5-Gln, P4-Glu, P4-Thr, P4-Arg, P4-Gly, P2-Ala, P2-Val, P1-Gly, P1-Val, P2'-Pro, and P2'-Ser. The method for the generation and the list of the expression vectors coding for the different VEEV PR fluorescent fusion protein substrates were described in **Section 5.1** and in **Table 12**, respectively. All the substrates tested in the specificity study was fused to mEYFP.

As a first step, we have compared the cleavability of the different variants to that of the wild-type SFV-1 substrate by using the 96-well plate adaptation of the developed Ni-NTA agarose bead-based protease assay. For this purpose, the expression of the analyzed substrates has been successfully optimized to smaller volume (15 mL) in contrast to the regularly applied volume (50 mL) in the previous experiments with HIV-1 PR and TEV PR. The results of the microplate-based experiments revealed that the substrate conversion rates for all of the P1' variants have found to be substantially lower compared to the wild-type (**Figure 33 A**). Regarding the VEEV nsP2pro-2 activity on variants modified at other positions, the rate of substrate conversion catalyzed by the enzyme have been remarkably increased in case of P4-Glu, P4-Thr but only slightly elevated in case of P4-Arg substitution compared to wild-type substrate. Based on the unpaired *t*-test, the difference to the wild-type in case of P4-Glu and P4-Thr variants were regarded as statistically significant, in contrast for P4-Arg, where the difference was not statistically significant. Moderate decrease in the substrate conversion rates has been observed on P4-Gly, P1-Gly and P2'-Pro, P2'-Ser variants compared to the wild-type, respectively, while the activity of VEEV nsP2pro-2 has been almost completely or completely inhibited in substrates possessing P5-Gln, P2-Ala and P2-Val substitutions, respectively (**Figure 33 B**). For all these cases, the statistical probe resulted in statistically significant differences compared to the wild-type. The results of one-way ANOVA were in accordance with that of the unpaired *t*-test in each case. Parallel to the *in vitro* specificity measurement, *in silico* calculations were also performed in order to predict changes of enzyme-substrate interactions. The mutations that correspond the *in vitro* investigated amino acid substitutions were introduced into the modeled complex of VEEV nsP2pro (encompassing 469-767 residues) and a peptide substrate representing wild-type SFV-1 cleavage site (EYHAGA↓GVVETP). The corresponding results are summarized in **Table 16** . The modeled structure of the VEEV nsP2pro complexed with the wild-type SFV-1 is illustrated in **Figure 34.**
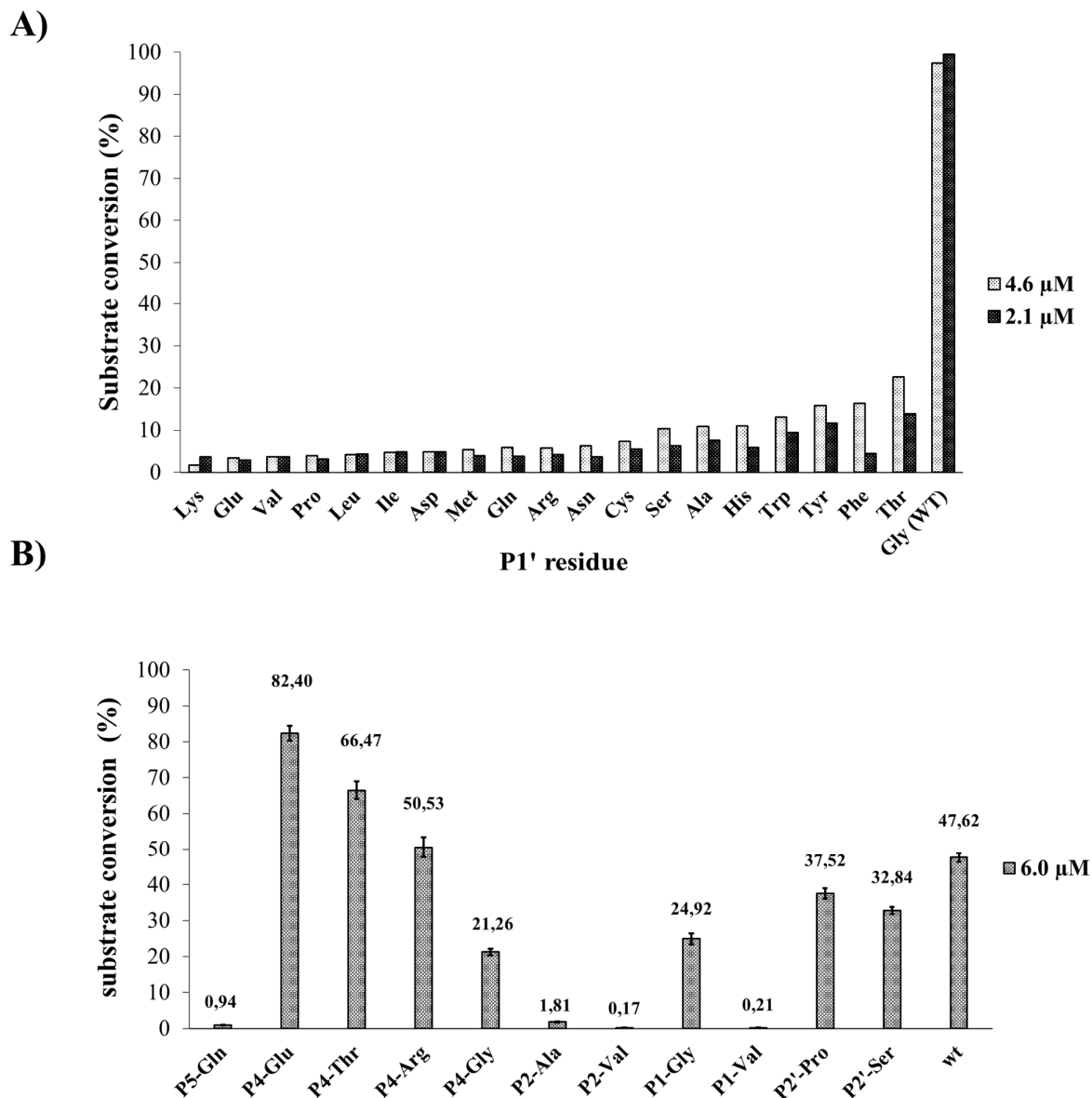
A)

B)

Figure 33.   Cleavage of substrate variants containing SFV nsP1/nsP2 cleavage site by VEEV nsP2pro-2. A) P1' variants were cleaved at 4.6 µM (white) and 2.1 µM (black) final enzyme concentrations at substrate concentration between 1.4-2.6 µM and 3.1-5.0 µM, respectively. The runs were performed on different plates, on different days, from substrate batches expressed and handled under same conditions but on different days. In the plate of each run no parallel samples were measured. Capital letters on the X axis indicate the P1'residue. B) P5, P4, P1, and P2 variants were cleaved at 6.0 µM final enzyme concentration. The substrate concentration was ranging between 3.0 -5.3 µM. Error bars represent SD, n=4. wt=wild-type.

Table 16. *In silico* **predicted folding free energy changes**. The predicted change of the free energy is listed, the ΔΔG (kcal/mol) changes in the complex of VEEV PR (encompassing 469-767 residues) with SFV-1 (EYHAGA↓GVVETP) substrate upon point mutations were determined by using DynaMut web server. ΔΔG >0 changes were regarded as stabilizing. Results were obtained by calculations performed by Dr. János András Mótyán.

SFV-1 variant	ΔΔG (kcal/mol)	SFV-1 variant	ΔΔG (kcal/mol)
P5-Gln	-0.222	P1'-Arg	1.166
P4-Glu	0.175	P1'-Leu	1.83
P4-Thr	-0.133	P1'-Lys	1.336
P4-Arg	-0.2	P1'-Glu	1.506
P4-Gly	-0.349	P1'-Tyr	2.082
P2-Ala	-0.536	P1'-Ile	1.085
P2-Val	0.495	P1'-Gln	1.05
P1-Gly	0.188	P1'-Asn	1.962
P1-Val	1.207	P1'-Asp	0.971
P1'-Thr	1.15	P1'-Ala	1.107
P1'-Val	1.314	P1'-Met	0.883
P1'-Pro	0.875	P1'-Trp	2.477
P1'-His	1.269	P1'-Phe	2.085
P1'-Cys	1.718	P2'-Pro	0.427
P1'-Ser	1.731	P2'-Ser	-0.068
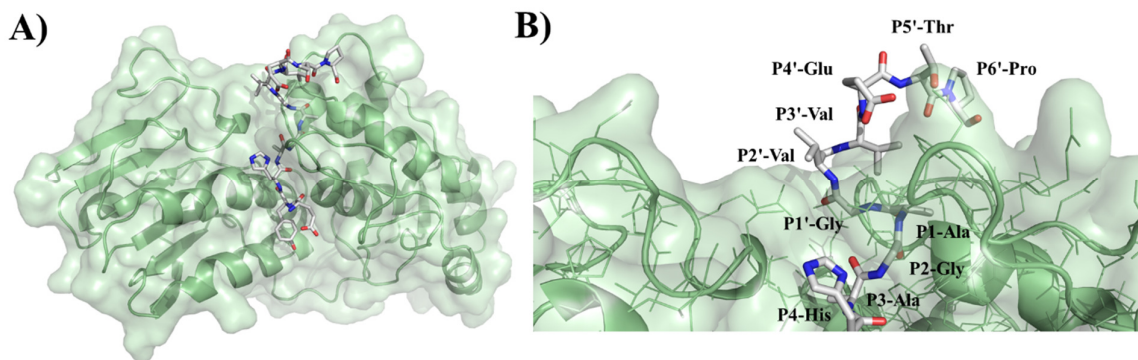
Figure 34. **VEEV nsP2pro complexed with wild-type SFV-1. A)** Modeled structure of VEEV PR (encompassing 469-767 residues) complexed with EYHAGA*GVVETP substrate. **B)** SFV-1 substrate bound to the active site is enlarged, substrate residues are labelled. Figure was prepared by Dr. János András Mótyán, using PyMOL Molecular Graphics System - Version 1.3 Schrödinger, LLC).

5.7. Kinetic measurements of VEEV nsP2pro-2

On SFV-1 substrate variants, in case of which the rate of substrate conversion allowed, kinetic measurements were performed using the developed microcentrifuge tube-based Ni-NTA magnetic bead-based assay platform to characterize and compare each proteolytic reaction by measured kinetic parameters.

In case of the expression of the desired fusion protein substrate variants for the kinetic measurements, the rate of matured mEYFP was improved by arresting the protein translation of the IPTG-induced cell with tetracycline in the middle of the 4-hour-long incubation period.

Time course studies on the wild-type SFV-1 substrate (His$_6$-MBP-EYHAGA↓GVVETP-mEYFP) have indicated that the activity of the VEEV nsP2pro-2 starts decreasing substantially after 15 minutes at 30 °C (**Figure 35**). Accordingly, the reaction time for the substrate-dependent kinetic measurement were needed to be set to 10 min. Therefore, kinetic parameters could only be determined for those substrates the cleavage of which resulted in reliably detectable fluorescent signal in concentration under enzyme saturation at 10 min reaction time (**Figure 36**).

His$_6$MBP-EYHAGA↓GVVETP-mEYFP
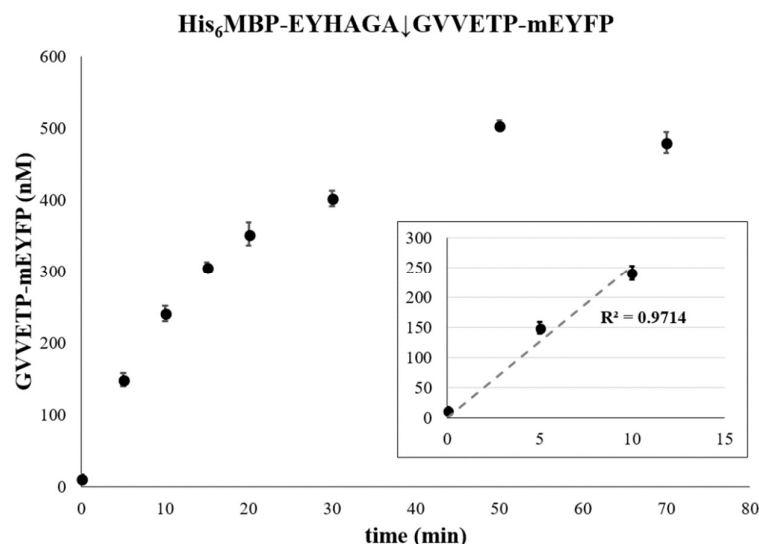
Figure 35. **Time course studies of VEEV nsP2pro-2.** The amount of the C-terminal fluorescent product released from His$_6$-MBP-EYHAGA↓GVVETP-mEYFP (at 21.2 µM final concentration) upon cleavage by VEEV nsP2pro-2 (at 3.0 µM final total concentration) is illustrated in the function of time. Zoom-in graph represents the timeframe, in which the enzyme activity is close to linear with time. Error bars represent SD, n=2.
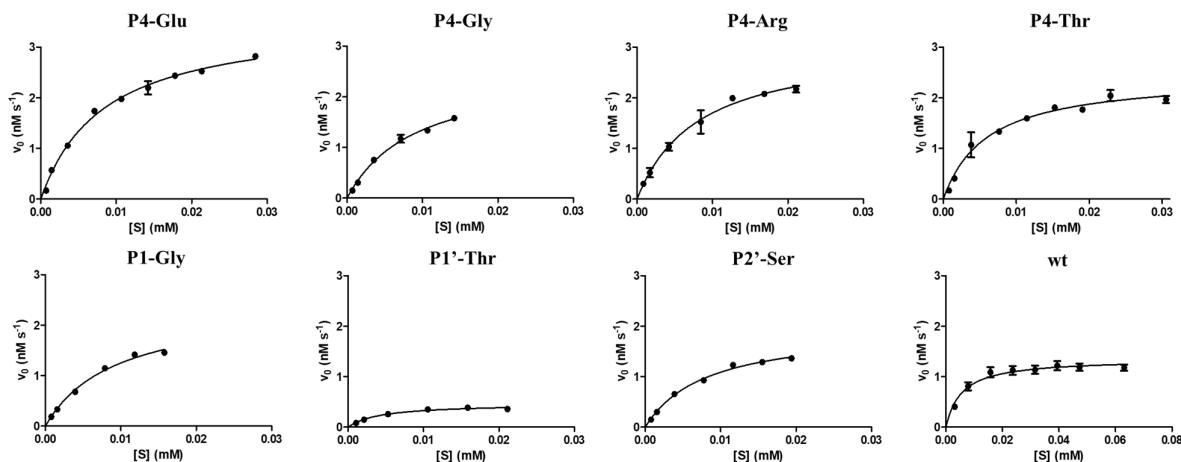
Figure 36. **Michaelis–Menten plots of velocity versus substrate concentration**. Graphs illustrate the cleavage results of His$_6$-MBP-EYHAGA↓GVVETP-mEYFP substrate (wild-type=wt) and its P4-Glu, P4-Thr, P4-Arg, P4-Gly, P1-Gly, P2'-Ser and P1'-Thr variants by VEEV nsP2pro-2. The hydrolysis of the substrates was ≤ 20 % at each concentration point. Error bars represent SD, n=2. The SD values are indicated on each graph, but in some cases they were smaller than the marker itself.

Appling the listed conditions, we have successfully determined the k_{cat}, K$_M$ and k_{cat}/K$_M$ values for VEEV nsP2pro-2 on the wild-type and P4-Glu, P4-Thr, P4-Ala, P4-Gly, P1-Gly, P1'-Thr, and P2'-Ser variants of His$_6$-MBP-SFV-1-mEYFP recombinant substrate (**Table 17**). In agreement with results of the microplate-based specificity results, P4-Glu variant showed the highest catalytic efficiency, which was approximately 5-fold higher compared to the wild-type substrate and this difference was determined as statistically significant. Variant P4-Thr and P4-Arg were found to be processed 3.7- and 1.7-fold more efficiently compared to wild-type SFV-1, respectively. For P4-Thr, this was considered as a significant difference, while for P4-Arg the difference was considered to be not statistically significant. The k_{cat}/K$_M$ values of variants including P4-Gly, P1-Gly, and P2-Ser were calculated to be within the lower error range of the k_{cat}/K$_M$ value of the wild-type cleavage site and accordingly the statistical probe did not result in any significant difference. In contrast, the catalytic efficiency on P1'-Thr variant was found to be approximately 4-fold lower than that of the native sequence, which was considered as a statistically significant difference.

Kinetic parameters in the literature for VEEV PR on the wild-type SFV-1 sequence were compared to the herein measured data. It can be concluded that the k_{cat}/K$_M$ value of the different measurements are within the same order of magnitude, while both k_{cat} and K$_M$ values determined for VEEV nsP2pro-2 on wild-type His$_6$-MBP-SFV-1- mEYFP were substantially lower compared to those measured by Zhang *et al.* (**2009**) and Hu *et al.* (**2016**) for VEEV nsP2pro-1 using oligopeptide substrates and recombinant CFP/YFP FRET substrates, respectively.

79

Table 17. **Comparison of VEEV nsP2pro kinetic parameters.** Kinetic parameters of VEEV nsP2pro-2 were measured on His_6-MBP-FP fusion protein substrates containing the wild-type and modified SFV nsP1/nsP2 cleavage site sequences and are compared to historical VEEV nsP2pro-1 kinetic parameters determined on different substrate types containing wild-type SFV nsP1/nsP2 cleavage site. Modified residues are marked in red. * and ** indicated values published previously by Hu *et al.* (**2016**) and Zhang *et al.* (**2009**), respectively.

VEEV nsP2pro-2			
His₆-MBP-mEYFP-based fusion protein substrate			
Cleavage site sequence	k_{cat} (s⁻¹)	K_M (mM)	k_{cat}/K_M (mM⁻¹ s⁻¹)
EYEAGA↓GVVETP	0.0025 ± 7.0E-05	0.0083 ± 6.0E-04	**0.29 ± 0.024**
EYTAGA↓GVVETP	0.0013 ± 6.0E-05	0.0060 ± 8.0E-04	**0.22± 0.032**
EYRAGA↓GVVETP	0.0008 ± 4.0E-05	0.0081 ± 0.001	**0.10 ± 0.015**
EYGAGA↓GVVETP	0.0005 ± 3.0E-05	0.0091 ± 0.001	**0.05 ± 0.008**
EYHAGG↓GVVETP	0.0005 ± 3.0E-05	0.0092 ± 0.001	**0.05 ± 0.007**
EYHAGA↓GSVETP	0.0005 ± 2.0E-05	0.0081 ± 8.0E-04	**0.06 ± 0.006**
EYHAGA↓TVVETP	0.00006 ± 4.0E-06	0.0042 ± 8.0E-04	**0.02 ± 0.003**
EYHAGA↓GVVETP (wt)	0.0003 ± 1.0E-05	0.0056 ± 9.0E-04	**0.06 ± 0.010**
VEEV nsP2pro-1			
recombinant CFP/YFP FRET fusion protein substrate			
DVEELEYHAGA↓GVVETP*	0.033 ± 0.002	0.23 ± 0.02	**0.144 ± 0.01**
AETGVVDVDVEELEYHAGA↓GVVETP*	0.050 ± 0.003	0.6 ± 0.2	**0.083 ± 0.03**
Oligopeptide substrate			
EYHAGA↓GVVETP**	0.016 ± 0.001	0.58 ± 0.09	**0.028 ± 0.005**

6. Discussion

Due to their central role in the lifecycle and homeostasis of living organisms and their inevitable biotechnological importance, proteases and their substrates are intensively investigated in the field of academics, therapeutics and industry. Accordingly, the establishment and improvement of powerful, cost-efficient and labor-saving methods for the characterization of their catalytic activity and specificity are of great demand.

This study aimed the development, testing and application of an affinity chromatography-based fluorescent protease assay platform which is based on the use of recombinant fusion protein substrates. In the designed system, each substrate comprises N-terminal His$_6$ and MBP fusion tags, a TEV PR processing site, a cleavage site of interest introduced via a 'cloning cassette' of the expression plasmid, and a C-terminal fluorescent protein. The His$_6$ tag enables the binding of the substrates Ni-NTA-coated magnetic beads surface. Upon proteolytic cleavage, both uncleaved substrates and N-terminal cleavage products remain attached to the beads, while the C-terminal fluorescent products are released into the supernatant. The fluorescent proteins can be detected in the supernatant after their separation from the magnetic beads. The principle of the assay has already been utilized earlier for the development of similar protease assay by Patel *et al*. (**2001**) Chaparro-Riggers *et al* (**2005**). These methods are also based on recombinant protein substrates containing an affinity tag (for substrate immobilization) and a fluorescent fusion tag (for detection) besides the proteolytic cleavage sites. The bead-attached substrates and the protease cleavage-released products can be quantified by fluorimetry. The substrate designed by Patel et al. (**2001**) contains cleavage site for TEV protease, while the assay of Chaparro-Riggers *et al.* has been used for the determination of kinetic parameters for the studied enzyme, although the linearity of the detected fluorescent *versus* protein concentration was not verified, and the methodology by of protein content determination after elution of the substrate by imidazole was not described. In contrast, our system may provide a novel tool for the investigation of proteases by combining and improving the advantages of some already existing methods by offering a true separation-based assay, offering the determination of enzyme kinetic parameters (k_{cat}, K_M) supported with a detailed substrate quantification procedure. Our assay offers a low volume HTS-compatible format and examines the linearity of the fluorescence *versus* substrate concentration in a wide concentration range at different buffer conditions.

The great advantages of our assay platform are provided by the highly beneficial properties of the recombinant substrates. In accordance with the strong nature of His$_6$-tag and Ni-NTA interaction, the assay is characterized by low level of impurities and high signal-to-noise ratio, while by the integration of MBP into the construct the substrates are less prone to the possible influence of surface-immobilization on the enzyme kinetic parameter as in this format the cleavage sites of protease of

interest is not located in the close proximity of the affinity binding site to the Ni-chelate surface. Additionally, MBP highly improves the folding and the solubility of recombinant substrates. The interdomain linker between the MBP and the FP provides sufficient accessibility of the cleavage site. Additionally, it contains the built-in canonical TEV PR cleavage site, as well, which makes investigation of TEV PR possible, and may also serve as a control cleavage site which can be used in the identification of cleavage position (**Gazda et al., 2020**). It is also the interdomain linker that embrace the cleavage site of interest, inserted at the DNA level onto the 'cloning cassette' of the substrate expression plasmid. The flexibility of the 'cloning cassette' allows the insertion of wide variety of the sequences by a simple one-step PCR reaction of designed oligonucleotide primers or by random mutagenesis. The inexpensive and efficient generation of the fluorescent fusion substrates harboring the cleavage site of interest, makes the system especially attractive for the generation of substrate libraries.

The C-terminal end of the protein substrate contains a fluorescent protein, which provides highly sensitive detection of the intact substrates and released C-terminal cleavage product. Furthermore, FP expression can be followed easily during expression, purification, and working procedures. By calibration of substrates' fluorescence under different buffer conditions, both the substrates and the generated C-terminal products can be quantified. This is the prerequisite for accurate determination of kinetic parameters. In addition, a single calibration curve can be applied for a set of substrates bearing the same fluorescent tag, which simplifies the quantification of series of substrates during specificity measurements. In our experiments expression vectors of the substrates were designed to code for several different type of fluorescent proteins including ECFP, mTurquoise2, EGFP, mEYFP, mApple, and mCherry. Since 3′ and 5′ ends of the DNA sequences of the most commonly used fluorescent proteins are highly similar, listed expression vectors can be prepared by using only a pair of oligonucleotide primers (N1 and C). In order to avoid substrate aggregation, we used substrates possessing monomeric fluorescent protein forms including mTurquoise2, mEYFP, mApple, and mCherry. The flexibility of the platform allows the selection of the fluorescent proteins that most suit the experimental purposes and the given instrumentation.

The utility of Ni-NTA magnetic bead-based protease assay platform developed on the recombinant substrates have been successfully demonstrated in a variety of applications. Time course, substrate-dependent and inhibition studies performed on TEV and HIV-1 PRs have demonstrated that the system is suitable for the determination of enzyme kinetic parameters. In both cases, the determined k_{cat}/K_M values were found to be highly comparable with those ones obtained by using synthetic oligopeptide-based protease assay systems, while individual k_{cat} and K_M values were substantially lower. This difference can be interpreted by the conformational differences of the

substrates, as oligopeptide substrates represents the binding sites in an extended form, in contrast to recombinant fusion protein substrate, where the cleavage site is likely to be folded in a different conformation and its flexibility can be affected by the different protein domains.

Additionally, the system has proven to be applicable for investigation of different reaction conditions such as enzyme activity at different alkaline pH conditions, as we have demonstrated in case of TEV PR. However, the utility of the Ni-NTA magnetic bead-based platform for these kinds of purposes may be limited by compatibility of the magnetic beads with the buffer components and also by the pH-dependent nature of the effective attachment of the substrates to the bead-surface. According to the manufacturer's handbook, monomer proteins start to be released from the bead surface at pH 5.9, whereas multimers are released at pH 4.5. We have observed increased spontaneous dissociation of substrates at pH 6.0, in some cases lower pHs may also be applied, however, the rate of spontaneous substrate dissociation at each reaction must be followed by appropriate control reactions and must be considered at the evaluation of the results.

Beside the described 'on-bead applications', it is also possible to study proteolytic cleavage of the purified substrates by in-solution digestion. This option may be of use if the applied buffer conditions may circumvent the interactions between the His$_6$ tag and the affinity beads. In this case, the uncleaved substrates can be separated from the cleavage fragments by PAGE analysis. According to our results, at non-denaturing conditions the fluorescence components are readily visible by naked eye and by blue or UV light illumination immediately after electrophoretic separation on polyacrylamide gels, even in the presence of SDS.

If denaturing conditions are applied during sample preparation, the fluorescent species can be detected after in-gel renaturation, *i.e.* by the removal of SDS from the gel. To our knowledge, a protocol for the detection of fluorescent proteins after a denaturing SDS-PAGE analysis, including heat-treatment and washing out the SDS from the gel, has not been published until now. We found that application of subsequent washing steps with distilled water is sufficient for at least partial renaturation of the proteins, by which the fluorescence of the previously denatured substrates and products were partially restored and were successfully detected in the gel by UV imaging. The advantage of protein denaturation/renaturation in this protease assay is that the cleavage products can be separated from each other based on their molecular weight, and their native charge or shape does not affect their migration during electrophoresis, however after renaturation the appropriate bands can be identified based on their fluorescence. This can be highly beneficial if the molecular weights of other cleavage products (or contaminants) that have no fluorescence closely resemble those of the fluorescent products. Relatively small (~2 kDa) differences can also be detected by separation of cleavage products, which may be indicative if cleavages occur at alternative sites, as well. Notably,

there were differences observed in the renaturation ability of applied fluorescent proteins, as dsRed-derived fluorescent proteins (mApple and mCherry) showed substantially lower renaturation ability compared to the GFP-derived ones including mTurqouise2 and mEYFP. According to Gross *et al.* (**2000**) the acylimine bond of the dsRed chromophore is irreversibly degraded upon denaturation. The acylimine bond, being responsible in part for the red shift of mApple (**Shu et al., 2006**) may be disrupted upon denaturation and this change is supposed to be responsible for the lower renaturation ability of mApple and mCherry. Detailed analysis of the changes induced by denaturation may help understanding the differences between the sensitivity of the fluorescent proteins, however, the investigation of renaturation abilities of the different fluorescent proteins was out of the scope and extent of this study. Therefore, the renaturation abilities of proteins should be considered during the experimental design and if required fluorescent proteins with higher stability need be chosen or in-gel renaturation needs to be further optimized to make the detection of less-stable proteins sensitive enough.

The fluorescent fusion substrates may potentially be useful in additional applications. For example, the recombinant substrates may be applied for the detection of protease activities of crude cell lysates. However, it needs to be tested whether the substrates are susceptible for cleavage by endogenous cellular proteases at unwanted sites. The susceptibility of a recombinant substrate towards digestion by trypsin have been tested (**Mótyán et al., 2018**), and their results implied proper folding of the expressed substrates, because no degradation of the fusion partners was observed. This indicates that MBP and the fluorescent protein are not likely to be processed by non-specific host cell proteases, however, the interdomain linkers can be sensitive towards proteolysis due to their accessibility.

We have proven the utility of the system in one of our on-going projects in collaboration with MCL, Center for Cancer Research, NCI (MD, USA) aiming to characterize alphavirus nsP2 protease *in vitro* activity and specificity. In this present study we aimed to generate two VEEV PR constructs in which the sequence of C-terminal protease domain was extended by (i) the whole N-terminal domain (VEEV nsP2) and by (ii) Ala436-Met457 region (VEEV nsP2pro-2). Although the experimental conditions were not investigated extensively during bacterial expression in Rozetta cells, we have found that at 30 °C for 4 hours incubation after IPTG induction, both the His$_6$- and the His$_6$-MBP fusioned VEEV nsP2pro-2 forms were overexpressed, however only the His$_6$-MBP-fusioned form was confirmed to be soluble. These results are consistent with the previously described phenomena, that N-terminal MBP-fusion facilitates water-solubility of the recombinant proteins and improves folding (**Kapust et al., 1999; Fox et al., 2002, 2003**). In case of full length VEEV nsP2, both His$_6$- and His$_6$-MBP-attached forms were not pronouncedly overexpressed and none of the

constructs were found to be soluble. Unfortunately, no improvement has been achieved by lowering the temperature to 18 °C and lengthen of the incubation time to overnight after IPTG induction. Based on this, we believe that in case of the His$_6$ –fused form the lack of solubility was due to inherent structural properties and folding, which may have been improved by the attachment of MBP to the construct as it was seen in the case of His$_6$-MBP-nsP2pro-2, however in this case it is highly likely that the relatively large size of His$_6$-MBP-VEEV nsP2 (~131 kDa) overwhelms the advantageous effect of MBP on the folding and hindered successful bacterial expression in Rozetta cells at the investigated fermentation conditions.

VEEV nsP2pro-2 have been successfully purified from its His$_6$-MBP attached form, at the very same purification conditions and laboratory instruments that were applied during the purification of nsP2pro-1 by Zhang *et al.* (**2009**). Ni-NTA magnetic bead-based time and substrate-dependent measurement were used to calculate the kinetic parameters of VEEV nsP2pro-2 on mEYFP-fused fusion protein containing P6-P6' residues of the wild-type SFV-1 cleavage site.

The kinetic parameters were compared to the historical kinetic results of the VEEV nsP2pro-1 on the very same cleavage site sequence modelled by oligopeptide substrates (**Zhang et al., 2009**) or CFP/YFP-based recombinant FRET substrates (**Hu et al., 2016**). The difference seen between the K$_M$ values is most probably occurred as a consequence of the conformational dissimilarities among the applied substrate types, as it has been observed in the case of TEV PR and HIV-1 PR kinetic measurements as well. The k_{cat} and k_{cat}/K$_M$ values cannot be compared directly to each other due to the lack of any selective inhibitor, by which the exact active enzyme concentration could be determined at the different assay conditions. However, if enzyme activity is considered as 100 % at each reaction conditions, k_{cat}/K$_M$ values are within the same order of magnitude, which suggests that the catalytic efficiency of VEEV nsP2pro is similar on the different substrate constructs comprising the wild-type cleavage site sequence of SFV-1. Basically, these results indicate that the attachment of the highly conserved segment built by 436-457 residues to VEEV nsP2pro domain did not cause any substantial change in the activity of the VEEV protease domain *in vitro*.

We have used VEEV nsP2pro-2 enzyme for the *in vitro* characterization VEEV nsP2pro specificity. For this purpose, Ni-NTA bead-based platform has been successfully adapted to a 96-well microplate format and was used to screen the activity of VEEV protease on a series of SFV-1 recombinant fusion protein substrate variants designed based on the modification of the residues in P5, P4, P2, P1, P1' and P2' positions. Following the micro-plate-based specificity studies, kinetic parameters were determined for the enzyme on the variants with well-measurable substrate conversion rates using the Ni-NTA magnetic bead-based assay platform. The results of the kinetic measurements were in good accordance with the conversion rates of the microplate-based study,

indicating the suitability of the 96-well-based adaptation of Ni-NTA magnetic bead-based assay system for ranking the hits based on their catalytic efficiency if the utilized substrate concentration is in the dynamic range of the enzyme of interest.

Specificity screening results are interpreted in the light of the corresponding *in silico* studies performed by Dr. János András Mótyán. He had predicted the stability changes of the SFV-1 - VEEV nsP2pro (encompassing 469-767 residues) model complex upon the different amino acid substitutions of SFV-1 by using DynaMut server.

In our *in vitro* microplate-based specificity study, the P1' site was investigated to the greatest extent, owing to its close proximity to the site of the cleavage and as it represents the N-terminal residue of the released product. The results indicate no strong correlation between the predicted energy changes and the *in vitro* observed conversion rates. Furthermore, no direct correlation was found between the hydrophobicity or volume of P1' residues and *in vitro* P1' specificity, therefore most probably the overall hydrophobicity of the entire recognition site may rather be a determinant of P1' specificity. These findings are in accordance with those of Russo *et al*. (**2010**), who described that, the protease contacts mainly the main-chain atoms of the substrate at the S1' site. However, the interactions seem to be not independent from the side-chain, we found besides the wild-type Gly, the most preferred P1' residues have mainly polar (Thr, Ser) or aromatic (Tyr, Trp, Phe) side-chain. In accordance with this, most considerable *in silico* energy changes were predicted for these latter aromatic mutants. It is highly likely that aromatic moiety enables favorable hydrophobic interactions with a main-chain atom of Leu-665 residue of S1' subsite.

Upon the substitution of P5-Tyr to Gln, the *in vitro* observed conversion was negligible in accordance with the *in silico* predicted destabilizing change. It is very likely that wild-type P5-Tyr forms hydrogen-bonds (H-bond) with both Ser-731 and Ser-701 at S5 subsite, while similar interactions are not formed in the P5-Gln variant.

In case of P4 position the wild-type His was replaced by Glu, Thr, Arg and Gly residues. The *in silico* predicted changes of free energy among the produced P4 mutant showed linear correlation (R^2=0.89) with the *in vitro* obtained k_{cat}/K_M values. Among all the tested variants, the catalytic efficiency of VEEV nsP2pro was shown to be the highest on P4-Glu mutant, which is actually the wild-type P4 residue of VEEV nsP1/nsP2 and nsP2/nsP3 cleavage site and its side-chain is supposed to be connected to Lys-706 residue of the S4 subsite *via* a salt bridge (**Russo et al., 2010**; **Hu et al., 2016**). The other tested P4 variants were processed with significantly higher (P4-Thr) or with the same (P4-Arg, P4-Gly) efficiency as compared to that of the wild-type, but the processing was significantly less efficient than in case of P4-Glu, which can be interpreted by the lack of side-chain-mediated polar interactions at S4 site in case of Thr, Arg and Gly residues at P4.

The *in vitro* determined conversion rates for both P2-Val and P2-Ala variants were marginal compared to wild-type comprising a highly conserved Gly residue recognized by the 'glycine specificity motif' of VEEV nsP2pro, similarly to the mechanism described for other cysteine proteases (**Golubtsov et al. 2006**). We hypothesized that both (i) the relatively larger sizes of Ala and Val residues and (ii) the lack of a H-bond formed by backbone atom of Gly may contribute to the lower preference of these mutants.

The conversion rate of P1-Gly variant was half that of the wild-type P1-Ala. The lower conversion of Gly compared to Ala, can be interpreted by the preference of hydrophobic S1-P1 interaction at this site. In contrast, although Val has larger volume and higher hydrophobicity than Ala, and accordingly *in silico* analysis predicted more favorable interactions for this variant compared to Gly, still the *in vitro* observed conversion for this mutant was negligible. This may be interpreted by the findings of Russo *et al.* (**2010**), the branched side-chains have lower flexibility that may make their binding to S1 site less favorable. Furthermore, they described that covariances of residues indicate relatively high tolerance for amino acid substitutions at P1 site. Correspondingly, the *in vitro* determined k_{cat}/K$_M$ values were very close to each other in case of the wild-type and the P1-Gly mutant.

It has already been described that mainly S4-S1' sites play the most crucial role in substrate recognition (**Lulla et al., 2006**; **Russo et al., 2010**). Accordingly, corresponding P4-P1' sites have been extensively investigated, however, the information about P2' site is limited. Based on the model structure, S2' site is not a well-defined pocket, because P2'-P6' residues of the bound substrate are solvent-exposed. Here we modified the wild-type P2'-Val residue to Ser and to Pro. *In silico* calculated energy changes predicted slightly lower interaction between the protease and the substrate upon P2'-Ser mutation, while elevated free energy for P2'-Pro variant. In contrast, the *in vitro* observed conversion values were fairly close to each other for the two P2 site-mutants and were about 1.5-fold lower compared to the wild-type. Kinetic measurements, performed only on the Ser variant, resulted in a k_{cat}/K$_M$ value that is practically identical to that of the wild-type, although both the residue volume and hydrophobicity of the wild-type P2'-Val is higher than that of the Ser. These findings can be interpreted partly (i) by the previously described fact, that solvent-exposed nature of P2' side-chain does not enable formation of polar interactions at S2' site and (ii) by backbone- rather than side-chain-mediated nature of S2'-P2' interactions. The latter concept is supported by the findings of Hu *et al.* (**2016**), who suggested favorable backbone interactions for the P1'-P6' sites.

7. Summary

In summary, we presented the development and the utility of a Ni-NTA magnetic bead-based HTS-compatible fluorescent protease assay platform using recombinant protein substrates. The substrates are composed of N-terminal His$_6$ and MBP fusion tags, TEV PR cleavage site sequence, the recognition site of the protease of interest and a C-terminal fluorescent protein. In the expression plasmid of the substrate, there is a cloning cassette between the TEV PR recognition site and FP domain. This cassette allows the one-step insertion of a short dsDNA sequence, which can be generated *via* random mutagenesis or oligonucleotide primers and can code for the protease cleavage site of interest. The expression of the fluorescent substrate was successfully demonstrated at different scales, with or without arresting the protein translation with tetracycline. To test the on-bead applicability of the expressed substrates, enzyme kinetic measurements, inhibition, and pH optimum studies were performed by using TEV and HIV-1 proteases. Beside the microbead-based fluorimetric studies, substrates and their cleavage products were also assayed by PAGE. We found that the denatured proteins could be renatured after reducing SDS-PAGE, and the fluorescent proteins were detected in the gel upon blue light and/or UV illumination. Interestingly, different fluorescent proteins were found to have different ability for renaturation. The standardized working instruction and a corresponding tutorial video of the Ni-NTA magnetic bead-based assay and PAGE analysis were prepared to support the execution of the assay procedure.

The platform has been further optimized to microplate-based specificity studies and has been successfully applied to characterize the *in vitro* activity and specificity of the purified VEEV nsP2pro-2 construct using mEYFP-fused SFV-1 substrate variants modified at P5, P4, P2, P1, P1' and P2' positions. Kinetic parameters were determined for seven variants and for the wild-type. The results indicated that VEEV nsP2pro-2 has processed P4-Glu, P4-Thr variants with significantly higher, P4-Arg, P4-Gly, P1-Gly and P2'-Ser variants with not significantly different, while P1'-Thr variant with significantly lower efficiency compared to wild-type substrate. VEEV nsP2pro-2 did not show improved catalytic performance on the wild-type SFV-1 cleavage site compared to VEEV nsP2pro-1, previously purified and examined by our collaborators.

Beside its herein verified academical applications, the developed assay platform can be adaptable to HTS and automation-based environment and thus can be of use in pharmaceutical protease substrate discovery, inhibitor screening and/or drug development, whereas the results of VEEV PR specificity studies could contribute to the investigation of alphaviruses or other important Group IV viruses.

8. Összefoglalás

Munkánk célja egy rekombináns fehérje szubsztrátokra optimalizált Ni-NTA mágneses gyöngy-alapú fluoreszcens proteáz vizsgálati módszer kifejlesztése és alkalmazása volt. Az tervezett szubsztrátok tartalmaznak egy N-terminális His_6 és egy MBP fúziós címkét, a TEV PR és a vizsgálni kívánt proteáz hasítási szekvenciáját, valamint egy C-terminális fluoreszcens fehérjét (FP). A szubsztrátokat expresszáló vektorba egy klónozó kazettát terveztünk, mely lehetővé teszi a vizsgálni kívánt hasítási szekvenciát kódoló, rövid DNS szakaszok egylépésben történő beillesztését.

A mágneses gyöngyhöz immobilizált szubsztrátok alkalmazhatóságának igazolására enzimkinetikai, gátlási és pH optimum meghatározási vizsgálatokat végeztünk a TEV és HIV-1 PR-okra. A mikrocső-alapú mérésekhez kidolgozott módszert a mérési térfogat csökkentése és az áteresztőképesség növelése érdekében mikrotiter lemezre is optimalizáltuk. A fluorimetriás mérések mellett a szubsztrátok és a termékek elválasztását elektroforézissel is elvégeztük, és kidolgoztunk egy eljárást, mellyel a redukáló körülmények között végzett SDS-poliakrilamid gél elektroforézist követően a denaturált fluoreszcens fehérjék a gélben renaturálhatóak, majd a fluoreszcencia alapján detektálhatóak. A különböző fluoreszcens fehérjék újrafeltekeredési képessége eltérő volt. A mágneses gyöngy-alapú és a gél elektroforézissel történő vizsgálatokhoz egy részletes módszerleírást, valamint egy kapcsolódó oktatóvideót készítettünk, melyek lehetővé teszik a módszer elsajátítását és kivitelezését. Az általunk kifejlesztett vizsgálati módszer tovább optimalizálható nagy áteresztőképességű és automatizált mérésekhez, és alkalmazható proteáz szubsztrátok azonosítására, gátlószerek vizsgálatára és fejlesztésére, akár gyógyszeripari környezetben is.

Célunk volt továbbá egy katalitikusan hatékonyabb VEEV PR enzimforma létrehozása is. A kifejlesztett módszert alkalmazva vizsgáltuk a létrehozott VEEV nsP2pro-2 proteáz *in vitro* aktivitását és specificitását. Ehhez olyan His_6-MBP-mEYFP rekombináns szubsztrátokat hoztunk létre, melyek az SFV-1 (EYHAGA↓GVVETP) hasítóhely vad típusú, illetve P5, P4, P2, P1, P1', vagy P2' variánsait tartalmazták. A vad típusú szubsztrát és hét variánsa esetében enzimkinetikai paramétereket is meghatároztunk. A mérések alapján a VEEV nsP2pro-2 enzimforma a P4-Gln és P4-Thr variánsokat szignifikánsan jobban, míg a P1'-Thr variánst szignifikánsan kisebb katalitikus hatékonysággal hasította a vad típusú szubsztráthoz képest, míg a P4-Arg, P4-Gly, P1-Gly és P2'-Ser mutánsok esetében nem tapasztaltunk szignifikáns különbséget. VEEV nsP2pro-2 nem bizonyult hatékonyabbnak a korábban már vizsgált VEEV nsP2pro-1 formához képest, azonban a specificitási vizsgálatok eredményei hozzájárulhatnak a további, alfavírusokat vagy egyéb IV. genetikai csoportba tartozó vírusokat célzó kutatásokhoz.

9. References

1. Adams MJ, Antoniw JF, Beaudoin F. Overview and analysis of the polyprotein cleavage sites in the family Potyviridae. *Mol Plant Pathol*. 2005;6(4):471-487.

2. Askin SP, Morin I, Schaeffer PM. Development of a protease activity assay using heat-sensitive Tus–GFP fusion protein substrates. *Anal Biochem*. 2011;415(2):126-133.

3. Bahudhanapati H, Zhang Y, Sidhu SS, Brew K. Phage Display of Tissue Inhibitor of Metalloproteinases-2 (TIMP-2): Identification Of Selective Inhibitors Of Collagenase-1 (Metalloproteinase 1 (Mmp-1)). *J Biol Chem*. 2011;286(36):31761-31770.

4. Balleza E, Kim JM, Cluzel P. Systematic characterization of maturation time of fluorescent proteins in living cells. *Nat Methods*. 2018;15(1):47-51.

5. Barrett AJ, Rawlings ND. Evolutionary lines of cysteine peptidases. *Biol Chem*. 2001;382(5):727-733.

6. Barrett AJ. Protein degradation in health and disease. Ciba Foundation Symposium 75, 1980. p1. Excerpta Medica, Amsterdam.

7. Bazan JF, Fletterick RJ. Viral cysteine proteases are homologous to the trypsin-like family of serine proteases: structural and functional implications. *Proc Natl Acad Sci USA*. 1988;85(21):7872-7876.

8. Berger A, Schechter I. Mapping the Active Site of Papain with the Aid of Peptide Substrates and Inhibitors. *Philos Trans R Soc Lond B Biol Sci*. 1970;257(813):249-264.

9. Bergmann M, Ross WF. On proteolytic enzymes X. The enzymes of papain and their activation. *J. Biol. Chem*. 1936;114:717-726.

10. Blommel PG, Fox BG. Fluorescence anisotropy assay for proteolysis of specifically labeled fusion proteins. *Anal Biochem*. 2005;336(1):75-86.

11. Boeneman K, Mei BC, Dennis AM, et al. Sensing Caspase 3 Activity with Quantum Dot–Fluorescent Protein Assemblies. *J Am Chem Soc*. 2009;131(11):3828-3829.

12. Boulware KT, Jabaiah A, Daugherty PS. Evolutionary optimization of peptide substrates for proteases that exhibit rapid hydrolysis kinetics. *Biotechnol Bioeng*. 2010:n/a-n/a.

13. Branchini BR, Rosenberg JC, Ablamsky DM, Taylor KP, Southworth TL, Linder SJ. Sequential bioluminescence resonance energy transfer–fluorescence resonance energy transfer-based ratiometric protease assays with fusion proteins of firefly luciferase and red fluorescent protein. *Anall Biochem*. 2011;414(2):239-245.

14. Bronze MS, Huycke MM, Machado LJ, Voskuhl GW, Greenfield RA.Viral agents as biological weapons and agents of bioterrorism. *Am J Med Sci*. 2002;323(6):316-325.

15. Callahan BP, Stanger MJ, Belfort M. Protease Activation of Split Green Fluorescent Protein. *Chem Eur J of Chem Bio*. 2010;11(16):2259-2263.

16. Carrington JC, Dougherty WG. Processing of the tobacco etch virus 49K protease requires autoproteolysis. *Virology*. 1987a;160(2):355-362.

17. Carrington JC, Dougherty WG. Small nuclear inclusion protein encoded by a plant potyvirus genome is a protease. *J Virol*. 1987b;61(8):2540-2548.

18. Cesaratto F, Burrone OR, Petris G. Tobacco Etch Virus protease: A shortcut across biotechnologies. *J Biotechnol*. 2016;231:239-249.

19. Cesaratto F, López-Requena A, Burrone OR, Petris G. Engineered tobacco etch virus (TEV) protease active in the secretory pathway of mammalian cells. *J Biotechnol*. 2015;212:159-166.

20. Chaparro-Riggers JF, Breves R, Michels A, Maurer K-H, Bornscheuer U. A GFP-based assay for the determination of hydrolytic activity and substrate specificity of subtilisins under washing conditions. *Journal of Molecular Catalysis B: Enzymatic*. 2005;35(1-3):74-77.

21. Choi SI, Song HW, Moon JW, Seong BL. Recombinant enterokinase light chain with affinity tag: Expression fromSaccharomyces cerevisiae and its utilities in fusion protein technology. *Biotechnol Bioeng*. 2001;75(6):718-724.

22. Cordingley MG, Callahan PL, Sardana VV, Garsky VM, Colonno RJ. Substrate requirements of human rhinovirus 3C protease for peptide cleavage in vitro. *J Biol Chem*. 1990;265(16):9062-9065.

23. Cushman DW, Pluscec J, Williams NJ, Weaver ER, Sabo EF, Kocy O, Cheung HS, Ondetti MA. Inhibition of angiotensin-coverting enzyme by analogs of peptides from Bothrops jararaca venom. *Experientia*. 1973;29(8):1032-1035.

24. Das PK, Merits A, Lulla A. Functional Cross-talk between Distant Domains of Chikungunya Virus Non-structural Protein 2 Is Decisive for Its RNA-modulating Activity. *J Biol Chem*. 2014;289(9):5635-5653.

25. Deardorff ER, Forrester NL, Travassos da Rosa AP, et al. Experimental Infection of Potential Reservoir Hosts with Venezuelan Equine Encephalitis Virus, Mexico. *Emerg Infect Dis*. 2009;15(4):519-525.

26. Diamond SL. Methods for mapping protease specificity. *Curr Opin Chem Biol*. 2007;11(1):46-51.

27. Dougherty WG, Parks TD, Cary SM, Bazan JF, Fletterick RJ. Characterization of the catalytic residues of the tobacco etch virus 49-kDa proteinase. *Virology*. 1989;172(1):302-310.

28. Dougherty WG, Parks TD. Post-translational processing of the tobacco etch virus 49-kDa small nuclear inclusion polyprotein: identification of an internal cleavage site and delimitation of VPg and proteinase domains. *Virology*. 1991;183(2):449-456.

29. Drag M, Salvesen GS. Emerging principles in protease-based drug discovery. *Nat Rev Drug Discov*. 2010;9(9):690-701.

30. Fehér A, Boross P, Sperka T, Miklóssy G, Kádas J, Bagossi P, Oroszlan S, Weber IT, Tőzsér J. Characterization of the murine leukemia virus protease and its comparison with the human immunodeficiency virus type 1 protease. *J Gen Virol*. 2006;87(5):1321-1330.

31. Fenner G. Das Genauigkeitsmaß von Summen, Differenzen, Produkten und Quotienten der Beobachtungsreihen. *Naturwissenschaften*. 1931;19:310.

32. Ferrer-Miralles N, Domingo-Espín J, Corchero J, Vázquez E, Villaverde A. Microbial factories for recombinant pharmaceuticals. *Microb Cell Fact*. 2009;8(1):17.

33. Fox JD, Routzahn KM, Bucher MH, Waugh DS. Maltodextrin-binding proteins from diverse bacteria and archaea are potent solubility enhancers. *FEBS Lett*. 2003;537(1-3):53-57.

34. Fox JD, Waugh DS. Maltose-Binding Protein as a Solubility Enhancer. In: *E. Coli Gene Expression Protocols.* Vol 205. New Jersey: Humana Press; 2002:99-118.

35. Freedberg DI, Ishima R, Jacob J, et al. Rapid structural fluctuations of the free HIV protease flaps in solution: Relationship to crystal structures and comparison with predictions of dynamics calculations. *Protein Science*. 2009;11(2):221-232.

36. Freund J, Kellner R, Konvalinka J, Wolber V, Krausslich H-G, Kalbitzer HR. A possible regulation of negative factor (Nef) activity of human immunodeficiency virus type 1 by the viral protease. *Eur J Biochem*. 1994;223(2):589-593.

37. Fujinaga M, Cherney MM, Oyama H, Oda K, James MNG. The molecular structure and catalytic mechanism of a novel carboxyl peptidase from Scytalidium lignicolum. *Proc Natl Acad Sci USA*. 2004;101(10):3364-3369.

38. Gaedigk-Nitschko K1, Schön A, Wachinger G, Erfle V, Kohleisen B. Cleavage of recombinant and cell derived human immunodeficiency virus 1 (HIV-1) Nef protein by HIV-1 protease. *FEBS Lett*. 1995 Jan 9;357(3):275-278.

39. Gallo RC, Montagnier L. AIDS in 1988. *Sci Am*. 1988 Oct;259(4):41-48.

40. Gammon ST, Villalobos VM, Roshal M, Samrakandi M, Piwnica-Worms D. Rational design of novel red-shifted BRET pairs: Platforms for real-time single-chain protease biosensors. *Biotechnol Progress*. 2009;25(2):559-569.

41. Gasteiger E, Hoogland C, Gattiker A, et al. Protein Identification and Analysis Tools on the ExPASy Server. In: Walker JM, ed. The Proteomics Protocols Handbook. Totowa, NJ: Humana Press; 2005:571-607.

42. Gautam Sarath, Zeece MG, Penheiter AR. Protease assay methods. In Proteolytic enzymes: a practical approach. Edited by Beynon RJ, Bond JS. Oxford ; New York : Oxford University Press, 2001.

43. Gazda LD, Joóné Matúz K, Nagy T, Mótyán JA, Tőzsér J. Biochemical characterization of Ty1 retrotransPoson protease. *PLoS ONE*. 2020:15(1): e0227062.

44. Golubtsov A, Kääriäinen L, Caldentey J. Characterization of the cysteine protease domain of Semliki Forest virus replicase protein nsP2 by in vitro mutagenesis. *FEBS Lett.* 2006;580(5):1502-1508.

45. Gomez de Cedrón M, Ehsani N, Mikkola ML, García JA, Kääriäinen L. RNA helicase activity of Semliki Forest virus replicase protein NSP2. *FEBS Lett.* 1999;448(1):19-22.

46. Gross LA, Baird GS, Hoffman RC, Baldridge KK, Tsien RY. The structure of the chromophore within DsRed, a red fluorescent protein from coral. *Proc Natl Acad Sci USA.* 2000;97(22):11990-11995.

47. Grzonka Z, Jankowska E, Kasprzykowski F, Kasprzykowska R, Lankiewicz L, Wiczk W, Wieczerzak E, Ciarkowski J, Drabik P, Janowski R, Kozak M, Jaskólski M, Grubb A. Structural studies of cysteine proteases and their inhibitors. *Acta Biochim Pol.* 2001;48(1):1-20.

48. Hartley BS. Proteolytic Enzymes. *Annu Rev Biochem.* 1960;29:45-72.

49. Hebisch E, Knebel J, Landsberg J, Frey E, Leisner M. High Variation of Fluorescence Protein Maturation Times in Closely Related Escherichia coli Strains. Hofmann A, ed. *PLoS ONE.* 2013;8(10):e75991.

50. Hornak V, Okur A, Rizzo RC, Simmerling C. HIV-1 protease flaps spontaneously open and reclose in molecular dynamics simulations. *Proc Natl Acad Sci USA.* 2006;103(4):915-920.

51. Horton RA, Strachan EA, Vogel KW, Riddle SM. A substrate for deubiquitinating enzymes based on time-resolved fluorescence resonance energy transfer between terbium and yellow fluorescent protein. *Anal Biochem.* 2007;360(1):138-143.

52. Hu X, Compton JR, Leary DH, et al. Kinetic, Mutational, and Structural Studies of the Venezuelan Equine Encephalitis Virus Nonstructural Protein 2 Cysteine Protease. *Biochemistry.* 2016;55(21):3007-3019.

53. Ido E, Han HP, Kezdy FJ, Tang J. Kinetic studies of human immunodeficiency virus type 1 protease and its active-site hydrogen bond mutant A28S. *J Biol Chem.* 1991;266(36):24359-66.

54. Ivanov KI, Eskelin K, Lohmus A, Makinen K. Molecular and cellular mechanisms underlying potyvirus infection. *J Gen Virol.* 2014;95(Pt_7):1415-1429.

55. Johnson KM, Martin DH. Venezuelan equine encephalitis. *Adv Vet Sci Comp Med.* 1974;18:79–116.

56. Kantor R. Human Immunodeficiency Virus Reverse Transcriptase and Protease Sequence Database: an expanded data model integrating natural language text and sequence analysis programs. *Nucleic Acids Res.* 2001;29(1):296-299

57. Kapust RB, Tözsér J, Copeland TD, Waugh DS. The P1′ specificity of tobacco etch virus protease. *Biochem Biophys Res Commun.* 2002;294(5):949-955.

58. Kapust RB, Tözsér J, Fox JD. Tobacco etch virus protease: mechanism of autolysis and rational design of stable mutants with wild-type catalytic proficiency. *Protein Eng*. 2001;14(12):993-1000.

59. Kapust RB, Waugh DS. *Escherichia coli* maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. *Protein Sci*. 1999;8(8):1668-1674.

60. Kim Y-P, Daniel WL, Xia Z, Xie H, Mirkin CA, Rao J. Bioluminescent nanosensors for protease detection based upon gold nanoparticle–luciferase conjugates. *Chem Commun*. 2010;46(1):76-78.

61. Klein T, Eckhard U, Dufour A, Solis N, Overall CM. Proteolytic Cleavage—Mechanisms, Function, and "Omic" Approaches for a Near-Ubiquitous Posttranslational Modification. *Chem Rev*. 2018;118(3):1137-1168.

62. Kohl T, Heinze KG, Kuhlemann R, Koltermann A, Schwille P. A protease assay for two-photon crosscorrelation and FRET analysis based solely on fluorescent proteins. *Proc Natl Acad Sci USA*. 2002;99(19):12161-12166.

63. Koonin EV, Dolja VV, Krupovic M. Origins and evolution of viruses of eukaryotes: The ultimate modularity. *Virology*. 2015;479-480:2-25.

64. Kostallas G, Löfdahl P-Å, Samuelson P. Substrate Profiling of Tobacco Etch Virus Protease Using a Novel Fluorescence-Assisted Whole-Cell Assay. Jones D, ed. *PLoS ONE*. 2011;6(1):e16136.

65. Liu Y, Song Y, Madahar V, Liao J. Quantitative Förster resonance energy transfer analysis for kinetic determinations of SUMO-specific protease. *Anal Biochem*. 2012;422(1):14-21.

66. Liu Z, Wang Y, Brunzelle J, Kovari IA, Kovari LC. Nine crystal structures determine the substrate envelope of the MDR HIV-1 protease. *Protein J*. 2011;30(3):173-183.

67. López-Otín C, Bond JS. Proteases: Multifunctional Enzymes in Life and Disease. *J Biol Chem*. 2008;283(45):30433-30437.

68. López-Otín C, Overall CM. Protease degradomics: A new challenge for proteomics. *Nat Rev Mol Cell Biol*. 2002;3(7):509-519.

69. Luciw PA. Human immunodeficiency viruses and their replication. In Virology. Edited by Fields BN; Knipe DM; Howley PM. Philadelphia, Lippincott-Raven, 1996.

70. Lulla A, Lulla V, Tints K, Ahola T, Merits A. Molecular determinants of substrate specificity for Semliki Forest virus nonstructural protease. *J Virol*. 2006; 80(11):5413-5422.

71. Malakhov MP, Mattern MR, Malakhova OA, Drinker M, Weeks SD, Butt TR. SUMO fusions and SUMO-specific protease for efficient expression and purification of proteins. *J Struct Func Genom*. 2004;5(1/2):75-86.

72. Martin SF, Hattersley N, Samuel IDW, Hay RT, Tatham MH. A fluorescence-resonance-energy-transfer-based protease activity assay and its use to monitor paralog-specific small ubiquitin-like modifier processing. *Anal Biochem*. 2007;363(1):83-90.

73. Michaelis L, Menten ML. Die Kinetik der Invertinwirkung. *Biochemische Zeitschrift.* 1913;49:333–369

74. Miladi B, El Marjou A, Boeuf G, Bouallagui H, Dufour F, Di Martino P, Elm'selmi A. Oriented immobilization of the tobacco etch virus protease for the cleavage of fusion proteins. *J Biotechnol*. 2012;158(3):97-103.

75. Morrison JF. Kinetics of the reversible inhibition of enzyme-catalysed reactions by tight-binding inhibitors. *Biochim Biophys Acta.* 1969;185(2):269-286.

76. Nicholson LK, Yamazaki T, Torchia DA, Grzesiek S, Bax A, Stahl SJ, Kaufman JD, Wingfield PT, Lam PY, Jadhav PK, et al. Flexibility and function in HIV-1 protease. *Nat Struct Biol.* 1995 Apr;2(4):274-280

77. Nunn CM, Jeeves M, Cliff MJ, Urquhart GT, George RR, Chao LH, Tscuchia Y, Djordjevic S. Crystal structure of tobacco etch virus protease shows the protein C terminus bound within the active site. *J Mol Biol*. 2005;350(1):145-155.

78. Ong ILH, Yang K-L. Recent developments in protease activity assays and sensors. *Analyst*. 2017;142(11):1867-1881.

79. Parks TD, Leuther KK, Howard ED, Johnston SA, Dougherty WG. Release of proteins and peptides from fusion proteins using a recombinant plant virus proteinase. *Anal Biochem.* 1994 Feb 1;216(2):413-417.

80. Patel D, Frelinger J, Goudsmit J, Kim B. In Vitro Assay for Site-Specific Proteases Using Bead-Attached GFP Substrate. *BioTechniques*. 2001;31(5):1194-1203.

81. Pehrson JC, Weatherman A, Markwell J, Sarath G, Schwartzbach SD. Use of GFP as a Reporter for the Facile Analysis of Sequence-Specific Proteases. *BioTechniques*. 1999;27(1):28-32.

82. Peranen J, Kaariainen' L. Nuclear Localization of Semliki Forest Virus-Specific Nonstructural Protein nsP2. *J Virol*. 1990;64:9.

83. Pettit SC, Michael SF, Swanstrom R. The specificity of the HIV-1 protease. *Perspect Drug Discov Des.* 1993;1:69–83.

84. Phan J, Zdanov A, Evdokimov AG, et al. Structural Basis for the Substrate Specificity of Tobacco Etch Virus Protease. *J Biol Chem*. 2002;277(52):50564-50572.

85. Plummer TH Jr, Kimmel MT. An improved spectrophotometric assay for human plasma carboxypeptidase N1. *Anal Biochem*, 1980;108(2):348-353.

86. Prabu-Jeyabalan M, Nalivaika E, Schiffer CA. Substrate Shape Determines Specificity of Recognition for HIV-1 Protease: Analysis of Crystal Structures of Six Substrate Complexes. *Structure*. 2002;10(3):369-381.

87. Pritz S, Doering K, Woelcke J, Hassiepen U. Fluorescence lifetime assays: current advances and applications in drug discovery. *Expert Opin Drug Discov*. 2011;6(6):663-670.

88. Puente XS, Sánchez LM, Overall CM, López-Otín C. Human and mouse proteases: a comparative genomic approach. *Nat Rev Genet*. 2003;4(7):544-558.

89. Puhl AC, Giacomini C, Irazoqui G, Batista-Viera F, Villarino A, Terenzi H. Covalent immobilization of tobacco-etch-virus NIa protease: a useful tool for cleavage of the histidine tag of recombinant proteins. *Biotechnol Appl Biochem*. 2009;53(3):165-174.

90. Rawlings ND, Barrett AJ, Bateman A. MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res*. 2012;40(D1):D343-D350.

91. Rawlings ND, Barrett AJ, Finn R. Twenty years of the *MEROPS* database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res*. 2016;44(D1):D343-D350.

92. Rawlings ND, Bateman A. Pepsin homologues in bacteria. BMC Genomics. 2009;10(1):437.

93. Reichert E, Clase A, Bacetty A, Larsen J. Alphavirus antiviral drug development: scientific gap analysis and prospective research areas. *Biosecur Bioterror*. 2009;7(4):413-427.

94. Revers F, García JA. Molecular biology of potyviruses. Adv Virus Res. 2015;92:101-199.

95. Richardson P. The Determination and Use of Optimized Protease Substrates In Drug Discovery and Development. *CPD*. 2002;8(28):2559-2581.

96. Rikkonen M, Peranen J, Kaariainen L. ATPase and GTPase Activities Associated with Semliki Forest Virus Nonstructural Protein nsP2. *J Virol*. 1994;68:7.

97. Rupp JC, Sokoloski KJ, Gebhart NN, Hardy RW. Alphavirus RNA synthesis and non-structural protein functions. *J Gen Virol*. 2015;96(9):2483-2500.

98. Russo AT, Malmstrom RD, White MA, Watowich SJ. Structural basis for substrate specificity of alphavirus nsP2 proteases. *J Mol Graph Model*. 2010;29(1):46-53.

99. Russo AT, White MA, Watowich SJ. The Crystal Structure of the Venezuelan Equine Encephalitis Alphavirus nsP2 Protease. *Structure*. 2006;14(9):1449-1458.

100. Salvesen G, Rawlings ND. Handbook of Proteolytic Enzymes. London; Boston : Academic Press, 2013.

101. Sawicki DL, Perri S, Polo JM, Sawicki SG. Role for nsP2 Proteins in the Cessation of Alphavirus Minus-Strand Synthesis by Host Cells. *J Virol*. 2006;80(1):360-371.

102. Sawicki DL, Sawicki SG. A second nonstructural protein functions in the regulation of alphavirus negative-strand RNA synthesis. *J Virol*. 1993; 67(6): 3605–3610.

103. Schilling O, Overall CM. Proteome-derived, database-searchable peptide libraries for identifying protease cleavage sites. *Nat Biotechnol*. 2008;26(6):685-694.

104. Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. *Nat Methods*. 2012;9(7):671-675.

105. Seemüller E, Lupas A, Stock D, Löwe J, Huber R, Baumeister W. Proteasome from Thermoplasma acidophilum: a threonine protease. *Science*. 1995;268(5210):579-582.

106. Shaner NC, Steinbach PA, Tsien RY. A guide to choosing fluorescent proteins. *Nat Methods*. 2005;2(12):905-909.

107. Shen L, Tatham MH, Dong C, Zagórska A, Naismith JH, Hay RT. SUMO protease SENP1 induces isomerization of the scissile peptide bond. *Nat Struct Mol Biol*. 2006;13(12):1069-1077.

108. Shu X, Shaner NC, Yarbrough CA, Tsien RY, Remington SJ. Novel Chromophores and Buried Charges Control Color in mFruits. *Biochemistry*. 2006;45(32):9639-9647.

109. Smith DB, Johnson KS. Single-step purification of polypeptides expressed in Escherichia coli as fusions with glutathione S-transferase. *Gene*. 1988;67(1):31-40.

110. Sticha KR, Sieg CA, Bergstrom CP, Hanna PE, Wagner CR. Overexpression and large-scale purification of recombinant hamster polymorphic arylamine N-acetyltransferase as a dihydrofolate reductase fusion protein. *Protein Expr Purif*. 1997;10(1):141-153.

111. Strauss EG, De Groot RJ, Levinson R, Strauss JH. Identification of the active site residues in the nsP2 proteinase of Sindbis virus. Virology. 1992;191(2):932-940

112. Strauss JH, Strauss EG. The Alphaviruses: Gene Expression, Replication, and Evolution. *Microbiol Rev*. 1994;58(3): 491–562.

113. Sun P, Austin BP, Tözsér J, Waugh DS. Structural determinants of tobacco vein mottling virus protease substrate specificity: Structure of TVMV Protease/Substrate Complex. *Protein Science*. 2010;19(11):2240-2251.

114. Suopanki J, Sawicki DL, Sawicki SG, Kääriäinen L. Regulation of alphavirus 26S mRNA transcription by replicase component nsP2. *J Gen Virol*. 1998;79(2):309-319.

115. Tajima N, Kawai F, Park SY, Tame JR. A novel intein-like autoproteolytic mechanism in autotransPorter proteins. *J Mol Biol*. 2010;402(4):645-656.

116. Thorne N, Auld DS, Inglese J. Apparent activity in high-throughput screening: origins of compound-dependent assay interference. *Curr Opin Chem Biol*. 2010;14(3):315-324.

117. Tözsér J. Specificity of Retroviral Proteinases Based on Substrates Containing Tyrosine and Proline at the Site of Cleavage. *Pathol Oncol Res*. 1997;3(2):142-146.

118. Tőzsér J, Bláha I, Copeland TD, Wondrak EM, Oroszlan S. Comparison of the HIV-1 and HIV-2 proteinases using oligopeptide substrates representing cleavage sites in Gag and Gag-Pol polyproteins. *FEBS Letters*. 1991;281(1-2):77-80.

119. Tőzsér J, Tropea JE, Cherry S, et al. Comparison of the substrate specificity of two potyvirus proteases: Comparison of two potyvirus proteases. *FEBS Journal*. 2005;272(2):514-523.

120. Tropea JE, Cherry S, Nallamsetty S, Bignon C, Waugh DS. A generic method for the production of recombinant proteins in Escherichia coli using a dual hexahistidine-maltose-binding protein affinity tag. *Methods Mol Biol*. 2007;363:1-19.

121. Turk B. Targeting proteases: successes, failures and future prospects. *Nat Rev Drug Discov*. 2006;5(9):785-799.

122. Umezawa H, Aoyagi T, Morishima H, Matsuzaki M, Hamada M, Takeuchi T. Pepstatin, A New Pepsin Inhibitor Produced By Agtinomygetes. *J Antibiot*. 1970;23(5):259-262.

123. UNAIDS Data 2017. Joint United Nations Programme on HIV/AIDS

124. Vasiljeva L, Merits A, Auvinen P, Kääriäinen L. Identification Of A Novel Function Of The Alphavirus Capping Apparatus: Rna 5′-Triphosphatase Activity Of NsP2. *J Biol Chem*. 2000;275(23):17281-17287.

125. Vasiljeva L, Merits A, Golubtsov A, Sizemskaja V, Kääriäinen L, Ahola T. Regulation of the Sequential Processing of Semliki Forest Virus Replicase Polyprotein. *J Biol Chem*. 2003;278(43):41636-41645.

126. Vasiljeva L, Valmu L, Kääriäinen L, Merits A. Site-specific Protease Activity of the Carboxyl-terminal Domain of Semliki Forest Virus Replicase Protein nsP2. *J Biol Chem*. 2001;276(33):30786-30793.

127. Verma S, Dixit R, Pandey KC. Cysteine Proteases: Modes of Activation and Future Prospects as Pharmacological Targets. *Front Pharmacol*. 2016;7.

128. Vothknecht UC, Kannangara CG, von Wettstein D. Expression of catalytically active barley glutamyl tRNAGlu reductase in Escherichia coli as a fusion protein with glutathione S-transferase. *Proc Natl Acad Sci USA*. 1996;93(17):9287-9291.

129. Waarts BL. Cell entry mechanisms of alphaviruses: receptor interaction and membrane fusion in a liposomal model system. Groningen: s.n. University of Groningen.

130. Wang Y, Zhu G-F, Ren S-Y, Han Y-G, Luo Y, Du L-F. Insight into the structural stability of wild-type and mutants of the tobacco etch virus protease with molecular dynamics simulations. *J Mol Model*. 2013;19(11):4865-4875.

131. Waugh DS. An overview of enzymatic reagents for the removal of affinity tags. *Protein Expr Purif*. 2011;80(2):283-293.

132. Wlodawer A, Erickson JW. Structure-based inhibitors of HIV-1 protease. *Annu Rev Biochem*. 1993;62:543-585.

133. Woelcke J, Hassiepen U. Fluorescence- Based Biochemical Protease Assay Formats. In A Practical Guide to Assay Development and High-Throughput Screening in Drug Discovery. Ed. Taosheng Chen. CRC Press; 2009.

134. Wu C-H, Liu I-J, Lu R-M, Wu H-C. Advancement and applications of peptide phage display technology in biomedical science. *J Biomed Sci*. 2016;23(1):8.

135. Xia Z, Xing Y, So M-K, Koh AL, Sinclair R, Rao J. Multiplex Detection of Protease Activity with Quantum Dot Nanosensors Prepared by Intein-Mediated Specific Bioconjugation. *Anal Chem*. 2008;80(22):8649-8655.

136. Yi L, Gebhard MC, Li Q, Taft JM, Georgiou G, Iverson BL. Engineering of TEV protease variants by yeast ER sequestration screening (YESS) of combinatorial libraries. *Proc Natl Acad Sci USA s*. 2013;110(18):7229-7234.

137. Yilmaz NK, Schiffer CA. Drug Resistance to HIV-1 Protease Inhibitors: Molecular Mechanisms and Substrate Coevolution. In: Mayers D., Sobel J., Ouellette M., Kaye K., Marchaim D. (eds) Antimicrobial Drug Resistance. Springer, Cham. 2017.

138. Young CL, Britton ZT, Robinson AS. Recombinant protein expression and purification: A comprehensive review of affinity tags and microbial applications. *Biotechnol J*. 2012;7(5):620-634.

139. Zacharias DA, Violin JD, Newton AC, Tsien RY. Partitioning of lipid-modified monomeric GFPs into membrane microdomains of live cells. *Science*. 2002;296(5569):913-916.

140. Zacks MA, Paessler S. Encephalitic alphaviruses. *Vet Microbiol*. 2010;140(3-4):281-286.

141. Zhang B. Design of FRET-based GFP probes for detection of protease inhibitors. *Biochem Biophys Res Commun*. 2004;323(2):674-678.

142. Zhang D, Tözsér J, Waugh DS. Molecular cloning, overproduction, purification and biochemical characterization of the p39 nsP2 protease domains encoded by three alphaviruses. *Protein Expr Purif*. 2009;64(1):89-97.

143. Zhang G. Protease assays, in: G.S. Sittampalam, N.P. Coussens, K. Brimacombe, et al. (Eds.), Assay Guidance Manual [Internet], Eli Lilly & Company and the National Center for Advancing Translational Sciences, Bethesda (MD), 2012.106.

144. Zhou C, Yan Y, Fang J, Cheng B, Fan J. A new fusion protein platform for quantitatively measuring activity of multiple proteases. *Microb Cell Fact*. 2014;13(1):44.

145. Zhu K, Zhou X, Yan Y, Mo H, Xie Y, Cheng B, Fan J. Cleavage of fusion proteins on the affinity resins using the TEV protease variant. *Protein Expr Purif*. 2017;131:27-33.

10. List of publications prepared by the Kenézy Life Science Library

UNIVERSITY of DEBRECEN

Registry number: DEENK/375/2019.PL
Subject: PhD Publikációs Lista

Candidate: Beáta Bozóki
Neptun ID: SDC8DN
Doctoral School: Doctoral School of Molecular Cellular and Immune Biology

List of publications related to the dissertation

1. **Bozóki, B.**, Gazda, L., 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

2. **Bozóki, B.**, Mótyán, J. A., Miczi, M., Gazda, L., Tőzsér, J.: Use of Recombinant Fusion Proteins in a Fluorescent Protease Assay Platform and Their In-gel Renaturation.
   JoVE. 143, 1-15, 2018.
   DOI: http://dx.doi.org/10.3791/58824
   IF: 1.108

List of other publications

3. Mótyán, J. A., Miczi, M., **Bozóki, B.**, Tőzsér, J.: Data supporting Ni-NTA magnetic bead-based fluorescent protease assay using recombinant fusion protein substrates.
   Data in Brief. 18, 203-208, 2018.
   DOI: http://dx.doi.org/10.1016/j.dib.2018.03.031

4. Bagossi, P., Bander, P., **Bozóki, B.**, Tőzsér, J.: Discovery and significance of new Human T-lymphotropic viruses: HTVL-3 and HTLV-4.
   Expert Rev. Anti. Infect. Ther. 7 (10), 1235-1249, 2009.
   IF: 2.857

Total IF of journals (all publications): 6,472
Total IF of journals (publications related to the dissertation): 3,615

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.

20 November, 2019

11. Keywords

Recombinant fusion protein substrate, protease assay, Ni-NTA magnetic agarose beads, fluorescent protein, in-gel renaturation, human immunodeficiency virus type 1 protease, tobacco etch virus protease, venezuelan equine encephalitis virus nsP2 protease

12. Acknowledgements

My completion of this project could not have been accomplished without the support of my family. Special thanks to Kornél, Gergő and Patrik Csontos for allowing me time away from my family to do research and write. I am greatly beholden to my mother, my sister, my father and to my mother- and farther-in-law for their encouragement and the countless times they kept the children during my hectic schedules.

And the last but not at least, I exclusively dedicate this thesis to my grandparents Sándor Papp and Julianna Sápi for their never-ending faith in me, for their outstanding and exemplary attitude to life and for their strength that has been inspiring me when I face any difficulties.

13. Other Posters

2009: Bozóki B, Bagossi P, Boross P and Tőzsér J. Comparative Studies of the substrate specificity of alphavirus proteases. XIVth Annual Conference of Hungarian Biochemistry Society in Budapest.

2010: Bozóki B., Bander P., Matúz K., Kádas J., Bagossi P., Boross P. Tőzsér J. Studies on the substrate-specificity of alphavirus proteases. Viral Vectors in Gene Therapy: Applications & Novel Production methods, Advanced Symposium and EMBO Practical Course, University of Eastern Finland – A-I- Virtanen Institute, Kuopio, Finland.

2012: Bordán Zs., Bozóki B. and Tőzsér J. A method for the modification of alphaviral oligopeptide substrates expressed in a fusion form. European Medical Student's Conference" (EMESCO). Debrecen, Hungary.

2013: Bozóki B., Bordán Zs. and Tőzsér J. Recombinant mutagenized fusion protein substrates of alphavirus proteases. Hungarian Molecular Life Sciences Conference 2013. Siófok Hungary.

APPENDIX