

**Characterization of the Hemagglutinin Cleaving  
Transmembrane Serine Proteases Matriptase and  
TMPRSS2**

**Dissertation**

zur Erlangung des Doktorgrades  
der Naturwissenschaften  
(Dr. rer. nat.)

dem  
Fachbereich Pharmazie der  
Philipps-Universität Marburg  
vorgelegt

von  
**Aline Keils**  
M.Sc.

aus  
Gerolstein

**Marburg/Lahn 2019**



Die Untersuchungen zur vorliegenden Arbeit wurden auf Anregung von Herrn Prof. Dr. Torsten Steinmetzer am Institut für Pharmazeutische Chemie des Fachbereichs Pharmazie der Philipps-Universität Marburg in der Zeit von April 2014 bis September 2018 durchgeführt.

Die Arbeiten zu Kapitel 5 erfolgten zusätzlich auf Anregung von Frau Prof. Dr. Eva Friebertshäuser am Institut für Virologie des Fachbereichs Medizin der Philipps-Universität Marburg.

Eingereicht am 09.12.2019

Erstgutachter: Prof. Dr. Torsten Steinmetzer

Zweitgutachter: Prof. Dr. Eva Friebertshäuser

Tag der mündlichen Prüfung am 31.01.2020

Hochschulkennziffer: 1180



**Tout est bien qui finit bien.**



# Contents

<b>ABBREVIATIONS .....</b>	<b>A</b>
<b>1. RESEARCH RELEVANCE AND THESIS OVERVIEW .....</b>	<b>1</b>
<b>2. GENERAL INTRODUCTION.....</b>	<b>3</b>
2.1. INFLUENZA-A-VIRUS (IAV) .....	3
2.2. IAV SURFACE GLYCOPROTEIN HEMAGGLUTININ .....	8
2.3. TRYPSIN-LIKE SERINE PROTEASES .....	10
<b>3. PREPARATION AND CHARACTERIZATION OF RECOMBINANT MATRIPTASE SPDS.....</b>	<b>17</b>
3.1. INTRODUCTION TO MATRIPTASE PREPARATION AND CHARACTERIZATION.....	17
3.2. RESULTS AND DISCUSSION OF MATRIPTASE PREPARATION AND CHARACTERIZATION .....	25
3.3. SUMMARY OF MATRIPTASE PREPARATION AND CHARACTERIZATION .....	41
<b>4. COMPUTER ASSISTED MATRIPTASE INHIBITOR DESIGN AND THEIR CHARACTERIZATION.....</b>	<b>43</b>
4.1. INTRODUCTION TO MATRIPTASE INHIBITOR DESIGN .....	43
4.2. RESULTS AND DISCUSSION OF MATRIPTASE INHIBITOR DESIGN AND CHARACTERIZATION.....	59
4.3. SUMMARY OF MATRIPTASE INHIBITOR DESIGN AND CHARACTERIZATION .....	77
<b>5. CHARACTERIZATION OF TRUNCATED AND MUTATED TMPRSS2 VARIANTS .....</b>	<b>81</b>
5.1. INTRODUCTION TO TMPRSS2 CHARACTERIZATION.....	81
5.2. RESULTS AND DISCUSSION OF TMPRSS2 CHARACTERIZATION.....	89
5.3. SUMMARY OF TMPRSS2 CHARACTERIZATION .....	109
<b>6. CONCLUSION AND OUTLOOK .....</b>	<b>115</b>
<b>7. EXPERIMENTAL PART.....</b>	<b>117</b>
7.1. MATERIALS AND CONSUMABLES.....	117
7.2. MOLECULAR BIOLOGICAL METHODS .....	123
7.3. MICROBIOLOGICAL METHODS .....	128
7.4. CELL CULTURE METHODS.....	129
7.5. BIOCHEMICAL METHODS .....	130
7.6. BIOSTRUCTURAL METHODS.....	142
7.7. SYNTHESIS OF COMPOUNDS.....	146
7.8. COMPUTATIONAL METHODS.....	146
<b>SUMMARY IN GERMAN.....</b>	<b>149</b>
<b>BIBLIOGRAPHY .....</b>	<b>151</b>
<b>DANKSAGUNG.....</b>	<b>I</b>
<b>PUBLIKATIONEN.....</b>	<b>III</b>
<b>LEBENS LAUF.....</b>	<b>V</b>
<b>EIDESSTÄTTLICHE ERKLÄRUNG .....</b>	<b>VII</b>





# Abbreviations

°C	Degree Celsius
Å	Angstrom (Å = 10 <sup>-10</sup> m)
aa	Amino acid
AMC	7-amino-4-methylcoumarin
APS	Ammoniumpersulfate
ATCC	American Type Culture Collection
bp	Base pair
BSA	Bovine serum albumin
CDC	Centers for Disease Control and Prevention
Cha	Cyclohexylalanine
DAPI	4',6-diamidino-2-phenylindole
def	Deficient
dH <sub>2</sub> O	Deionized water
DMEM	Dulbecco's Modified Eagle Medium
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide triphosphate
DTT	Dithiothreitol
<i>E. coli</i>	<i>Escherichia coli</i>
EDTA	Ethylenediamine tetraacetic acid
EK	Enterokinase
ER	Endoplasmic reticulum
et al.	<i>Et alii</i>
FCS	Fetal calf serum
FITC	Fluorescein isothiocyanate
for	Forward
fXa	Factor Xa
GSH	Reduced glutathione
GSSG	Oxidized glutathione
HA	Hemagglutinin
HA <sub>0</sub>	Hemagglutinin precursor
HA <sub>1</sub> / HA <sub>2</sub>	Hemagglutinin subunit 1 / 2
HAI-I/ HAI-II	Hepatocyte growth factor activator inhibitor I and II
HEK	Human Embryonic Kidney (cells)
HPAIV	Highly pathogenic avian influenza A virus
<i>h</i> Phe	<i>Homo</i> -phenylalanine
HRP	Horseradish peroxidase
IAV	Influenza A virus
IB	Inclusion body
IF	Immunofluorescence
IgK	Immunoglobulin kappa
IMAC	Immobilized metal ion affinity chromatography
IPTG	Isopropyl β-D-1-thiogalactopyranoside
<i>k</i> <sub>cat</sub>	Catalytic rate constant
kDa	Kilodalton

$K_i$	Dissociation constant of the enzyme inhibitor complex
$K_M$	Michaelis-Menten constant
LB	Lysogeny broth
LDLRA	Low-density lipoprotein receptor A (domain)
LPAIV	Low pathogenic avian influenza A virus
M	Molar
MCS	Multiple cloning site
MDCK	Madin-Darby canine kidney (cells)
Mes	Methylsulfonyl
MES	2-(N-morpholino)ethanesulfonic acid
MI	Marburg inhibitor
MME	Monomethyl ether
MMS	Microseed matrix screening
N.D.	Not determined
NA	Neuraminidase
NMR	Nuclear magnetic resonance
nt	Nucleotide
NTA	Nitrilotriacetic acid
OD	Optical density
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PDB	Protein data bank
PEG	Polyethylene glycol
pNA	4-nitroaniline
PPMO	Peptide-conjugated phosphorodiamidate morpholino oligomer
PTM	Post translational modification
PVDF	Polyvinylidene fluoride
rev	Reverse
RFU	Relative fluorescence unit
RKI	Robert Koch Institute
RNA	Ribonucleic acid
rpm	Rounds per minute
RPMI	Roswell Park Memorial Institute (culture medium)
RT	Room temperature
SAR	Structure-activity relationship
SDS	Sodium dodecyl sulfate
SP	Signal peptide
SPD	Serine protease domain
SRCR	Group A scavenger receptor cysteine-rich (domain)
ss	Single stranded
TAE	Tris-acetate-EDTA (buffer)
TBS	Tris buffered saline
TEMED	Tetramethylethylenediamine
TGN	Trans-Golgi Network
TMPRSS	Transmembrane protease, serine, S1 family member
Tos	Toluenesulfonyl (Tosyl)
tPA	Tissue-type plasminogen activator
Tris	Tris(hydroxymethyl)aminomethane
TRITC	Tetramethylrhodamine
TTSP	Type II transmembrane serine protease

uPA	Urokinase-type plasminogen activator
UV/Vis	Ultraviolet–visible
v/v	Volume per volume
V <sub>max</sub>	Maximum velocity of non-inhibited enzyme reaction
w/v	Weight per volume
WHO	World Health Organization
wt	Wild type
$\alpha$	<i>Alpha</i> , used as anti (antibodies)
$\beta$ -ME	2-mercaptoethanol
$\Delta$	<i>Delta</i> , used as deficient or truncated

Abbreviations of chemical compounds are taken from the relevant references like Methods of Organic Chemistry (Houben-Weyl), March's Advanced Organic Chemistry and Greene's Protective Groups in Organic Synthesis.

### Amino acid code table

The abbreviations of the amino acids and their derivatives are used according to the recommendations of the International Union of Pure and Applied Chemistry (IUPAC) and International Union of Biochemistry and Molecular Biology (IUBMB) Joint Commission on Biochemical Nomenclature (JCBN) using the three or one letter code. Unless otherwise indicated, all amino acids and their derivatives have L-configurations.

Amino acid	Three letter code	One letter code	Amino acid	Three letter code	One letter code
Alanine	Ala	A	Leucine	Leu	L
Arginine	Arg	R	Lysine	Lys	K
Asparagine	Asn	N	Methionine	Met	M
Aspartic acid	Asp	D	Phenylalanine	Phe	F
Cysteine	Cys	C	Proline	Pro	P
Glutamic acid	Glu	E	Serine	Ser	S
Glutamine	Gln	Q	Threonine	Thr	T
Glycine	Gly	G	Tryptophan	Trp	W
Histidine	His	H	Tyrosine	Tyr	Y
Isoleucine	Ile	I	Valine	Val	V



# 1. Research relevance and thesis overview

*“Influenza is a serious global health threat that impacts all countries: every year, there are an estimated 1 billion cases, 3-5 million severe cases, and 290,000-650,000 influenza-related respiratory deaths worldwide. In this interconnected world, the next influenza pandemic is a matter of when, not if and a severe pandemic is believed by many experts to be potentially the most devastating global health event with far reaching consequences.”*

- Global Influenza Strategy 2019-2030, WHO <sup>[1]</sup> –

Influenza (short flu) is one of the commonest infections of humans. As a highly contagious acute respiratory disease caused by the influenza virus it affects millions of people every year. The characteristic symptoms include fever, headache, muscle pain, malaise, rhinitis and a non-productive cough but can also cause serious complications such as acute bronchitis, secondary bacterial pneumonia, otitis media or even death. The highest morbidity rates are found in infants and the elderly with over 65 years of age. Furthermore, pregnant women, individuals with chronic disorders and persons with immunosuppression are at higher risk <sup>[2, 3]</sup>. On a phylogenetic-based taxonomy, influenza viruses are assigned to the *Orthomyxoviridae* family. Three genera of influenza viruses are known to infect humans, including influenza type A, B and C with type A and B being responsible for most of the infections. Between these two, influenza type A is the most prevalent one causing both, epidemic and pandemic outbreaks. The latter results from major genetic changes of the influenza virus and occur every 10-50 years with severe consequences for the immunological naive population. Historically, the most devastating flu pandemic occurred in 1918 during and shortly after the first world war and is referred to as the "Spanish flu". Between 1918 and 1920 the H1N1 influenza A subtype affected approximately one third of the world's population and caused approximately 40 million deaths worldwide <sup>[4]</sup>. A novel H1N1 subtype (swine-origin H1N1pdm09) was responsible for the most recent flu pandemic in 2009. Heavy casualties were estimated between 123,000 and 203,000 in the last 9 months of 2009 <sup>[5]</sup>. Whereas these pandemic outbreaks are from a rather unpredictable nature, the seasonal epidemic influenza infections mainly take place annually between December and February and are strictly monitored by global or regional organizations, such as the WHO (worldwide), the CDC (US) or the RKI in case of Germany. The most effective prevention against seasonal flu viruses is vaccination. By the virus's capability to undergo constant continuous genetic changes, termed as antigenic drift or shift, the vaccine composition has to be annually updated based on predictions. Millions of people are affected by influenza

viruses every year. Nine million influenza-attributable medically attended acute respiratory illnesses that resulted in approximately 45,000 influenza-related hospital admissions were estimated for the influenza season 2017/2018 by the RKI in Germany <sup>[6]</sup>. Except for these direct costs of medical care, absenteeism from work results in a huge economic burden of seasonal influenza. 5.3 million physician-signed sick leaves, including the need to stay home cases in case of non-employed persons such as children and the elderly were estimated for the 2017/2018 influenza season in Germany. The actual influenza prevention strategy as well as the required medical care and the financial suffers through absenteeism making influenza a cost-intensive disease of the modern world. Not to forget the heavy casualties of around 290,000 – 650,000 every year <sup>[7]</sup>. New strategies for influenza treatment in order to protect the human population from the recent influenza epidemics and prepare them for the next imminent pandemic are in demand. In the “Global influenza strategy 2019-2030” the WHO defined the main goals as to:

- Reduce the burden of seasonal influenza,
- Minimize the risk of zoonotic influenza and
- Mitigate the impact of pandemic influenza.

Improved tools for prevention, detection and treatment of influenza are required to address unmet public health needs. To do so, influenza research has to be promoted, including the research to better understand the virus characteristics and host factors that drive the impact of influenza <sup>[1]</sup>.

The following work addresses the characterization of host cell factors, with the main focus on the type-II transmembrane serine proteases TMPRSS2 and matriptase that are involved in (but not limited to) the activation of several influenza A virus surface glycoprotein hemagglutinin subtypes. The cleavage of hemagglutinin by host proteases is essential for virus spread and infectivity. The inhibition of these relevant proteases therefore provides a promising therapeutic approach. In detail, this thesis is divided into a general introduction (chapter 2), three main chapters covering three different projects (chapters 3, 4 and 5), a conclusion and outlook (chapter 6) and the experimental section (chapter 7).

The three independent main chapters namely:

- I. Preparation and characterization of recombinant matriptase SPDs (Chapter 3),
- II. Computer-assisted matriptase inhibitor design and their characterization (Chapter 4),
- III. Characterization of truncated and mutated TMPRSS2 variants (Chapter 5),

are separated each in introduction, results and discussion and a chapter specific summary.

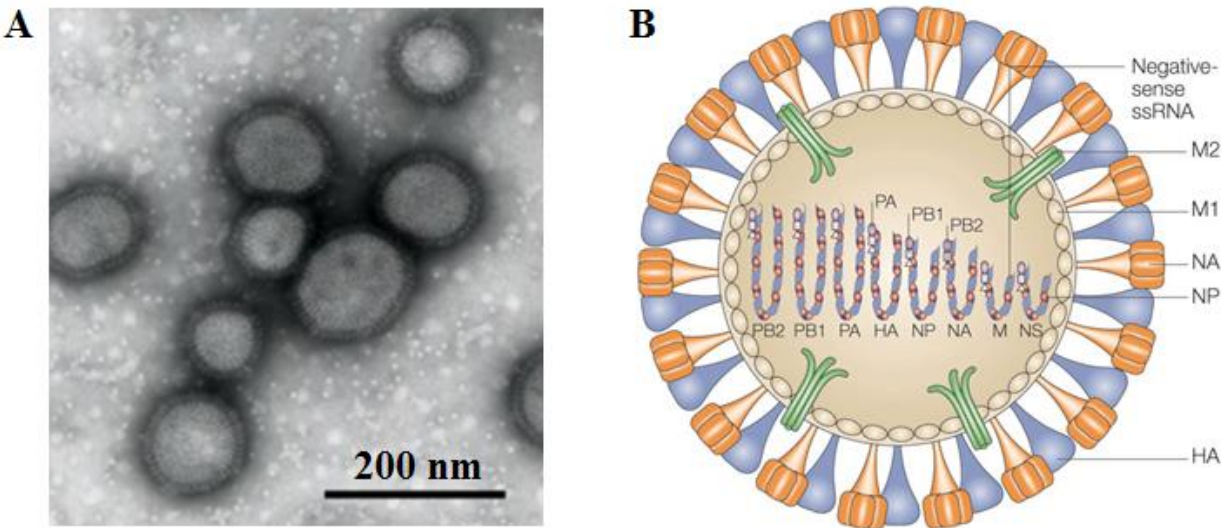
## **2. General introduction**

### **2.1. Influenza-A-virus (IAV)**

IAV is an enveloped virus with a genome of segmented negatively sensed single-stranded (ss) RNA. Based on the antigen variability of the surface glycoproteins hemagglutinin (HA) and neuraminidase (NA) IAVs are divided into several subtypes. So far, sequences of 18 HA and 11 NA subtypes have been identified <sup>[8]</sup>. Nearly all of the 144 theoretical combinations of H1-H16 and N1-N9 are found in aquatic wild birds, which were determined as the natural main reservoir. Subtypes H17 and H18 as well as N10 and N11 have been exclusively identified in bats suggesting that wild birds are not the only natural reservoir <sup>[8-10]</sup>.

#### **2.1.1. Virus and genomic structure**

The enveloped IAV particles possess a spherical (100 nm) (Figure 2-1 A) or more rarely a filamentous form (20 µm) <sup>[11]</sup>. The genome consists of 8 single-stranded negatively sensed RNAs ((-)ssRNA) encoding for at least 12 proteins that are required for virus survival and propagation (Table 2-1). The viral RNA is packed into viral ribonucleoprotein (vRNP) complexes where the virus RNA is associated with the viral heterotrimeric RNA polymerase (consisting of the polymerase basic protein 1 (PB1), polymerase basic protein 2 (PB2) and the polymerase acid protein (PA)) and is coated with multiple nucleoproteins (NPs). These RNP complexes are surrounded by multiple matrix protein 1 (M1) and located beneath the envelope. Besides HA and NA, the virus envelope contains the channel protein M2 (matrix protein 2). In addition, small quantities of non-structural proteins (NS) are found in the virion. A schematic illustration of a single virion is given in Figure 2-1 B.



**Figure 2-1: Spherical Influenza A virus particles.** **A:** Electron micrograph of IAV subtype H1N1. HA and NA are visible in form of "spikes" on the IAV surface. Figure was taken and modified from <sup>[11]</sup>. **B:** Schematic illustration of one influenza A virion. The major glycoproteins HA and NA, together with M2, are embedded into the virus membrane. The M1 protein coats the inner layer of the virus particle and interacts with the NP protein. This protein encapsulates the 8 (-)ssRNA genome segments and forms together with the polymerase complex (PB1, PB2 and PA) the viral ribonucleoproteins (vRNPs). Figure was taken from <sup>[12]</sup>.

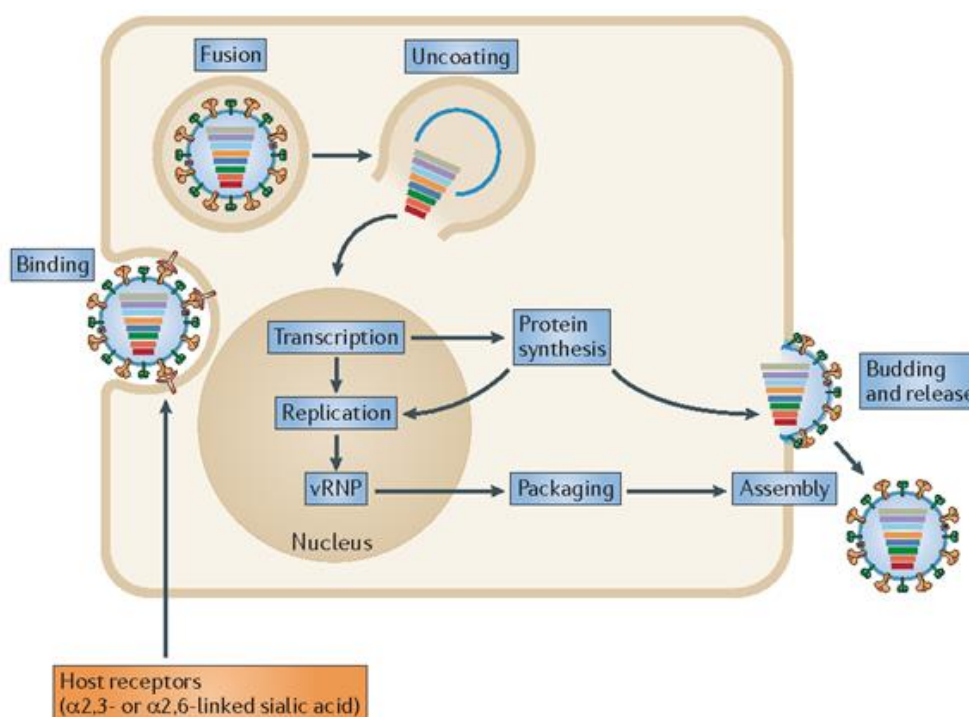
**Table 2-1: Influenza A virus genome segments, encoded proteins and their function** <sup>[13]</sup>

Segment	Protein	Description and Function
1	PB2	Polymerase subunit, binds to 5` cap structure of cellular mRNA
2	PB1	Catalytic subunit of polymerase complex, elongation of viral RNA
	PB1-F2	Virulence factor, induces apoptosis
	PB1-N40	Mediates balance between PB1 and PB1-F2
3	PA	Endonuclease activity, cap snatching
4	HA	Surface glycoprotein, receptor binding and membrane fusion
5	NP	RNA binding protein, major component vRNP, mediates nuclear import of vRNPs
6	NA	Surface glycoprotein, sialidase activity, virus release
7	M1	Matrix protein, interaction with vRNPs, nuclear export and budding
8	M2	Ion channel, important for uncoating and assembly
	NS1	Interferon antagonist and regulation of host gene expression
	NEP	Nuclear export of synthesized vRNPs (earlier name: NS2)



### 2.1.2. Replication cycle of influenza A viruses

The multistage IAV replication process includes binding, entry, fusion and uncoating, genome transcription, viral protein synthesis, assembly and finally egress/budding of progeny virions (Figure 2-2). Since transcription and replication of IAV take place in the nucleus, the viral RNPs have to be transported from the cell periphery to the nucleus of the cell for RNA synthesis.



**Figure 2-2: Schematic illustration of Influenza A virus replication cycle** including the virion binding to the host cell via  $\alpha 2.3$ - or  $\alpha 2.6$ -linked sialic acid, the entry, fusion and uncoating as well as genome transcription, viral protein synthesis, assembly and finally egress/budding of progeny virions. Figure was taken and modified from literature <sup>[14]</sup>.

The first step in IAV replication is the attachment of virions to the target cells <sup>[15]</sup> that is mediated by the binding of the homotrimeric virus surface glycoprotein HA to  $\alpha 2.3$ - or  $\alpha 2.6$ -linked terminal sialic acids on cellular host glycoproteins or glycolipids <sup>[16]</sup>. The virus is subsequently transported to the cell interior by endocytosis. The acidic pH within the endosomes triggers irreversible conformational changes in the HA causing the insertion of the HA<sub>2</sub> subunit into the endosomal membrane <sup>[17]</sup>. Furthermore, the low pH activates the ion channel activity of viral M2 envelope proteins, resulting in the proton transport into the virion interior. This acidification leads to the release of viral RNP complexes and after the uncoating step to its transport into the nucleus that is mediated by nuclear localization signals on the nucleoprotein <sup>[18-20]</sup>. The negatively sensed genomic ssRNA segments are transcribed into mRNA by the heterotrimeric viral polymerase complex and translated in the cytoplasm by the host's translation machinery. After being modified, the viral proteins NP, PA, PB1, and PB2

are transported back into the nucleus, where viral RNP complexes are formed and finally exported together with the M1 protein out to the cell surface. HA, NA and M2 are transported along the secretory pathway to the plasma membrane, where they wrap around the vRNP. Finally, virus progeny is released from cells by viral NA <sup>[13]</sup>.

### **2.1.3. IAV pathology**

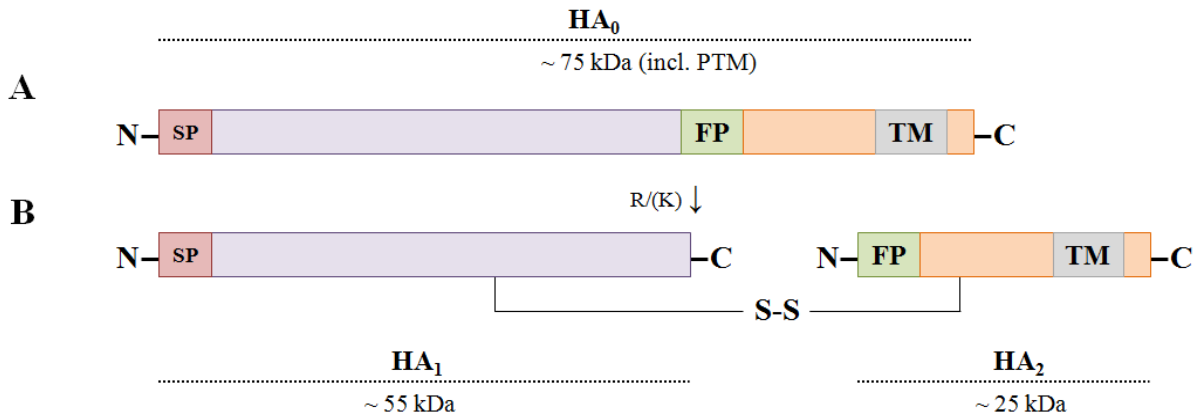
On the basis of molecular genetics and pathogenesis criteria, low pathogenic avian influenza viruses (LPAIV) and highly pathogenic avian influenza viruses (HPAIV) can be discriminated <sup>[21]</sup>. Whereas LPAIV usually induce no symptoms or only mild disease in birds, HPAIV cause severe disease with high mortality in poultry <sup>[22-24]</sup>. Generally, humans can be infected with both, HPAIV and LPAIV. However, infection with HPAIV usually requires direct contact with infected sick or dead birds. Zoonotic, non-human-adapted IAVs such as the HPAIV subtype H5N1 and H7N7 only occasionally infect humans but cause severe and frequently fatal diseases. Mild to severe, as well as fatal illness in humans have been also observed upon infection with the non-human adapted LPAIV virus H7N9 <sup>[25, 26]</sup>. Most of the known avian influenza viruses are classified as LPAIV. They are thought to be transmitted among non-human mammals, such as animals from domestic poultry or swine <sup>[27]</sup>. Currently, only 3 HA subtypes (H1, H2 and H3) and 2 NA subtypes (N1 and N2) with mainly H1N1 and H3N2 are circulating in humans and cause the annual epidemics. As the main target for neutralizing antibodies, NA and particularly HA are under intense immune-mediated selection pressure. Two mechanisms, namely antigenic drift and antigenic shift allow the IAV to emerge host immunity. Antigenic drift describes the continuous acquisition of point mutations in the influenza surface proteins NA and HA that pass unnoticed since the viral RNA polymerase complex lacks a proofreading function. These small genetic changes can accumulate over time and result in antigenically distinct virus strains. The partial escape in host immunity by antigenic drift causes the yearly epidemic IAV infections. Due to the segmented nature of the influenza gene, entire genomic segments can be exchanged between different IAV strains of different hosts with the result of antigenically novel virus strains. These reassortment events are termed antigenic shift <sup>[12]</sup>. Since the population is immunologically naive against new strains, pandemics might be the consequence. Pandemics occur only every 10 to 50 years <sup>[28]</sup> but come with more severe consequences compared to the annual epidemics. The first documented pandemic was the 1918 Spanish flu, caused by the transmission of an avian H1N1 virus to humans. The death of over 40 million people was the consequence <sup>[4]</sup>. Less severe pandemics were the Asian flu (H2N2) in 1957, the Hong Kong flu (H3N2) in 1968 <sup>[29]</sup> and the recent Swine flu (pH1N1) in 2009 <sup>[30]</sup>.

#### 2.1.4. Prophylaxis and therapeutics

The most efficient strategy to fight infectious diseases is vaccination. Although influenza vaccines are available since the 1930s<sup>[31, 32]</sup> the constant genetic changes entail a new vaccine composition on a yearly basis that is based on predictions. Besides of social distancing and a high personal hygienic standard, the yearly influenza vaccination is still the best option to prevent influenza infections and is highly recommended by the Centers for Disease Control and Prevention (CDC) and the World Health Organization (WHO) for high risk groups such as the elderly and children. Currently, trivalent and quadrivalent vaccines with either inactivated or live attenuated (i.e. less virulent) viruses are available. The recent influenza vaccines consist of two IAV strains (one pH1N1 and one H3N2 strain) and either one (trivalent) or two (quadrivalent) influenza B strains<sup>[33]</sup>. The vaccines are designed to stimulate the immune system to develop an adaptive response against viruses predicted to be circulating during the next influenza season. However, in case of a pandemic event, it will still take several months to develop a new effective vaccine, which rises the urgent need for novel and effective anti-influenza drugs<sup>[13]</sup>. In theory, anti-influenza drug strategies can be directed against every step in the viral propagation cycle (see 2.1.1). At present, three classes of symptom-reducing antiviral medication are approved by the Food and Drug Administration (FDA), including M2 ion channel inhibitors (Amantadine and Rimantadine), NA inhibitors (Oseltamivir, Zanamivir and Peramivir) and the very recently approved pro-drug baloxavir marboxil, an inhibitor of the endonuclease PA. These drugs are to be taken within 48 hours after the debut of influenza symptoms in order to reduce illness duration and severity<sup>[34]</sup>. The M2 ion channel inhibitors that only act on influenza A viruses are currently not recommended by CDC and WHO, since the latest H3N2 and pH1N1 swine flu strains showed amantadine and rimantadine resistance of 92 % to 100 % respectively<sup>[35, 36]</sup>. Also, in the case of NA, specific mutations confer drug resistance against the present NA inhibitors<sup>[35, 37-39]</sup>. New strategies against influenza are required to face the emerging resistance. Researchers focus on the development of new derivatives of existing drugs and on new targets with HA being one of the most appealing targets. HA targeting approaches focus on antibody-based therapy where neutralizing monoclonal antibodies bind to reserved regions of HA<sup>[40]</sup> or on small molecules HA inhibitors. Depending on the HA maturation stage, those inhibitors target HA glycosylation, proteolytic activation, attachment to host cell receptors and HA-mediated membrane fusion<sup>[13, 41]</sup>.

## 2.2. IAV surface glycoprotein hemagglutinin

Hemagglutinin is a type I transmembrane protein with an N-terminal signal peptide, a long ectodomain, a transmembrane region and a short cytoplasmic tail at the C-terminus (Figure 2-3). HA is synthesized as a single chain zymogen consisting of 560 amino acids. This polypeptide has a molecular weight of 60 kDa that increases upon glycosylation <sup>[17]</sup>.



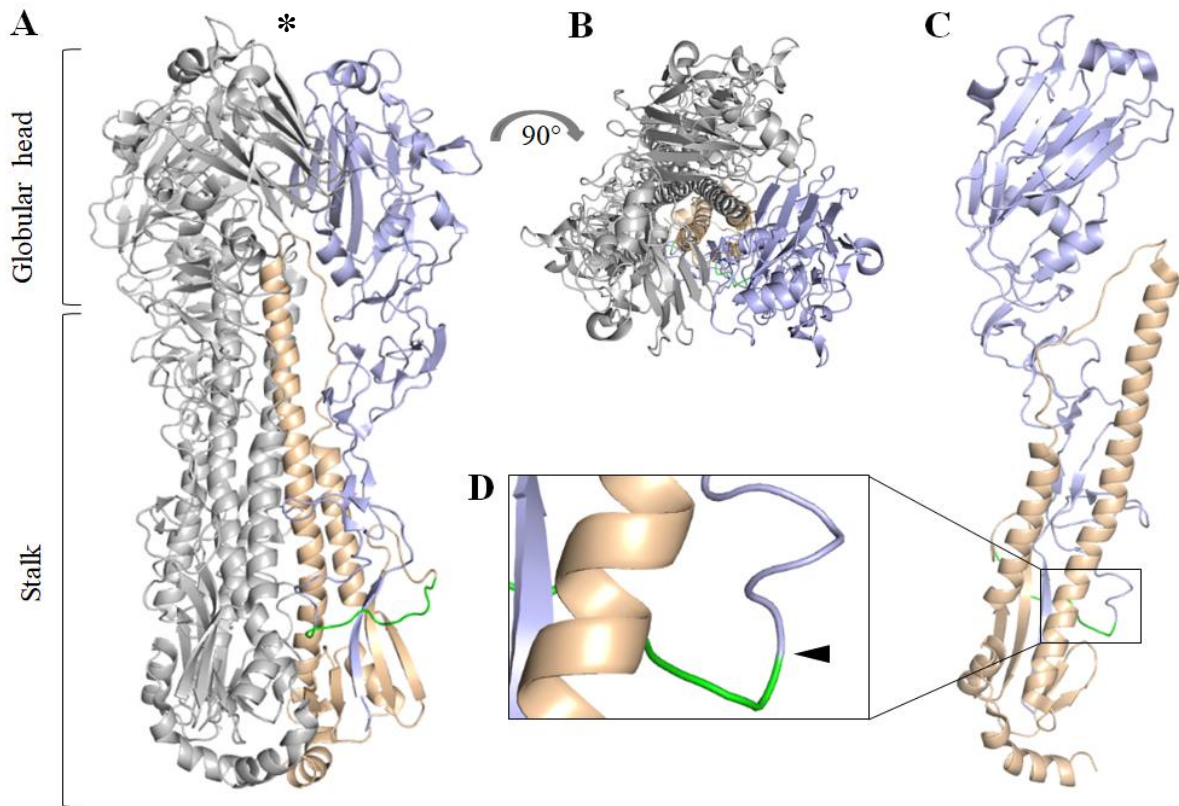
**Figure 2-3: Genetic organization of hemagglutinin.** **A:** HA is synthesized as the precursor HA<sub>0</sub> protein of 75 kDa (after modification) and consists of an N-terminal signal peptide (SP, red), the subunits HA<sub>1</sub> (violet) and HA<sub>2</sub> (orange) including the N-terminal fusion peptide (FP, green) as well as the C-terminal transmembrane domain (grey). **B:** Cleavage (arrow) after an arginine residue (or more rarely lysine) results in the introduction of a new C-terminus to the 55 kDa HA<sub>1</sub> and an N-terminus to the 25 kDa HA<sub>2</sub> subunit. The two subunits remain linked by a conserved disulfide bond (S-S) post cleavage.

The HA ectodomain contains the HA<sub>1</sub> subunit, a single arginine (or more rarely a lysine) connector residue, and the HA<sub>2</sub> subunit <sup>[42]</sup>. During modifications in the ER, high mannose N-glycans are added and HA homotrimers are formed. The position and number of N-glycosylation sites, hence the size of the HA<sub>0</sub> precursor protein, can vary with different IAV isolates <sup>[43]</sup>. HA<sub>0</sub> is cleaved by cellular proteases into the subunits HA<sub>1</sub> and HA<sub>2</sub>, either in the trans-Golgi network (TGN), on the cell surface or in endosomes. Both subunits remain tightly associated post-cleavage through a conserved disulfide bond <sup>[44]</sup>.

### 2.2.1. Three-dimensional structure of HA

HA homotrimers have a cylindrical shape (Figure 2-4 A and B) with approximate dimensions of 135 Å (length) x 35-70 Å (radius) and can be detected via electron microscopy in form of spikes that project externally <sup>[45, 46]</sup>. Each HA monomer has a globular head domain and a stalk-like stem domain (see Figure 2-4 C). The globular head is formed by the HA<sub>1</sub> residues 116 – 261, which are folded into a jelly-roll motif of eight stranded antiparallel β-sheets and sit on top of the triple-stranded coiled-coil of α-helices of the stem region <sup>[47, 48]</sup>. HA<sub>1</sub> contains the receptor-binding site and the major antigenic epitopes. Amino acid changes at these antigenic sites during infection are associated with antigenic change <sup>[17]</sup>. The stalk is composed of the N-

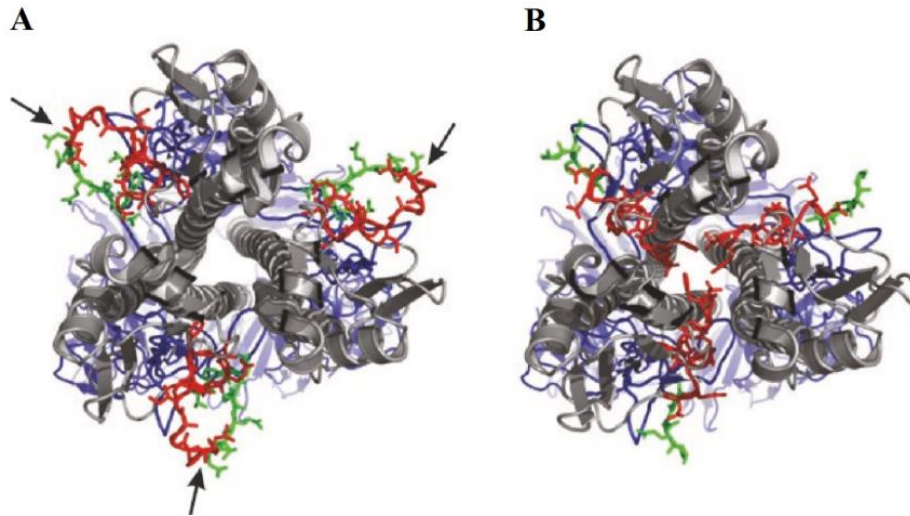
and C-terminus of HA<sub>1</sub>, the proteolytic cleavage site on the surface loop (Figure 2-4 D), and the HA<sub>2</sub> subunit containing the fusion protein that mediates the fusion of viral and endosomal host membranes during virus uptake <sup>[49]</sup>. The fusion peptide is a highly conserved region among all HA subtypes at the N-terminus of HA<sub>2</sub> and is only presented after proteolytic cleavage of the precursor HA<sub>0</sub> <sup>[50]</sup>.



**Figure 2-4: The structure of uncleaved influenza A virus hemagglutinin subtype H1.** Homotrimeric structure from the side (A) or the top view (B) orientation as well as the HA monomer (C) are color-coded as followed: HA<sub>1</sub> subunit in violet, HA<sub>2</sub> subunit in wheat with the fusion peptide in green. The globular head region contains the receptor-binding site (RBS) and is indicated by the star symbol. The stalk-like domain mainly consisting of HA<sub>2</sub>, the N- plus C- terminus of HA<sub>1</sub> and harbors the active cleavage site (D). Cleavage occurs after an arginine residue and is indicated with an arrowhead. Figure was made with PyMol on the basis of the protein data bank (PDB) code: 1RD8.

The proteolytic cleavage of HA<sub>0</sub> into the subunits HA<sub>1</sub> and HA<sub>2</sub> is an essential step in the hemagglutinin-mediated virus uptake and thus virus spread and infectivity <sup>[51, 52]</sup>. Major consequences of the activation cleavage are the generation of an HA<sub>1</sub> C-terminus and an HA<sub>2</sub> N-terminus, the latter containing the fusion peptide <sup>[53]</sup>. Post cleavage, the fusion peptide is hidden within the hemagglutinin homotrimers (see Figure 2-5 B). For HA fusion potential, several irreversible structural rearrangements of the HA, as a result of acidification, are required. At a pH of 5.0 – 6.0 (depending on the virus strain) the fusion peptide residues are relocated to the same end as the C-terminal end of the polypeptide chains close to the membrane

anchor domain <sup>[54]</sup>. This rod-like low pH structure is considered to be the most stable HA conformation.



**Figure 2-5: Homotrimeric structure of HA3 of A/Aichi/2/68 virus (H3N2 subtype).** **A:** Uncleaved H3-HA<sub>0</sub> with the cleavage loops in red and green sticks. Arrows indicating the cleavage site. **B:** Insertion of the HA<sub>2</sub> N-terminus into the trimer interior post cleavage. The HA<sub>1</sub> subunits are shown in blue and HA<sub>2</sub> subunits in gray. Figure was taken from literature <sup>[54]</sup>.

Depending on the cleavage motif of the respective HA strain and the host organism, different host proteases can catalyze the HA<sub>0</sub> activation cleavage. In general, the number of basic amino acid residues, namely arginine and lysine, and the cleavage site sequence determine the pathogenicity of the virus <sup>[42, 49, 55]</sup>. HAs of LPAIVs, including H1 (as in H1N1) and H3 (as in H3N2) have monobasic cleavage sites. In contrast, HPAIVs such as certain H5- and H7-HA possess multibasic cleavage sites. HAs with multibasic cleavage sites of R-X-R/K-R↓G-L-F or R-X-X-R↓G-L-F with X as a non-basic amino acid can be proteolytically activated by ubiquitous present furin and furin-like proteases. Since these proteases are found throughout the host body they contribute to severe systemic infections <sup>[49]</sup>. Host cell proteases that activate LPAIV are mainly trypsin-like serine proteases including soluble serine proteases such as tryptase Clara <sup>[56]</sup> or plasmin <sup>[57]</sup> and membrane-anchored serine proteases, including the transmembrane serine S1 member 2 (TMPRSS2) <sup>[58-62]</sup> and matriptase <sup>[63, 64]</sup>. The epidemic annual infections from LPAIV are restricted to the aerodigestive tract. Viral spread is therefore limited to these organs, which makes it easier to be controlled by the immune system.

### 2.3. Trypsin-like serine proteases

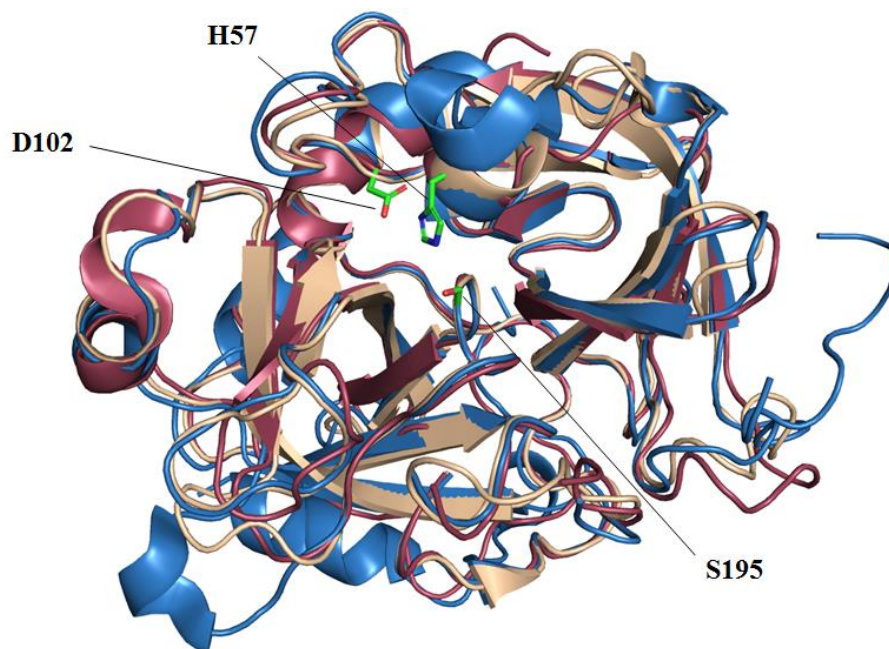
Proteases catalyze the hydrolytic cleavage of peptide bonds. Almost one third of these proteases have a nucleophilic serine residue (along with histidine and aspartate) at the active site and are therefore classified as serine proteases. Trypsin-like serine proteases represent the largest



family of serine proteases and play essential roles in many important biological processes including blood coagulation (thrombin and factor Xa), food digestion (trypsin), fibrinolysis (uPA, tPA, and plasmin) or immunity (tryptase). Due to their high structural similarities, the amino acid residue numbering of the respective SPD follows the chymotrypsin numbering, since the catalytic mechanism was elucidated by the X-ray structure of chymotrypsin in 1985 [65].

### 2.3.1. Characteristics of trypsin-like serine proteases

As mentioned above, members of the trypsin-like serine protease family share high structural similarities. They have two six-stranded asymmetrically arranged  $\beta$ -barrels that host the conserved catalytic triad of His57, Asp102 and Ser195 at their interface (Figure 2-6). Peptide bonds are preferentially cleaved from these proteases following the amino acid residues arginine or lysine, mainly caused by the presence of the negatively charged Asp189 at the bottom of the S1 pocket [66].



**Figure 2-6: Superimposition of three serine proteases of the S1A fold.** Chymotrypsin (red), thrombin (blue) and matriptase (wheat) share a high structural similarity. The catalytic triad of chymotrypsin (consisting of the residues H57, D102 and S195) is highlighted in sticks with green carbon atoms. Figure was made with PyMol on the basis PDB codes: 1CBW (chymotrypsin) 2GV6 (matriptase) and 3EQ0 (thrombin).

Trypsin-like serine proteases (and also chymotrypsin) are principally expressed as inactive zymogens that require catalytic activation by other proteases or by an autocatalytic manner. After cleavage of the peptide bond between residue 15 and 16 (Arg15 and Ile16 for chymotrypsinogen), the active enzyme form is stabilized by polar interactions between the new amino-terminal group of residues 16 and the conserved residue Asp194 (in the interior of the

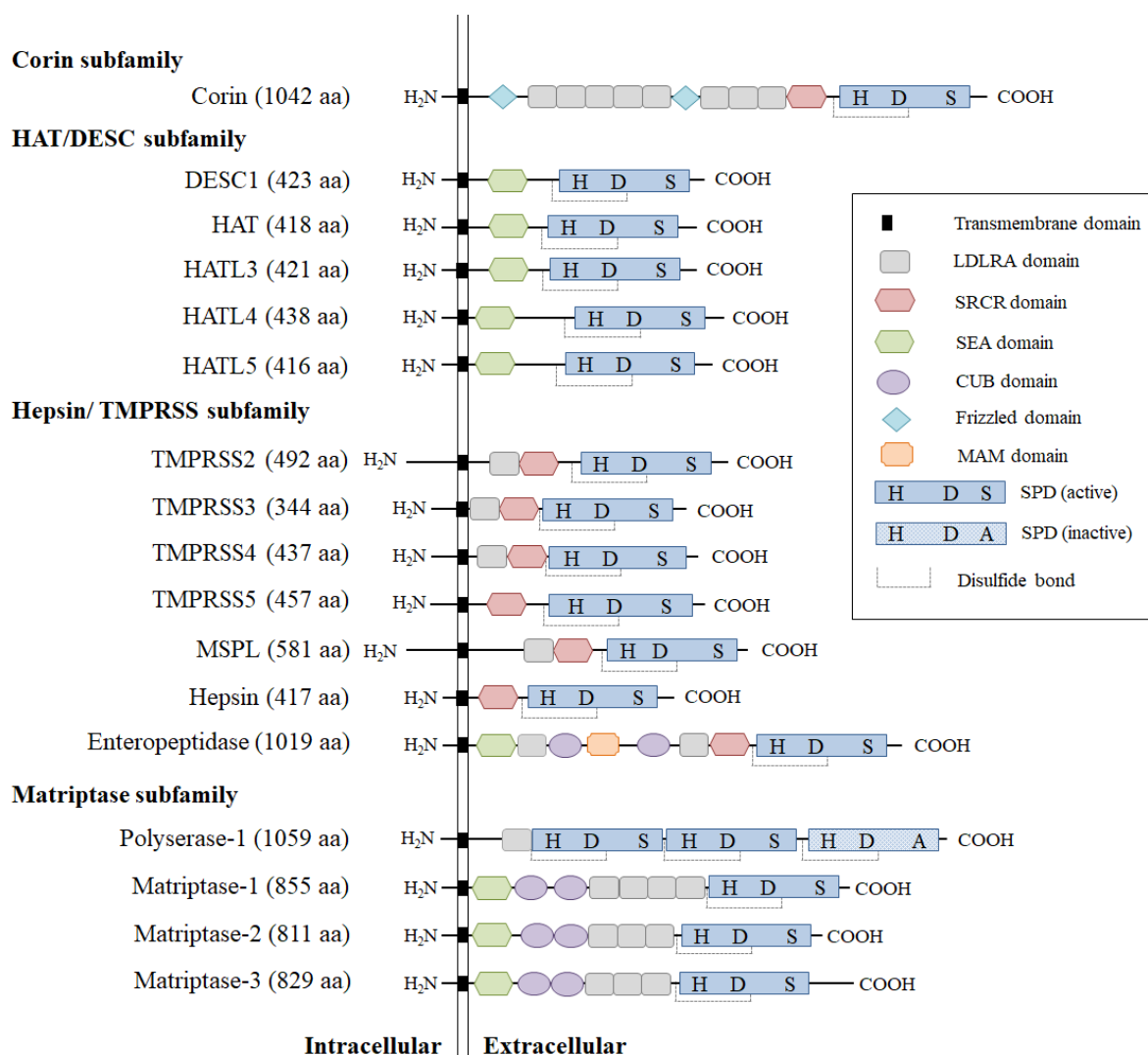
molecule). Since trypsin-like serine proteases control many physiological processes, the expression as inactive zymogen prevents uncontrolled proteolytic degradation and allows tight control of enzyme activity. In addition, zymogens can be stored in larger quantities and the cell can rapidly provide high amounts of active protease when required. The presence of endogenous inhibitor proteins, such as protease inhibiting serpins, including antithrombin and antitrypsin, and Kunitz-type serine protease inhibitors tightly control the enzyme activity. Due to the inactive zymogen form and inhibitor-bound enzymes, the protease activity does therefore not necessarily correlate with its expression level.

Many well-characterized members of the trypsin-like serine proteases, including trypsin itself, are soluble proteases that are secreted into the extracellular space. More recently, a subgroup of trypsin-like serine proteases had been identified that are anchored in the cell membrane by a transmembrane domain either on the C- (type I) or near the N-terminus (type II) or via a glycosylphosphatidylinositol (GPI) linkage.

### **2.3.2. Type-II transmembrane serine proteases**

With 17 members in humans, the type II transmembrane serine proteases (TTSPs) represent the largest group of membrane-anchored serine proteases <sup>[67, 68]</sup>. All TTSPs share the structural organization of an N-terminal cytoplasmic domain, a transmembrane domain and an extracellular domain comprising of a variable stem region of 1 – 11 protein domains of six different types and a serine protease domain (SPD) of the chymotrypsin-like S1A fold at the C-terminus. Based on similarities of the domain structures and amino acid sequences, the TTSPs have been divided into four subfamilies: the hepsin/TMPRSS, the matriptase, the corin and the HAT/DESC subfamily (Figure 2-7).





**Figure 2-7: Schematic overview of the TTSP subfamilies.** All TTSPs consist of an N-terminal cytosolic portion, a membrane anchoring domain and a variable stem region with the SPD at the C-terminus. Based on similarities of the extracellular portion, TTSPs are divided into 4 subfamilies: corin, HAT/DESC, hepsin/TMPRSS and the matriptase subfamily.

### *Hepsin/TMPRSS subfamily*

With seven members, including enteropeptidase, hepsin, TMPRSS2, TMPRSS3, TMPRSS4, TMPRSS5 and MSPL (mosaic serine protease large-form, also TMPRSS13) the hepsin/TMPRSS subfamily is the biggest representative. All members of the subfamily contain a group A scavenger receptor cysteine-rich (SRCR) domain upstream the SPD and almost all, except hepsin and TMPRSS5, have an additional low-density lipoprotein receptor A (LDLRA) domain. Enteropeptidase (also named enterokinase or EK) is the most prominent TTSP and is well-studied because of its essential role in food digestion. It possesses a more diverse stem region of seven domains including a single sea urchin sperm protein, enterokinase, and agrin (SEA) domain, two LDLRA domains, two C1r/s, urchin embryonic growth factor and bone morphogenetic protein 1 (CUB) domains, one meprin, A-5 protein, and receptor protein-

tyrosine phosphatase mu (MAM) domain as well as the family characteristic SRCR domain upstream the protease domain. Proteases of the hepsin/TMPRSS family are predominantly expressed in fetal liver and kidney, prostate, lung as well as in the respiratory and parts of the gastrointestinal tracts.

### ***HAT/DESC subfamily***

The HAT/DESC subfamily comprises of five members, which exhibit the simplest modular stem region structure of only one SEA domain. Compared to the wider distribution of the hepsin/TMPRSS subfamily, the expression of HAT/DESC family members is limited to the upper respiratory tract <sup>[68, 69]</sup>.

### ***Matriptase subfamily***

The matriptase subfamily contains three highly homologous proteases, namely matriptase-1 (or simply matriptase), matriptase-2 and matriptase-3 as well as polyserase-1 with a rather atypical structure. Matriptase, matriptase-2 and matriptase-3 have very similar stem regions with one SEA domain, two CUB domains and three (for matriptase-2 and matriptase-3) or four (for matriptase) LDLRA domains. Polyserase-1 has a unique structure with three tandem serine protease domains and the ability to generate three independent serine proteases (i.e. Serase-1, -2 and -3) the third of which is enzymatically inactive due to the lack of the catalytic serine residue <sup>[70]</sup>. Such as EK, matriptase is a very well-studied TTSP and was found to be broadly expressed by epithelial tissues in all organs examined <sup>[71, 72]</sup> and to be involved in several physiological processes as well as disorders and diseases. Expression of the other members is more restricted.

### ***Corin subfamily***

The smallest TTSP subfamily is the corin subfamily with corin as the single member so far. The stem region is composed of 11 domains, including two frizzled domains, eight LDLRA domains and one SRCR domain. Corin is mainly expressed in the heart and associated with hypertension, heart and renal diseases <sup>[73]</sup>.

### **2.3.3. Modulation of TTSPs activity**

The membrane anchoring of TTSPs contributes to cellular localization and trafficking. Membrane polarity and the epithelial to mesenchymal transition (EMT) influence the pericellular distribution. In polarized epithelial monolayers, plasma membranes are separated into apical and basolateral surfaces. Whereas matriptase is present on basolateral membranes <sup>[74, 75]</sup>, TMPRSS2 is localized to apical membranes <sup>[76, 77]</sup>. The loss of membrane polarity, as it occurs in tumor cells, is thereby associated with mislocation. Due to ectodomain shedding <sup>[78]</sup>,

also several soluble TTSPs, including EK <sup>[79]</sup>, matriptase <sup>[80, 81]</sup> and human airway trypsin-like protease (HAT) <sup>[82]</sup> have been detected.

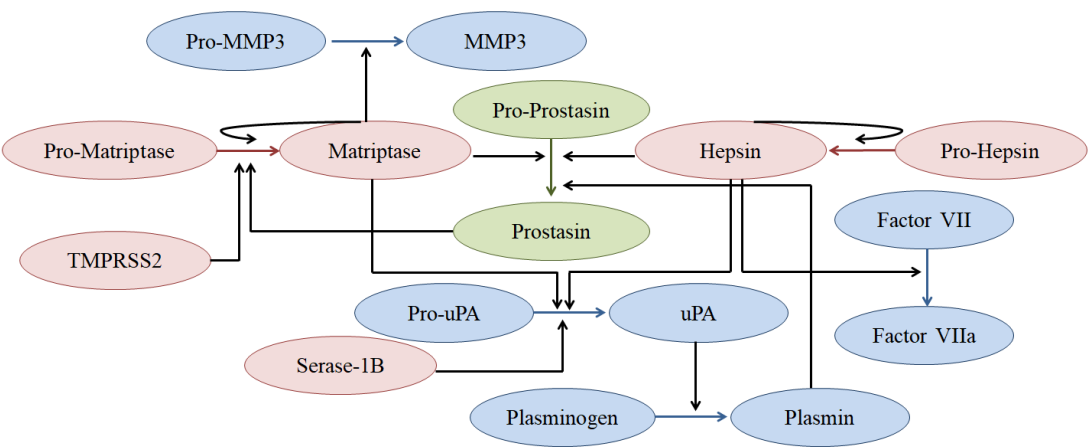
Not only the SPD, but also extracellular domains (stem region) were found to affect protease activity. In case of matriptase, and potentially other similar members, zymogen activation is not only dependent on the activation cleavage after the respective arginine residue but also requires proteolytic processing at a conserved glycine residue in the SEA domain. In addition, mutations or deletions within the LDLRA domain, as well as lacking PTMs such as N-linked glycosylation in the CUB domain were found to restrict matriptase activation <sup>[72, 83]</sup>. Another example represents corin: a single nucleotide polymorphism in the frizzled domain <sup>[84]</sup> as well as the glycosylation status of the SPD were found to affect zymogen activation <sup>[85, 86]</sup>. Crystallographic characterization of a hepsin variant containing the complete extracellular domain revealed that the SRCR domain was rigidly bound to the back of the SPD <sup>[87]</sup>. It has been discussed that the stem region is an interacting partner of the exposed SPD surface and contributes to an appropriate orientation of the active site cleft <sup>[88]</sup>.

Each stem domain seems to contribute uniquely to the respective TTSP cellular localization, activation and inhibition and plays an important, yet not fully identified, role in the regulation of TTSP mediated pericellular proteolysis. Profound insight into different TTSPs is given in several comprehensive reviews <sup>[68, 78, 89-91]</sup>. Although much work has been done since the EK structure has been identified as the first TTSP in 1994, and several TTSP members are already very well characterized, still little is known about other members such as TMPRSS2.

#### 2.3.4. The role of TTSPs in biological regulation

As most of the trypsin-like serine proteases, also zymogens of the TTSPs require proteolytic cleavage to achieve proteolytic activity. This zymogen activation produces a prompt and irreversible response to a physiological stimulus and initiates new physiological functions. Single-step activation reaction or a consecutive series (i.e. cascade) is possible. For the latter, at least two consecutive proteolytic reactions occur with one protease zymogen being the substrate of another previously activated protease. The involvement of serine proteases in zymogen cascades has been demonstrated quite early with the most prominent example EK. The protease is located in the duodenum and catalyzes the crucial conversion of trypsinogen to trypsin <sup>[92]</sup> which is essential for normal digestion and nutritional well-being <sup>[93]</sup>. Trypsin, in turn, initiates a cascade of proteolytic reactions in order to activate additional pancreatic zymogens such as chymotrypsinogen. *In vitro*, trypsin is able to cleave nearly every serine protease zymogen at the respective activation cleavage site, since it cleaves exclusively C-terminal to an arginine or lysine residue <sup>[94, 95]</sup>. However, *in vivo* activation is most likely

catalyzed by specific proteases or is a result of auto-activation, whereas the protease activities are often regulated by endogenous inhibitors. In the case of many TTSPs, these specific endogenous activating proteases are still unknown. Figure 2-8 gives an example of the involvement of some TTSPs (red) in the blood coagulation (factor VII activation by hepsin <sup>[96]</sup>), extravascular fibrinolysis (pro-uPA activation by matriptase <sup>[81]</sup>) and the metalloproteinase pathways (pro-MMP3 activation by matriptase <sup>[97]</sup>).



**Figure 2-8: Membrane anchored serine proteases participate in zymogen cascades.** *In vitro* and *in vivo* data suggest that TTSPs intersect the blood coagulation cascade (factor VII activation), fibrinolysis (pro-uPA activation) and the metalloproteinase pathways (pro-MMP3 activation). Red: TTSPs, green GPI anchored serine proteases, blue: non-membrane anchored serine proteases Figure was taken and modified from literature <sup>[89]</sup>.

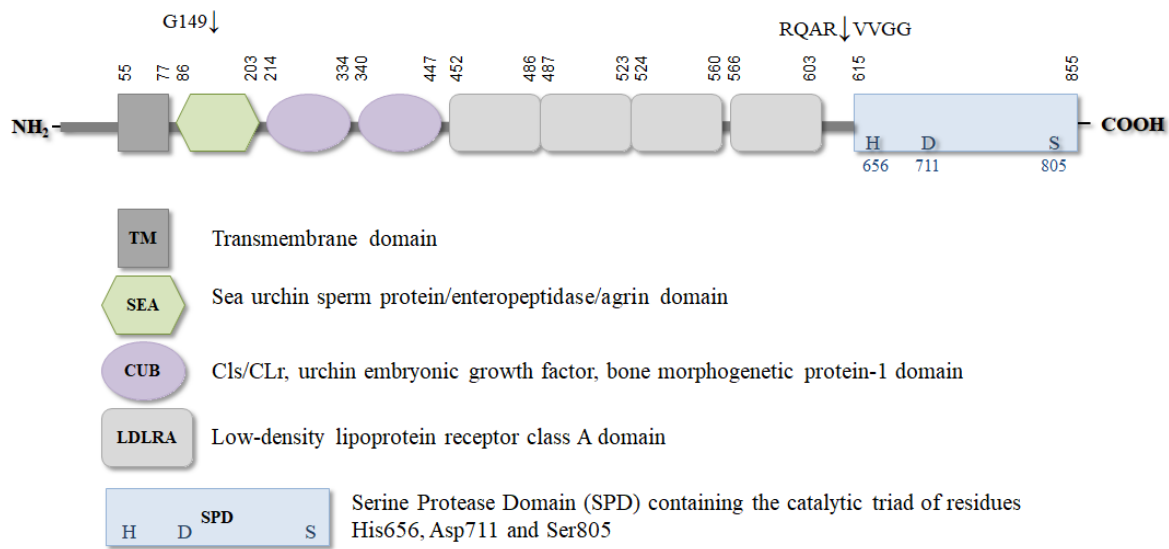
### 3. Preparation and characterization of recombinant matriptase SPDs

#### 3.1. Introduction to matriptase preparation and characterization

The membrane-anchored trypsin-like serine protease matriptase-1 belongs to the matriptase subfamily of TTSPs (Figure 2-7). Other names for matriptase-1 are membrane-type-serine-protease 1 (MT-SP1), tumor-associated differentially expressed gene-15 (TADG-15), TMPRSS14, suppression of tumorigenicity 14 protein and epithin for the mouse analogue. The gene designation is *suppression of tumorigenicity 14 (ST14)*. On protein level, the name matriptase will be used for sake of simplicity in this thesis.

##### 3.1.1. ST14 gene organization and expression

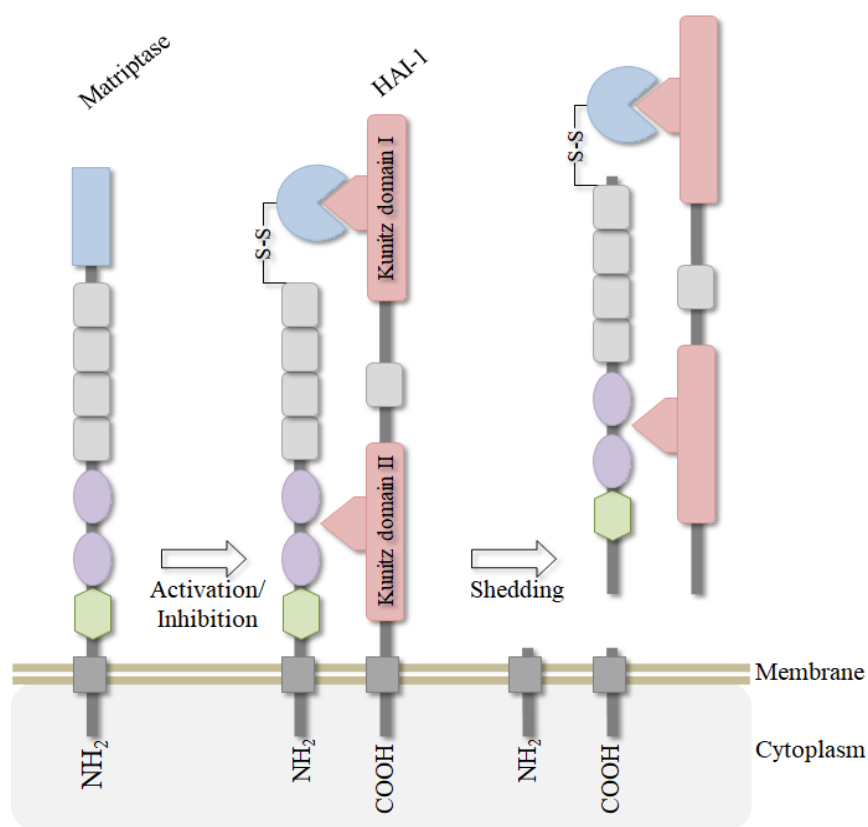
Matriptase is a glycoprotein of 855 amino acids (aa) that consists of a short N-terminal cytoplasmic tail (aa residues 1-54), a transmembrane domain (aa 55-77) and a C-terminal extracellular portion (aa 78-855) (see Figure 3-1).



**Figure 3-1: Schematic overview of the mosaic protein structure of matriptase.** Matriptase comprises of an N-terminal cytoplasmic tail, a transmembrane domain and a C-terminal extracellular portion that includes one SEA domain, two CUB domains, four LDLRA domains and the peptidase domain (SPD). Matriptase is expressed as a zymogen that requires two activation cleavages. Cleavage sites are indicated with an arrow (↓).

The extracellular portion comprises one SEA domain (residues 86-203), two CUB domains (aa 214-334 and aa 340-447), four LDLRA domains (residues 452-486, 487-523, 524-561, and 566-603) and the C-terminal trypsin-fold S1 SPD (residues 615-855) possessing the catalytic triad of His656, Asp711, and Ser805 (His57, Asp102 and Ser195 in chymotrypsinogen numbering). The non-catalytic modules appear to play an essential role in the cellular

localization, activation, endogenous inhibition and likely the specificity of matriptase against natural substrates <sup>[98]</sup>. The stem region also contains four putative asparagine-linked glycosylation sites and a potential integrin-binding RGD motif in the first CUB domain (UniProtKB - Q9Y5Y6). Matriptase is synthesized as a single-chain 95 kDa zymogen that requires physiological activation. The complex maturation process involves two sequential proteolytic cleavages as well as interactions with hepatocyte growth factor activator inhibitors (HAI-I and HAI-II). The first proteolytic processing occurs after the glycine residue Gly149 of the SEA domain within the endoplasmic reticulum (ER) or the Golgi apparatus. Hydrolysis of the peptide bond is thought to be of a conformation driven non-enzymatic and spontaneous manner as it has been shown for human MUC1 transmembrane mucin <sup>[99]</sup>. The SEA-cleaved 70 kDa zymogen remains attached to the membrane by non-covalent interactions <sup>[100]</sup> and is associated with HAI-I. This matriptase/HAI-I complex is translocated to the cell surface where the second proteolytic cleavage event occurs between Arg614 and Val615 within a highly conserved activation cleavage site of RQAR↓VVGG (for schematic representation see Figure 3-1 and Figure 3-2).



**Figure 3-2: Schematic representation of matriptase activation, inhibition and possible shedding from the cell surface.** During activation, matriptase is first cleaved after Gly149 in the SEA domain before being translocated to the cell membrane. After the autocatalytic activation cleavage after Arg614, a 120 kDa complex with HAI-1 will be simultaneously formed. Due to shedding, this inactive matriptase/HAI-1 complex can be released from the membrane to the extracellular space <sup>[98]</sup>.

Mutation of any catalytic triad residue prevents the activation of the SPD indicating an intrinsic autoproteolytic activation mechanism of matriptase [72]. The disulfide-linked two-chain fully active matriptase is subsequently inhibited by HAI-1 but can be shed from the cell surface [98]. Due to ectodomain shedding, inactive matriptase/HAI-I or HAI-II complexes were also found as soluble complexes in milk [101, 102].

The auto-activation cleavage seems to be influenced by the stem region, posttranslational modifications, the cellular localization of the protease and matriptase zymogen activity [103, 104]. Even though trypsin-like serine proteases are classically synthesized as inactive zymogens some exceptions exist. This includes pro-tPA, which has a significant zymogen catalytic activity due to special structural features [105]. Zymogen activity of matriptase has been discussed in the literature and was investigated by enzyme kinetic studies with a Biotin-Arg-Gln-Arg-Arg chloromethyl ketone (CMK) and a transgenic mouse model [104, 106, 107]. The difference in activity between a pseudozymogen (imitating the SEA-cleaved 70 kDa zymogen form) and the 27 kDa activated SPD (R614 cleaved) of rat-matriptase has been shown to be unusually small with only a 27-fold higher activity of the SPD using a synthetic peptide substrate [108, 109]. A more recent study of TAMBERG and colleagues confirmed these results: a zymogen-locked (R614A mutated) matriptase form possessed 3 % of matriptase-SPD activity (zymogenicity factor of 33-fold). Both matriptase forms were recombinantly expressed in *P. pastoris* (zymogen) or *E. coli* (SPD) and measured in enzyme kinetic experiments using a chromogenic peptide substrate [110]. The matriptase zymogen has been shown to possess adequate levels of catalytic activity to not only facilitate autoactivation (in order to achieve full activity) but also to carry out its endogenous function including the activation of other zymogen forms in proteolytic cascades.

### **3.1.2. Physiological function of matriptase**

Matriptase was first isolated in 1993 as a secreted gelatinase expressed by cultured human breast cancer cells [111]. Since then, matriptase expression has been found in many vertebrates. However, unlike most other TTSPs, matriptase is widely expressed in epithelial compartments of many embryonic and adult tissues. These include several organs such as skin, breast, pancreas, lung, intestine, reproductive and glandular epithelia [72, 103] as well as immune cells like monocytes, macrophages and lymphocytes [103, 112, 113]. More recently, matriptase mRNA has been detected in several regions of the brain with an enrichment in neurons and at higher levels in younger compared to older individuals [114]. The essential role of matriptase in the epithelia was initially discovered in matriptase null mice which revealed aberrant skin development resulting in death shortly after birth due to dehydration from compromised skin

integrity <sup>[115]</sup>. These mice are unable to process pro-filaggrin, an important protein in the arrangement of keratin filaments in the skin <sup>[116]</sup>. Furthermore, matriptase-deficient mice exhibited a defect in the formation of lamellar granules. These specialized secretory vesicles contain lipid material which is required for the formation of extracellular lipid lamellae within the cornified layer. Last, the mice displayed impaired formation of epidermal tight junctions within the granular layer <sup>[117, 118]</sup>.

### 3.1.3. Endogenous matriptase inhibitors

On protein level, matriptase activity is tightly regulated by the cognate cell surface Kunitz-type serine protease inhibitors HAI-I and HAI-II. The dysregulation of matriptase is associated with several diseases including osteoarthritis <sup>[119, 120]</sup>, ichthyosis and various kinds of epidermal cancer <sup>[121-124]</sup>. A mutation of the glycine residue Gly827 (216 after chymotrypsin numbering) to arginine (G827R) within the active site of the SPD, which is involved in the binding of the P3 residue of substrates and substrate-analogue inhibitors, led to significant reduction of matriptase activity *in vitro* and has been found in individuals diagnosed with autosomal recessive ichthyosis with hypotrichosis (ARIH). Characteristics of ARIH include thickened, scaly and shiny skin with follicular hypoplasia <sup>[71, 125, 126]</sup>. Hypomorphic mice show high similarities to the phenotype of patients with ARIH, but in contrast to matriptase null mice, they have a normal survival <sup>[71]</sup>. Besides HAIs, the secreted serpins antithrombin, alpha1-antitrypsin and alpha2-antiplasmin complexes were reported to inhibit matriptase activity especially in cells with low expression levels of HAIs, such as leukocytes <sup>[102]</sup>.

### 3.1.4. Substrate specificity

The substrate specificity of matriptase was determined by the topology of the binding pocket based on several crystal structures by using a positional scanning-synthetic combinatorial library (PS-SCL) and by substrate phage techniques. The preferred cleavage sequences were found to be P4(Arg/Lys)-P3(Xaa)-P2(Ser/Ala)-P1(Arg)↓P1'(Ala) and P4(Xaa)-P3(Arg/Lys)-P2(Ser/Ala)-P1(Arg)↓P1'(Ala), with Xaa being a non-basic amino acid <sup>[81, 127-129]</sup>. Thus, in contrast to trypsin, matriptase does not indiscriminately cleave protein or peptide substrates after lysine or arginine residues but prefers recognition of additional basic residues especially in P4 or P3 position. Based on predictions, the first matriptase substrates were identified *in vitro*, including the pro-form of urokinase-type plasminogen activator (pro-uPA), the pro form of hepatocyte growth factor (pro-HGF) and the protease-activated receptor 2 (PAR-2). All these extracellular surface proteins contain the preferred P4-P1' matriptase cleavage site of PRFK↓I, KQLR↓V and SKGR↓S for pro-uPA, pro-HGF and PAR-2 respectively <sup>[130]</sup>.



### 3.1.5. Matriptase and cancer

Over the last decade, several substrates had been shown to be activated by matriptase *in vitro* of which many are linked to cancer progression. Examples include the previously mentioned substrates pro-HGF <sup>[131]</sup>, pro-uPA <sup>[81, 131, 132]</sup> and PAR-2 <sup>[81, 133]</sup>, but also the pro macrophage-stimulating protein (pro-MSP-1) <sup>[112]</sup>, the matrix metalloproteinase-3 (MMP-3) <sup>[97]</sup> and the epidermal growth factor receptor (EGFR) <sup>[134]</sup>. For the latter, matriptase has shown to cleave the EGFR extracellular domain in FT-293 cells at the same sites it is cleaved by prostasin. Matriptase and prostasin can activate the pro-forms of each other, respectively (Figure 2-8). By this cleavage, the ligand-binding domains are hampering e.g. mAB drugs (such as Erbitux) to be effective. EGFR dysregulation in either way (up- or down-regulation) is causative for many different types of solid tumors. A significant induction of matriptase expression was however not observed in FT-293 cells <sup>[134]</sup>.

The dysregulation of matriptase has been associated to a wide range of benign and malignant tumors of epithelial origin such as breast cancer <sup>[80, 111]</sup>, prostate carcinoma <sup>[135]</sup>, ovarian cancer <sup>[136]</sup>, cervical cancer <sup>[124]</sup>, gastrointestinal cancers and kidney tumors <sup>[103, 137]</sup>. For most cancers (e.g. breast, cervical, ovarian and prostate cancer), matriptase overexpression at mRNA or protein level has been linked with tumor progression <sup>[121-124, 135, 136, 138]</sup>. In contrast, other studies on breast, gastric, and colorectal cancer <sup>[139-141]</sup> reported a significant down-regulation of matriptase expression levels. In total, approximately 85 % of all cancers originate from epithelial cells. This indicates the potential value of matriptase as a prognostic marker and as a possible therapeutic target for cancer treatment. Most proteases involved in carcinogenesis are expressed by the connective tissue, supporting the epithelia. Unlike them, matriptase is expressed by the transformed epithelial cells themselves <sup>[142]</sup>. A recently published review of the LIST group gives a more detailed summary of the role of matriptase in cancer <sup>[143]</sup>. Besides its relevance in cancer, matriptase has been identified as a critical driver for other diseases, including osteoarthritis <sup>[144]</sup> and virus infections such as the influenza viruses H1N1 and H9N2 <sup>[63, 145]</sup> and the human immunodeficiency virus (HIV) <sup>[146]</sup>.

### 3.1.6. Matriptase in direct and indirect IAV activation

Matriptase is able to cleave the hemagglutinins of certain influenza viruses directly or potentially contributes to virus infections by proteolytical activation of other proteases, which subsequently can cleave hemagglutinin <sup>[64]</sup>. For virus infectivity, the influenza virus hemagglutinin must undergo proteolytic activation either in the TGN (secretory pathway), at the cell surface, in the extracellular space or within endosomes. The HA-activation is a prerequisite that the virus achieves fusion competence. Matriptase was detected in human

bronchial epithelial cells and was characterized as a potential protease for proteolytic activation of hemagglutinin in the human respiratory tract <sup>[145]</sup>. In addition, soluble matriptase was found in nasal wash samples of healthy human donors. Based on the ability to be shed from basolateral and apical faces of polarized epithelial cells, matriptase may also serve as a functional extracellular protease <sup>[64]</sup>. Since the influenza virus infects from the apical surface and the membrane-bound form of matriptase primarily resides on the basolateral face of epithelial cells, it has been suggested that the shedded active matriptase form is more likely to be involved in HA cleavage in intact epithelia.

Several studies demonstrated that matriptase efficiently cleaves HA subtypes H1 and H9 (of H1N1 and H9N2) but also shows low activity towards H2- and H3-HA (of H2N2 and H3N2) <sup>[63, 64, 145]</sup>. The knockdown of matriptase in Calu-3 cells (lung cells) led to reduced growth of H1N1 strains <sup>[145]</sup>. Virus replication of H9N2 was demonstrated in the kidney of chickens, where matriptase is highly expressed, and associated to increased mortality <sup>[147]</sup>. BARON and colleagues data support H9-HA activation in two kidney cell lines (MDCK-II and chicken embryo kidney (CEK) cells) of some H9N2 strains and additionally gave evidence of a possible matriptase involvement in H9-HA cleavage: Overexpressed H9-HA with R-S-S-R or R-S-R-R motif was cleaved by recombinant matriptase SPD *in vitro* <sup>[63]</sup>. Virus replication of some H9N2 strains was also detected in the brain of mice <sup>[148]</sup> which is compatible with matriptase expression in neural progenitor cells and neurons <sup>[149]</sup>. In a more recent study, matriptase, among seven other serine proteases, was investigated for cleavage of various peptides mimicking the cleavage motifs of distinct HA subtypes *in vitro*. Only matriptase (besides the ubiquitously cleaving trypsin) showed cleavage activity towards H10-HA subtype (H10N8) <sup>[150]</sup>. Whereas matriptase cleavage activity towards H10 was only weak, another study showed 80 % HA<sub>0</sub> cleavage in H10 expressing cells by TMPRSS2 <sup>[151]</sup>. In accordance with the latter study, recent *in vivo* data even suggest a TMPRSS2 dependency of H10-HA cleavage <sup>[152]</sup>. Besides H9N2, H7 and H5-HA of H7N9 and H5N1 HPAIV strains containing the multibasic cleavage motif R-K-K-R were efficiently cleaved by matriptase in the study of STRAUS and WHITTAKER. H7N9 and H5N1 are highly infectious IAV types causing systemic infections in birds. H7- and H5-HA were thought to be mainly activated by furin and furin-like proprotein convertases <sup>[153]</sup> and are only rarely transmitted to humans so far. However, if transmitted, the resulting infections lead to severe disease exhibiting a mortality rate of 60 % in the case of H5N1. The H5N1 is one of the most critically observed influenza strains by the authorities and is discussed as a possible elicitor for the next influenza pandemic <sup>[154, 155]</sup>. Whereas H7N9 was suggested to be also mainly activated by TMPRSS2 *in vivo*, HPAIV H5N1 was still successfully activated in TMPRSS2 deficient mice lungs resulting in animal death <sup>[61]</sup>. The first H9N2

lineage that caused human infection showed genetic relations to isolates from H5N1 IAV<sup>[156]</sup>. This, and the fact, that H9N2 isolates had shown dual or human-like sialic acid receptor specificity classify H9N2 as a strain with a high potential to infect humans<sup>[157]</sup>. Thus, cleavage activity of matriptase towards the discussed HA subtypes makes matriptase also an interesting target for influenza drug development not only to reduce mortality in poultry but also to be prepared for a future pandemic outbreak in humans.

Besides TTSPs, including matriptase, TMPRSS2, TMPRSS4 and HAT, extracellular soluble serine proteases such as tryptase clara<sup>[56]</sup> and extracellular thrombolytic proteases such as plasmin, uPA and plasma kallikrein were also detected to activate certain HAs *in vitro*. In addition, some of these proteases, were found in the respiratory tract when the tissue was damaged due to influenza virus infections<sup>[158]</sup>. Matriptase, in turn, has shown cleavage activity towards some of these zymogen forms *in vitro*<sup>[64, 131]</sup>, underlining also an indirect impact on influenza virus activation.

### **3.1.7. Recombinant production of matriptase in *E. coli***

For *in vitro* investigations of enzyme activity, ligand binding, crystallization, protein interactions and other functions, it is crucial to produce sufficient amounts of pure matriptase. The isolation of enzymes from its natural source oftentimes suffers from low amounts of the protein, resulting in cost-intensive and laborious purification protocols with unsatisfying results. Hence, recombinant protein production (RPP) has become the method of choice. In the last years, various expression systems were developed based on different host organisms including prokaryotes, yeast, insect- or mammalian cells. Each system has advantages and disadvantages and no "golden" standard protocol is available.

Recombinant production and purification of the matriptase SPD have been reported from different groups including the expression in *E. coli*<sup>[159-161]</sup>, and the two yeast organisms *Saccharomyces cerevisiae* (*S. cerevisiae*)<sup>[162]</sup> and *Pichia pastoris* (*P. pastoris*)<sup>[163]</sup>. The protein expression in *E. coli* is, compared to higher organisms, inexpensive, offers rapid culture times and the ability to achieve high biomass and high protein yields. Furthermore, the system can be easily modified and many well-investigated expression systems are commercially available. In case of matriptase and many other eukaryotic proteins, the main advantage to choose yeast cells over *E. coli* is their ability to perform posttranslational modifications (PTMs), including disulfide-bond (DB) formation, a major PTM that dictates the correct fold of a protein, provides protein stability and biological activity. Among others, the lacking availability of disulfide bond formation of *E. coli* is one reason that the overproduction of matriptase leads to the formation of so-called inclusion bodies, particles of densely packed and denaturized, hence inactive

protein molecules that accumulate in the *E. coli* cytoplasm. The protein refolding from denaturated inclusion bodies into the bioactive form is a cumbersome procedure and often results in poor recovery of active protein or is even not possible at all. However, the successful recovery from *E. coli* inclusion bodies of several transmembrane serine proteases has been described in the past, including matriptase <sup>[159, 160]</sup>.

Various modified *E. coli* expression systems are nowadays available to overcome the initial drawbacks, such as inability to form DBs, for example by redirecting the proteins to the bacterial periplasm, an oxidizing compartment with enzymes catalyzing DB formation and their isomerization <sup>[164, 165]</sup>. However, this approach often results in very low protein amounts which require intensive optimization. In the interest of efficiency, it is advantageous to utilize the simplest system possible for protein production. In our laboratories, solely the recombinant protein production using *E. coli* as a host organism is established. Furthermore, active matriptase has already been successfully recovered and purified from inclusion bodies. As a consequence, *E. coli* was also the host of choice in this thesis.

### **3.1.8. Aims for matriptase production and characterization**

Matriptase is one of the best investigated TTSPs and has been linked to many diseases including cancer and influenza virus infections. This chapter describes the recombinant production of different matriptase variants in *E. coli*, including the characterization and crystallization.

In detail, the goals of this chapter are:

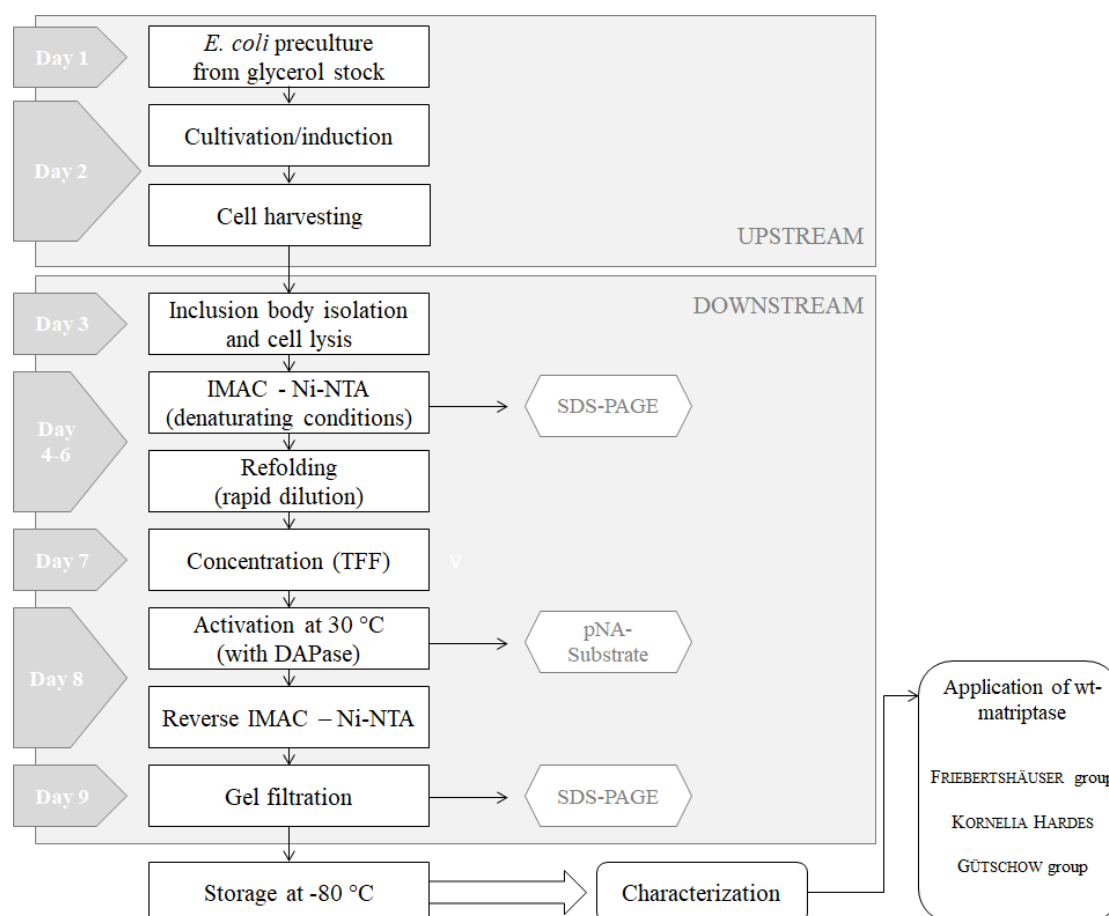
- the preparation of wild-type matriptase SPD using an existing expression system and protocol <sup>[159]</sup>,
- the establishment of a novel more cost-efficient expression system of a C122S mutated matriptase SPD variant (collaboration with VIKTOR MAGDOLEN, TU Munich),
- the adequate preparation of pure C122S-matriptase for further experiments,
- the comparison of both matriptase variants using enzyme kinetic parameters,
- and the establishment of C122S-matriptase crystallization in our facilities.

## 3.2. Results and discussion of matriptase preparation and characterization

The heterologous production of recombinant proteins in *E. coli* is a common method to produce large quantities of the target protein, which is essential for its proper functional and structural characterization. So far, all protocols for the preparation of matriptase using *E. coli* resulted in the formation of inclusion bodies. As a result, an additional refolding step is required. Even though protocols for isolation, purification and refolding of matriptase exist, such protocols need to be established and optimized in each laboratory independently. In this chapter, two different matriptase constructs were used for protein production in combination with two different commercially available *E. coli* expression systems. In both cases, a tag of six histidines was added N-terminally to the SPD domain for purification purposes. As a TTSPs, matriptase needs a free N-terminus of the amino acid residue 16 (chymotrypsinogen numbering) for proper folding of the active site and hence activity. Therefore, the following purification and refolding as inactive form, matriptase was activated by removing all amino acid residues upstream of residue Val615, which corresponds Val16 according to the chymotrypsinogen numbering. In doing so, also the six histidines (6×His-tag) were removed after initial affinity purification.

### 3.2.1. Production of wild-type (wt) matriptase

For the expression of wt-matriptase using the pET/*BL21* expression system (Merck, Darmstadt), a glycerol stock from 2008 was used. The cloning of the expression plasmid and a protocol for the preparation of the recombinant protein had been described earlier <sup>[159]</sup>. In the course of this thesis, matriptase was isolated, purified, refolded and activated based on this protocol with only slight modifications as described in the experimental part (Chapter 7). For an overview, the workflow is summarized in Figure 3-3. General activity of the prepared wt-matriptase was visually monitored using the chromogenic substrate Mes-DCha-Gly-Arg-pNA (synthesized in-house). Since the protocol has already been established in our lab and the production of wt-matriptase did not show any deviations from earlier documented procedures, the individual preparation steps are not shown in detail. Following the last chromatographic gel filtration step, active wt-matriptase was flash-frozen in liquid nitrogen at a protein concentration of 0.3 mg/ml and stored at -80 °C.



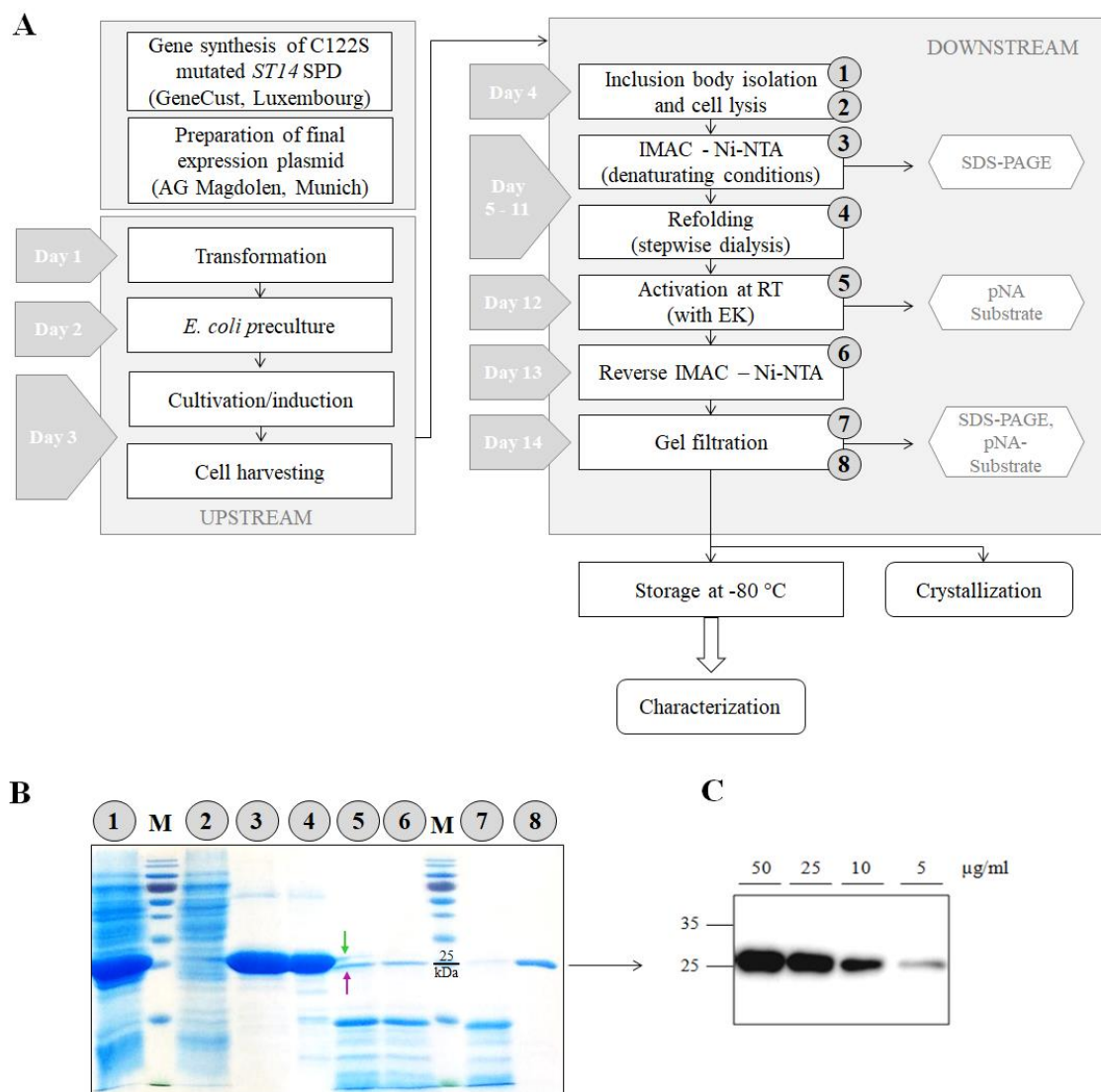
**Figure 3-3: Workflow of the wt-matriptase production.** Wt-matriptase producing *E. coli* was cultivated from a glycerol stock and purified from inclusion bodies, refolded, concentrated and activated following instructions as described earlier <sup>[159]</sup>. After final gel filtration, wt-matriptase was stored longterm at -80 °C until it was used in further experiments. TFF: Tangential flow filtration.

The wt-matriptase batch prepared within this thesis was applied in various experiments for the HA activation of certain influenza strains by DR. KORNELIA HARDES from our group and the group of PROF. FRIEBERTSHÄUSER (Institute of Virology, Department of Medicine, Philipps Universität Marburg). Furthermore, in the group of PROF. GÜTSCHOW (University Bonn) wt-matriptase was used to demonstrate the efficacy of a fluorescent-labeled phosphono bisbenzguanidine as an activity-based probe for matriptase by electrophoresis and fluorescence detection in HPLC <sup>[166]</sup>. These experiments verified the activity and high purity of the prepared wt-matriptase.

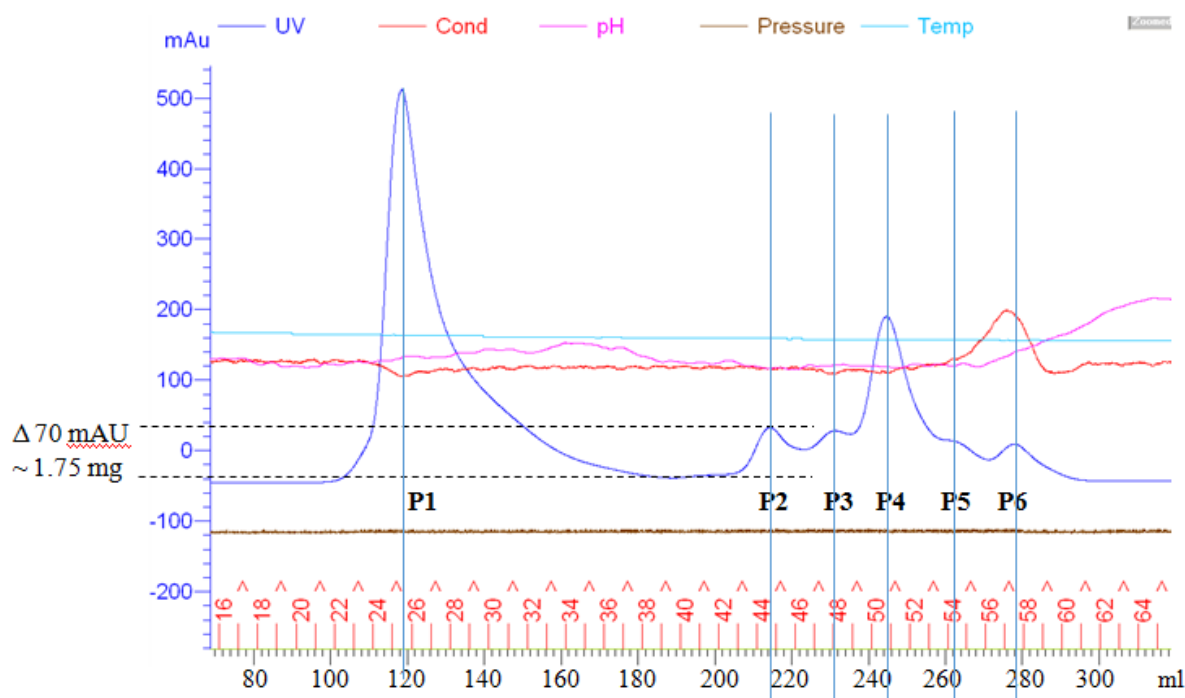
### 3.2.2. Production and characterization of C122S-matriptase

For the expression of C122S-matriptase SPD, *M15[PREP4]* *E. coli* cells were transformed with the respective pQE-30\_C122S\_matriptase expression plasmid (kindly prepared and provided by VIKTOR MAGDOLEN, TU Munich). C122S-matriptase was isolated, purified, refolded and activated following a newly established protocol (see Chapter 7). Figure 3-4 A gives a brief overview of all steps of the C122S-matriptase preparation process including the in process

monitoring in form of SDS-PAGE (Figure 3-4 B) and the analysis of the final product by western blotting (Figure 3-4 C). A typical gel-filtration chromatogram as the final step in C122S-matriptase preparation is shown in Figure 3-5.



**Figure 3-4: Description of the C122S-matriptase production process and its monitoring.** **A:** Workflow of the C122S-matriptase preparation process including all main steps: creation of the expression plasmid, the upstream process (transformation and cultivation) and the downstream process including the purification, refolding and activation. **B:** SDS-PAGE-analysis to monitor C122S-matriptase production. Samples were taken during every matriptase preparation step, separated in a 15 % SDS gel and stained with 1x Roti®Blue solution overnight. The photo was taken after 30 min incubation in the destaining solution. Lane 1-8: 1: cell lysate (8 M urea); 2: flow through after 1<sup>st</sup> nickel column; 3: elution peak of 1<sup>st</sup> nickel column; 4: after refolding via step wise dialysis (urea free); 5: activation with 5xHis-tagged enteropeptidase; 6: flow through of 2<sup>nd</sup> nickel column (active matriptase; reverse IMAC); 7: 1<sup>st</sup> peak of gel filtration chromatography (120 ml); 8: 2<sup>nd</sup> peak of gel filtration chromatography containing active matriptase (after 215 ml; immediate yellow staining by turnover of substrate Mes-DCha-Gly-Arg-pNA). Green arrow: non-activated 6xHis-tagged matriptase (calculated size: 28.4 kDa); pink arrow: after removal of 6xHis-tag (activated form; calculated size: 26.4 kDa). M: PageRuler Prestained 10-180 kDa marker. **C:** Western Blot Analysis. After manufacturing, 10 µl of variable concentrations (50, 25, 10 and 5 µg/ml) of the final C122S-matriptase product were separated in a 12 % SDS Gel and blotted to a PDFV membrane. After incubation with a monoclonal α-matriptase primary and HRP-conjugated secondary antibody, matriptase was visualized using the substrate SuperSignal™West Dura in the ChemiDoc™ Imaging System. One single concentration-dependent signal appeared at approximately 26-27 kDa.



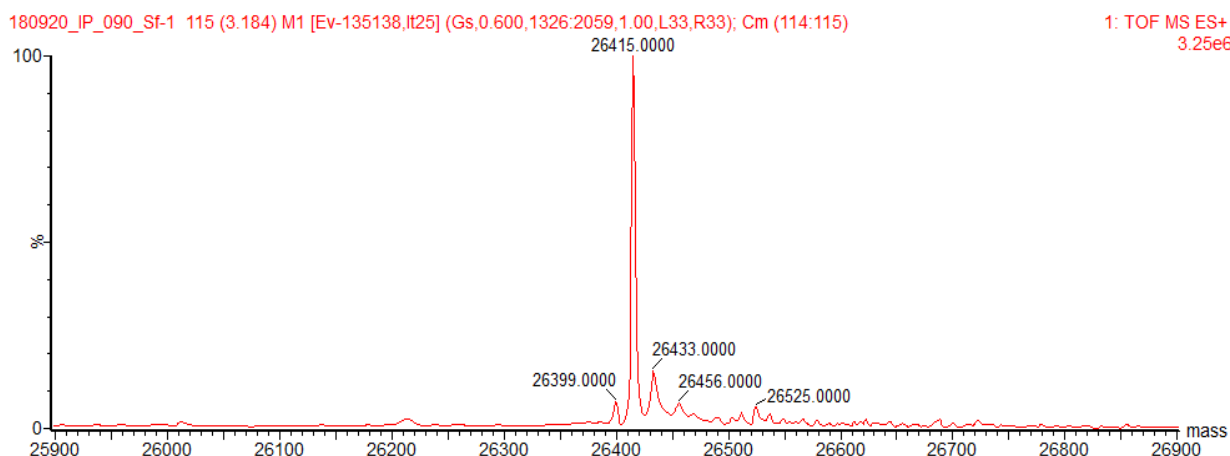
**Figure 3-5: Gel filtration chromatogram of purified, refolded and activated C122S-matriptase using a Superdex75 prep grade column.** Blue: UV absorption at 254 nm; red: conductivity; pink: pH; light blue: temperature; brown: pressure. The x-axis is given in ml, the y-axis shows UV absorption in mAU. The sample was separated into six peaks (P1-P6). Peak 2 (P2) was visually identified as the peak with the highest matriptase activity using the Mes-DCha-Gly-Arg-pNA substrate. Peak 3 showed low matriptase activity, and no activity was detected in fractions of peaks 1, 4, 5 and 6.

The SDS-gel (Figure 3-4 B) shows that C122S-matriptase was successfully prepared following the above-mentioned workflow (Figure 3-4 A). The denaturated 6xHis-tagged protein can be isolated from the crude cell lysate by IMAC (lane 1-3), refolded by stepwise dialysis (lane 4) and activated by incubation with recombinant 5xHis-tagged enteropeptidase (lane 5). The activated protein was separable from the non-activated portion by reverse IMAC (lane 6) and finally be cleared up by gel filtration from misfolded or digested protein pieces as a result of the activation step (lane 7-8). The gel filtration chromatogram (Figure 3-5) shows six different peaks, with the first and largest peak after 120 ml (P1) and several non-completely separated multiple peaks at 215 ml (P2), 230 ml (P3), 245 ml (P4), 262 ml (P5) and 278 ml (P6). All peaks were collected separately and tested for activity by incubation with the substrate Mes-DCha-Gly-Arg-pNA. The faster the initial colorless solution turns yellow, the more catalytically active matriptase is expected to be in the sample. The fractions containing peak 2 (peak maximum at 215 ml) showed immediate yellow staining, indicating the highest concentration of active matriptase. Fractions covering peak 3 stained light yellow after several seconds. Fractions of peak 1, 4, 5 and 6 did not show yellow staining after incubation with the chromogenic substrate, implying that these fractions do not contain active matriptase. The first peak (P1) is the by far largest one and therefore expected to be composed of molecules of the



largest size. However, after SDS-PAGE analysis, the peak was identified as a mixture of several smaller protein fragments between 10 kDa and 15 kDa (Figure 3-4 B, lane 7). A possible explanation for the early elution is the aggregation of these small protein fragments and hence the “imitation” of a larger molecule in the gel filtration chromatogram.

The single protein band in lane 8 of Figure 3-4 B represents peak 2 after gel filtration. It was finally verified as matriptase in western blot experiments using a specific monoclonal  $\alpha$ -matriptase primary antibody (Figure 3-4 C). Only a single concentration-dependent signal is visible at the same height of approximately 27 kDa. This final C122S-matriptase product was further analyzed via positive electrospray ionization time-of-flight mass spectrometry (TOF MS ES+). The analysis was performed using C122S-matriptase in a concentration of 1 mg/ml in storage buffer by employees of the bioanalytic service division of the chemistry department at Marburg University. As shown in Figure 3-6, the most prominent peak of 26415 m/z exactly matches the expected calculated mass of 26421 Da, based on the C122S-matriptase primary amino acid sequence under consideration of the formation of three disulfide bridges within the protease domain ( $-6$  m/z).



**Figure 3-6: TOF MS ES+ spectrum of C122S-matriptase.** Purified C122S-matriptase was used to perform a mass TOF MS ES+ spectrometry by the MarMass bioanalytic service division of the chemistry department at Marburg University. The most prominent peak represents matriptase with a mass of 26.415 kDa. The exact mass matches the calculated mass of 26.421 kDa (based on the amino acid sequence using ExPASy ProtParam) under consideration of the three essential disulfide bridges within the matriptase peptidase domain ( $-6$  m/z).

### 3.2.3. Comparison of wt-matriptase and C122S-matriptase production efficiency

Both matriptase variants were produced according to their respective protocols. During the process, protein concentration was monitored by UV/Vis spectroscopy at reasonable steps to compare the final yields and to reveal "weak" spots of each procedure. Since the C122S-matriptase production cycle was performed several times during this thesis, only one representative cycle has been chosen for comparison.

Production of wt-matriptase started from 6 liters *E. coli* culture, leading to approximately 150 mg protein after a first IMAC purification step with Ni-NTA. During the following steps, including the refolding, desalting and activation, accompanied by filtration, concentration or dialysis steps around two-thirds of precipitated protein were removed. In a second reverse IMAC step again using Ni-NTA as resin, the properly refolded and active wt-matriptase without the N-terminal 6xHis tag was separated from the non-activated protein still containing the six histidines N-terminally. Only approximately one-tenth of the starting protein concentration could be recovered in this step. After a final gel filtration chromatography step using a SEC75 column, possibly aggregated or multimerized protein molecules as well as other impurities were removed and monomeric wt-matriptase molecules were collected in 1 x PBS. In total, 6 mg of pure and active wt-matriptase was prepared from 6 liters initial bacterial culture. This corresponds to a protein recovery of 4 % (Table 3-1).

**Table 3-1: Monitoring of protein yields during the wt-matriptase workflow**

Process step *6 liters <i>E. coli</i> culture	Total protein [mg]	Overall yield [%]
After IMAC (1 <sup>st</sup> nickel column)	150	100
After Refolding and concentration	133	89
After desalting and filtration	48	32
After activation and filtration	43	29
After 2 <sup>nd</sup> nickel column and concentration	14.5	10
After gel filtration (SEC75)	6	4

In the case of C122S-matriptase, approximately 180 mg protein had been isolated via IMAC, starting from 12 liters bacterial culture. After the separation of activated and non-activated C122S-matriptase, 55 % (circa 100 mg) of the initial protein quantity was recovered by reverse IMAC. Following the final gel filtration step, the fractions with active C122S-matriptase (detected using Mes-DCha-Gly-Arg-pNA as substrate) were pooled. Starting from 180 mg protein, 3.5 mg monodispersed and active C122S-matriptase could be recovered in a standard purification cycle, representing 2 % of the initial starting protein concentration (Table 3-2)

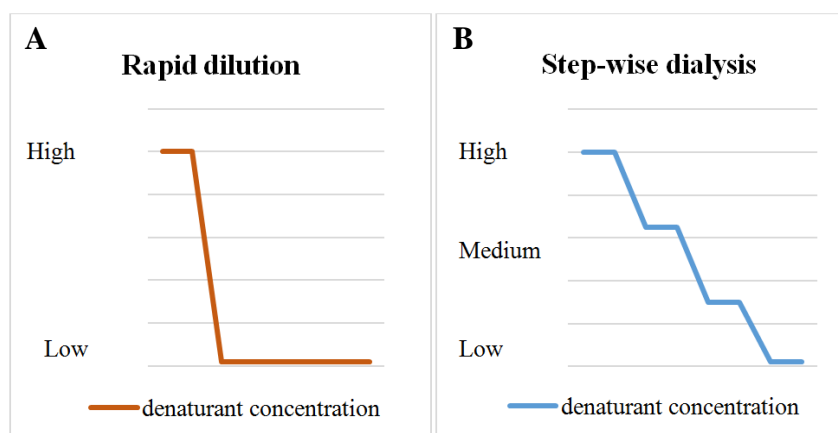
**Table 3-2: Monitoring of protein yields during the C122S-matriptase workflow.**

Process step *12 liters <i>E. coli</i> culture	Total protein (mg)	Overall yield (%)
After IMAC (1 <sup>st</sup> nickel column)	180	100
After dialysis and filtration (refolding)	170	94
After activation and 2 <sup>nd</sup> nickel	100	56
After gel filtration (Sec 75)	3.5	2

The production of the wt-matriptase from the pET-*BL21(DE3)* expression system in combination with the rapid dilution refolding strategy was, with a 4 % recovery, more fruitful than the production of C122S-matriptase using the pQ30E-*M15* expression system and the stepwise dialysis strategy for refolding (2 % recovery). In addition, the total protein amount per liter bacterial culture was higher in the case of the pET-*BL21* system.

The main difference between the two expression systems is the promoter, a short nucleic acid sequence upstream of the target gene, controlling the attachment of the RNA-polymerase to the DNA. The promoter directly controls the amount of generated transcript. In pQE-30 plasmids (C122S-matriptase), gene expression is under control of the IPTG inducible T5 promoter, which is recognized by *E. coli* RNA polymerase of any strain. In pET vectors, gene expression is under control of the IPTG inducible T7 promoter that is only recognized by the T7 RNA polymerase but not the *E. coli* RNA polymerase. The T7 polymerase needs to be provided in another plasmid and is commonly placed in the  $\lambda$ DE3 genome, such as in case of all *BL21(DE3)* bacterial strains from Novagen (Merck, Darmstadt). The T7 polymerase is highly active and exclusively dedicated to the production of target gene transcripts, while the T5 promoter competes with other host promoters <sup>[167]</sup>. Up to 50 % of the total cell protein has been reported to be represented by the target protein when using the T7 promoter - T7 RNA polymerase system <sup>[168, 169]</sup>, making it a popular system for recombinant protein production. Nevertheless, it is also possible that very high transcription levels can direct energy fluxes towards mRNA production and away from protein expression <sup>[170, 171]</sup>. Surely, the promoter could be one possible explanation for higher wt-matriptase yields. However, also age, fitness and colony-to-colony variations are considerable reasons. Other bacterial strains, such as *BL21(DE3)Gold* and *BL21(DE3)RIPL* cells were tested briefly as an expression host for C122S-matriptase and showed more promising results regarding isolated quantity per liter bacterial culture than the *M15* cells. For future expression experiments it is therefore recommended to screen for an appropriate *E. coli* strain in order to increase the protein production and reduce expression costs for C122S-matriptase. In the end, both expression systems deliver sufficient protein to start purification and the main optimization focus should address the recovery yield after refolding of the inclusion bodies. Many different factors can influence the refolding success rate, including protein concentration, refolding temperature, pH and the chosen refolding method including the refolding buffer composition. The main cause of low yields of properly refolded proteins is non-specific aggregations, which derive from interactions of hydrophobic cores between proteins during the refolding process. Regarding protein concentration, it is generally accepted that the lower the final protein concentration, the greater the refolding efficiency. In addition, it has been found that protein aggregation predominantly occurs at medium

concentrations of denaturants due to the formation of an unstable intermediate form, which is more likely to result in misfolded or aggregated protein. In Figure 3-7 A and B, the standard refolding methods rapid-dilution (orange) and stepwise dialysis (blue) are graphically described.



**Figure 3-7: Graphic description of one-step rapid-dilution and stepwise dialysis protein refolding methods.**

**A:** In rapid dilution approaches, the denaturant is removed rapidly by diluting the denatured protein sample usually 1:50 - 1:100 in a denaturant free buffer. Parallel to the denaturant concentration, the protein concentration decreases equally. **B:** In stepwise dialysis refolding approaches, the denaturant is removed stepwise by multiple buffer exchanges. After each buffer exchange, the protein solution is allowed to equilibrate. During the procedure, protein concentration remains constant and buffer components are either removed or introduced.

In rapid-dilution, the concentrated and denatured protein is introduced rapidly into a denaturant free refolding buffer creating only a brief aggregation-promoting environment in the beginning. In contrast, during stepwise dialysis the denaturant concentration is lowered slowly resulting in a longer phase with critical middle denaturant concentration. However, the stepwise buffer exchange allows control over the conditions and timepoint when protein is exposed to the critical environment. Especially the composition of the refolding buffer is a key step in protein refolding that has been discussed thoroughly in literature <sup>[172-174]</sup>. Solvent conditions modulate various properties of proteins in aqueous solutions, including solubility, stability and the potential interaction with other molecules. In general, two main types of additives can be classified: folding enhancers (i.g. kosmotropes as protein stabilizers) and aggregation suppressors (chaotropes as aggregation inhibitors), with folding enhancer enhancing protein-protein interactions, and aggregation suppressors reducing side-chain interactions <sup>[75, 175]</sup>. Table 3-3 gives detailed information about the buffer compositions used for wt-matriptase in the rapid-dilution method and C122S-matriptase during the stepwise refolding. The main differences are a gradual introduction of L-arginine (Buffer 1 and 2) during stepwise dialysis in the case of C122S-matriptase refolding and the regulation of the redox-state to arrange proper conditions for correct DB formation.

**Table 3-3: Buffer compositions for wt-matriptase and C122S-matriptase refolding.**

Refolding buffer	Wt- matriptase	C122S-matriptase		
		Buffer 1	Buffer 2	Buffer 3
Urea	---	4 M	2 M	1 M
Tris-HCl	0.05 M	0.05 M	0.05 M	0.05 M
L-arginine	0.5 M	0.25 M	0.5 M	0.5 M
NaCl	0.1 M	0.1 M	0.1 M	0.1 M
CaCl <sub>2</sub>	0.02 M	0.02 M	0.02 M	0.02 M
EDTA	0.001 M	0.001 M	0.001 M	0.001 M
Cysteine-HCl	0.0005 M	---	---	---
GSH	---	---	0.0025 M	0.0015 M
GSSG	---	---	0.00025 M	0.00015 M
Tween-20	---	0.005 %	0.005 %	0.005 %
<b>Refolding method</b>	Rapid dilution	Stepwise dialysis		
<b>Final buffer</b>	0.05 M Tris-HCl 0.15 M NaCl	0.05 M Tris-HCl 0.1 M NaCl 0.005 % Tween-20		

L-arginine is customarily used as low molecular chaotrope that has shown to inhibit intermolecular hydrophobic interactions leading to subsequent precipitation <sup>[172, 176]</sup> and to increase protein solubility. Typically used L-arginine concentrations in refolding buffers are 0.5 M-2 M. For both matriptase variants, the minimum recommended concentration of 0.5 M L-arginine was used. In the case of C122S-matriptase L-arginine was adjusted to 0.5 M to stabilize the protein when denaturant concentration was lowered.

The redox-state of the refolding solution is regulated by the addition of folding enhancers and is a critical step during refolding for correct disulfide bond formation. The mechanism of disulfide bond formation in proteins involves a series of thiol/disulfide exchange reactions. The DB formation can only occur when the thiol groups (SH) of two cysteine residues are positioned in a correct steric conformation to enable the thiol-disulfide exchange <sup>[177]</sup>. Once formed, they restrict the number of possible folding conformations for a given protein, resulting in a decreased entropy and increased thermostability <sup>[178]</sup>. *In vivo*, the disulfide bond formation is mediated by specialized enzymes, including the protein disulfide isomerase (PDI) in the ER. During *in vitro* renaturation, the DB formation is standardly mediated using redox shuffling systems to not only drive the formation by creating an oxidizing environment but also to enable wrongly formed disulfide bonds to be reduced again with the goal to achieve the most stable form over the kinetically favored folding form. Commonly used redox shuffling systems include reduced and oxidized glutathione (GSH/GSSG) or cysteine/cystine which are combined in a molar ratio from 1:1 to 10:1 in concentrations between 1-15 mM <sup>[179]</sup>. In the case of matriptase recovery, the GSH/GSSG redox shuffling system as used for C122S-matriptase does

not seem to be beneficial over simply cysteine (the electron donor from the cysteine/cystine redox shuffling system) and air oxygen as used for the wt-matriptase refolding process. No clear advantage of stepwise introduction and removal of buffer components during stepwise dialysis for matriptase recovery could be confirmed under the given parameters. However, plenty of further stabilizing additives are available including kosmotropes, such as glycerol or polyethylene glycols (PEGs). Glycerol stabilizes proteins by enhancing the hydrophobic interactions caused by an increase in the solvent ordering around the proteins. An increasing glycerol concentration correlates with an increase in protein stability. PEGs have a stabilizing effect by binding to the intermediate form during protein folding or interact with hydrophobic side chains of the denatured protein <sup>[180, 181]</sup>. The identification of the optimal combination of refolding buffer and protein concentration is essential, but also very time-consuming as mostly achieved using trial and error approaches. To address this, high-throughput (HT) automated refolding screenings have been developed <sup>[182-184]</sup> allowing to screen multiple combinations of buffer and protein concentrations at the same time. However, success is often difficult to determine when quantified standardly by turbidity only. Both matriptase variants from this thesis require an additional activation step after successful refolding. Therefore, the HT-approach, as a quick screening method, is difficult and would be more suitable for matriptase variants containing the autocatalytic cleavage site where protein activity in addition to turbidity can be used for quantification to identify suitable buffer components. In terms of production costs, the limiting factor for the wt-type matriptase production following the given protocol of <sup>[159]</sup> is the activating enzyme TAGZyme DAPase (QIAGEN, Hilden). To process 50 mg of recombinant protein, the usage of 2.5 U TAGZyme DAPase Enzyme (costs approximately 500 € (2017)) is recommended. In the preparation run of wt-matriptase, around 50 mg protein remained after the refolding step and represented the base for the calculations of the activation-mix formulation. With 500 € per 3.8 mg properly folded and processed active matriptase solely for the activation, the production costs of wild-type matriptase are significantly higher compared to the production of C122S-matriptase. The decisive factor of C122S-matriptase production costs is L-arginine with an average price of approximately 100 € per kg. Even though C122S-matriptase production requires circa 2.5 times more L-arginine (depending on buffer volume) than wt-matriptase manufacturing, the cost factor does not overcome costs for DAPase. The activating enzyme for C122S-matriptase (5xHis-tagged recombinant enteropeptidase) was kindly provided by Viktor Magdolen (TU Munich) but can also be produced in-house since the respective plasmid (Addgene) and a detailed manual are commercially available <sup>[185]</sup>. This do-it-yourself option is missing in the case of the TAGZyme

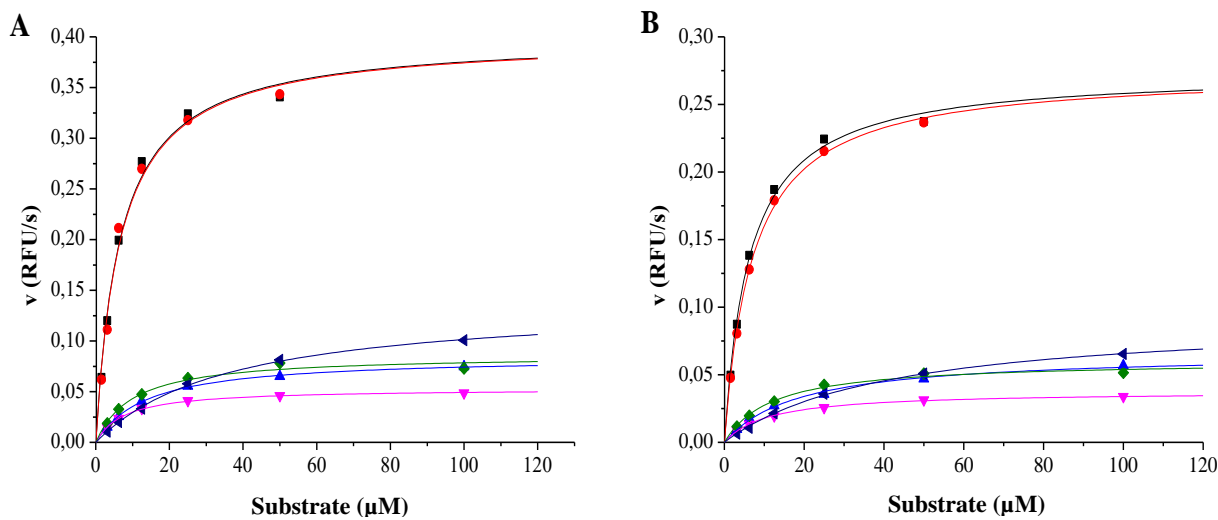
DAPase from QIAGEN. To sum up, the higher recovery rate of 4 % compared to the 2 % in case of C122S-matriptase does not compensate for immense activation costs of wt-matriptase. Ultimately, the functionality and quality of the prepared enzyme are vital. Therefore, both enzymes were characterized in enzyme kinetic measurements.

### 3.2.4. Functional comparison of matriptase variants using enzyme kinetic parameters

Both matriptase variants (wt and C122S-matriptase) were compared in enzyme kinetic measurements using six different fluorogenic AMC substrates (synthesized in our group) containing arginine in P1 position. At a constant enzyme concentration, the AMC release was determined at seven different substrate concentrations. In Figure 3-8, the steady-state velocities (Y-axis) were plotted versus the substrate concentrations (X-axis). The MICHAELIS-MENTEN equation was applied to determine the kinetic parameters  $K_M$  and  $V_{max}$  from the hyperbolic curves and the enzymatic efficiency is described by the ratio of  $V_{max}/K_M$ . The results are summarized in Table 3-4.

Due to un-evaluable results in active site titration experiments using the fluorogenic substrate 4-Methylumbelliferyl 4-guanidinobenzoate (MUGB), the determination of  $k_{cat}$  was not possible for matriptase samples. As described for several other trypsin-like serine proteases<sup>[186-188]</sup>, this method can generally be used to determine the active protein portion in a sample using a 4-Methylumbelliferone calibration line. Since the method was successfully used under the same conditions for thrombin as an in-house positive control, general equipment or method failure can be excluded. The term  $V_{max}/K_M$  is proportional to the term  $k_{cat}/K_M$  and describes the enzyme efficiency with the highest value representing the best substrate for an enzyme. Due to an unsuccessful active site titration, the concentrations of both matriptase preparations used for kinetic measurements were adjusted to a final concentration of 0.2 nM based on the protein content (UV measurement at 280 nm). The two matriptase variants (C122S-matriptase in Figure 3-8 A and wt-matriptase in Figure 3-8 B) show the same preferences for each tested substrate, respectively. For both, the lowest enzymatic efficiency was found for the substrate MI-1319 and the highest  $V_{max}/K_M$  values were identified for MI-507. Although enzymatic efficiency for MI-508 is at the same level, substrate MI-507 (Mes-DArg-Pro-Arg-AMC) was chosen for further measurements, since it had been also used for matriptase testing in recent studies in our lab<sup>[128]</sup>. These kinetic results prove on the one hand, that the single site-mutation C122S does not affect the catalytic property of wt-matriptase and on the other hand that the two refolding strategies result in active enzyme of equal quality. Only marginal differences in performance can be observed and might be based on batch-to-batch variations or be caused by the different ages of both preparations. Whereas the slightly less active wild-type matriptase batch was

produced in 2014 at the beginning of the work, the C122S-matriptase batch used for the enzyme kinetic measurement was produced more recently.



**Figure 3-8: Determination of the enzyme kinetic parameters  $K_M$  and  $V_{max}$  for both matriptase variants with AMC substrates.** Six different AMC substrates were measured at six concentrations using a constant amount of matriptase. The kinetic parameters  $K_M$  and  $V_{max}$  were determined using the MICHAELIS-MENTEN equation. Substrates MI-505(◄), MI-506 (▼), MI-1314 (◆) and MI-1319 (▲) were measured at 1.56  $\mu\text{M}$  - 50  $\mu\text{M}$ , whereas MI-507 (■), MI-508 (●) were measured at a concentration between 3.13  $\mu\text{M}$  - 100  $\mu\text{M}$ . A: C122S matriptase, B: wt-matriptase.

**Table 3-4: Enzyme kinetic parameters of the tested AMC-substrates.**

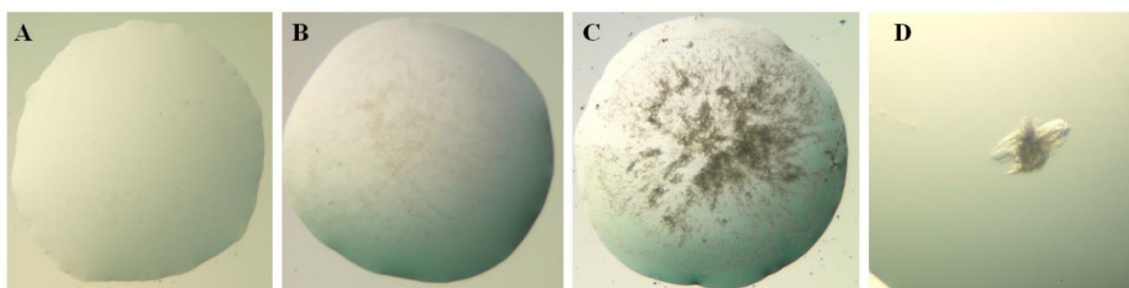
		$K_M$ ( $\mu\text{M}$ )		$V_{max}$ ( $\Delta\text{RFU/s}$ )		$V_{max}/K_M$ ( $\text{RFU}/\mu\text{M}\cdot\text{s}$ )	
MI	Sequence P4-P1'	C122S	wt	C122S	wt	C122S	wt
505	H-DArg-Pro-Arg-AMC	14.3 ( $\pm 3.275$ )	18.8 ( $\pm 2.253$ )	0.09 ( $\pm 0.010$ )	0.07 ( $\pm 0.001$ )	0.0062	0.0037
506	H-DArg-Gly-Arg-AMC	7.7 ( $\pm 0.844$ )	11.2 ( $\pm 0.101$ )	0.05 ( $\pm 0.007$ )	0.04 ( $\pm 0.000$ )	0.0065	0.0036
507	Mes-DArg-Pro-Arg-AMC	6.5 ( $\pm 1.184$ )	6.3 ( $\pm 0.001$ )	0.40 ( $\pm 0.086$ )	0.27 ( $\pm 0.003$ )	0.0620	0.0430
508	Mes-DArg-Gly-Arg-AMC	6.8 ( $\pm 1.048$ )	7.1 ( $\pm 0.383$ )	0.40 ( $\pm 0.065$ )	0.27 ( $\pm 0.001$ )	0.0588	0.0380
1314	Mes-DArg-Val-Arg-AMC	10.0 ( $\pm 1.178$ )	12.0 ( $\pm 0.739$ )	0.09 ( $\pm 0.014$ )	0.06 ( $\pm 0.000$ )	0.0090	0.0050
1319	H-DhPhe-Pro-Arg-AMC	36.8 ( $\pm 6.955$ )	41.7 ( $\pm 4.855$ )	0.14 ( $\pm 0.023$ )	0.09 ( $\pm 0.003$ )	0.0038	0.0022

**3.2.5. Crystallization experiments using C122S-matriptase**

So far, binding modes of in-house matriptase inhibitors could only be investigated in complex with the serine proteases thrombin and trypsin since the availability of sufficient quantities of matriptase was limited. In order to establish matriptase crystallization in our facilities, C122S-

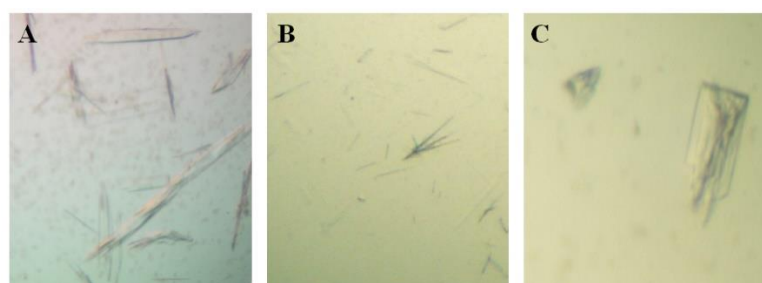


matriptase was produced and concentrated to 5 – 10 mg/ml, incubated with either benzamidine or an inhibitor of the 3-amidinophenylalanine type (compound **17** in Table 4-1) and used in hanging drop crystallization experiments in combination with various crystallization buffers. Since properly purified monodispersed C122S-matriptase is still not superabundant for large screenings, first attempts included testing buffer conditions plus slight variations of already published matriptase crystallization conditions <sup>[159, 162, 189-192]</sup>. These screenings resulted mostly in drops being empty (Figure 3-9 A) or containing light (possibly crystalline, Figure 3-9 B) or dark precipitate (Figure 3-9 C).



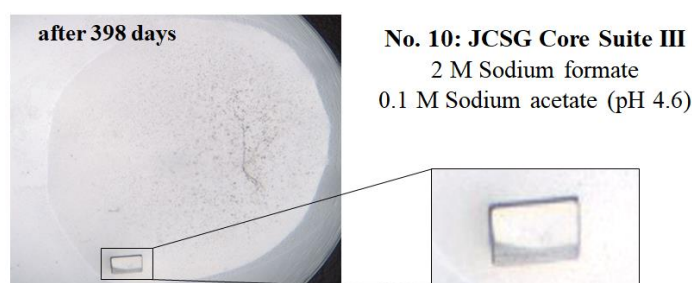
**Figure 3-9: C122S-matriptase crystallization set up resulting in empty drop (A) or drop containing light precipitate (B), dark precipitate (C) or an overgrown protein crystal (D).** C122S-matriptase at 5 mg/ml was subjected to crystallization experiments using the hanging drop vapor technique. Drops were monitored over time (up to 1 year) using a light microscope. Figures were taken 8 weeks after drop set up and incubation at 18 °C. Crystallization conditions for 5 mg/ml matriptase/10 mM benzamidine complex were described earlier including conditions used for **A** and **B** <sup>[159]</sup> as well as **C** <sup>[189]</sup>. For the crystallization experiment in **D** C122S-matriptase was incubated with 1 mM of the 3-amidinophenylalanine-derived inhibitor **17**. Crystallization buffer composed of 100 mM Tris-HCl (pH 8.0), 200 mM MgCl<sub>2</sub> and 30 % (w/w) PEG 5000.

Since polyethylene glycol (PEG) was used as precipitant in several earlier protocols for matriptase crystallization <sup>[159, 190-192]</sup> a self-made PEG screen (screening for PEG 3000 (25/30/35 %), 5000 (25/30/35 %), 6000 (20/25/30 %) and 8000 (16/20/24 %) in combination with 200 mM MgCl<sub>2</sub> or 200 mM Li<sub>2</sub>SO<sub>4</sub>) was used. Crystallization of the C122S-matriptase/inhibitor complex resulted in one strongly overgrown protein crystal (Figure 3-9 D). As crystal growth was not repeatable at the very same and varying (pH and precipitant concentration) conditions, the overgrown crystal was used for subsequent seeding experiments. 96 different crystallization buffers from two commercially available screens “The PEGs Suite” and “The PEGs II Suite” (QIAGEN) were chosen. Buffers with a pH of 6.5 (0.1 M MES), 7.5 (0.1 M HEPES) and 8.5 (0.1 M Tris-HCl) containing precipitant only or precipitant plus a salt were selected. Crystallization drops composed of 4 mg/ml C122S-matriptase/1 mM inhibitor, crystallization buffer and seeding stock (undiluted) in a ratio 3:2:1 were prepared and monitored. Several conditions lead to drops containing crystalline materials in form of needles or small splitter like pieces ( Figure 3-10 A-C) but no conditions could be identified leading to single well-shaped protein crystals suitable for X-ray analysis.



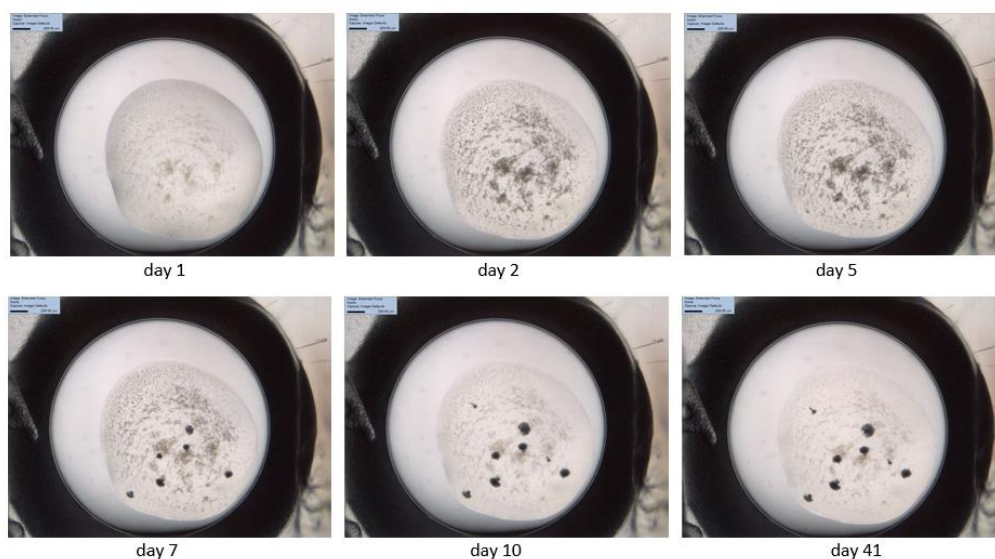
**Figure 3-10: Seeding experiments with C122S-matriptase resulting in needles and splitter like crystalline material.** Seeding was performed as described in the literature <sup>[193]</sup> using the hanging drop vapor technique. Drops were monitored using a light microscope. Figures were taken 4 weeks after drop set up and incubation at 18 °C. **A:** 0.1 M MES (pH 6.5), 0.2 M MgCl<sub>2</sub>, 25 % (w/v) PEG 4000 **B:** 0.1 M Tris-HCl (pH 8.5), 0.8 M LiCl, 32 % (w/v) PEG 4000; **C:** 0.1 M MES (pH 6.5), 25 % (w/v) PEG 550 MME.

Although these results were more promising, C122S-matriptase is still too limitedly available for larger in-house optimizations. Therefore, C122S-matriptase was also subjected to a high throughput nanodrop screening at the crystallization laboratory MarXtal (Marburg, Germany), where small amounts of protein are sufficient to screen against multiple different buffer compositions in a nanoscale. Using a crystallization robot, three commercially available kits containing 96 conditions each were screened for optimal crystallization conditions of C122S-matriptase in combination with benzamidine or inhibitor. Each condition was photographed by the Rock Image system over time in order to monitor crystal growth. After one year of incubation at 18 °C, a single box-shaped C122S-matriptase/benzamidine crystal appeared that was suitable for X-ray analysis (Figure 3-11).

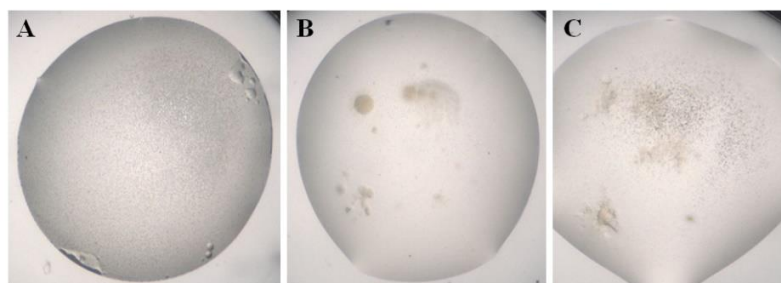


**Figure 3-11: Single box-shape C122S-matriptase crystal grown after 1-year incubation at 18 °C.** 7 mg/ml C122S-matriptase/10 mM benzamidine was mixed in a ratio of 1:1 with the respective crystallization buffer. After incubation for 1 year a box shape single well-defined protein crystal appeared.

After a much shorter incubation period of 7 days, walnut-like C122S-matriptase/benzamidine crystals appeared in several tested conditions (Figure 3-12). This crystal form is suitable for optimizations and was subjected to seeding experiments (described below). In addition, few optimization worthy C122S-matriptase/inhibitor crystallization conditions could be identified (Figure 3-13) but were not further followed.



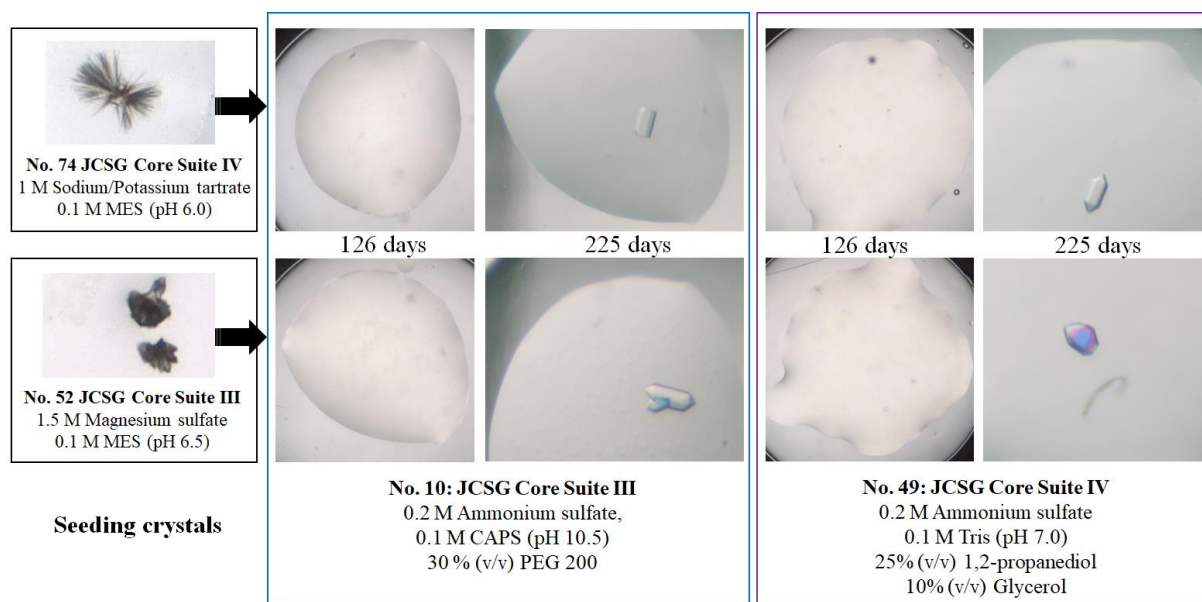
**Figure 3-12: Walnut like C122S-matriptase/benzamidine crystals grow after 5-7 days in high throughput nanodrop crystallization experiments.** 7 mg/ml C122S-matriptase/10 mM benzamidine was mixed 1:1 with 100 mM CAPS (pH 10.5), 200 mM  $\text{Li}_2\text{SO}_4$ , 1.2 M  $\text{NaH}_2\text{PO}_4$  and 0.8 M  $\text{K}_2\text{HPO}_4$  (No. 2 of JCSG Core IV screen (QIAGEN)) and incubated in a sitting-drop experimental set up over 41 days. Crystals of a walnut shape appear after 7 days. After 10 days the crystals stopped growing and no further changes were visible for the next 31 days of incubation.



**Figure 3-13: Crystallization conditions for C122S-matriptase/inhibitor complex that are suitable for optimization.** 5 mg/ml C122S-matriptase/1 mM inhibitor was mixed 1:1 with respective crystallization buffers. **A:** Microcrystalline material visible using the crystallization buffer JCSG Core IV buffer No. 53 (3.6 M Sodium formate, 10 % Glycerol), **B:** Quasi-crystals in condition 33 of JCSG Core IV screen (0.1 M Tris-HCl (pH 8.0), 1.6 M Ammonium sulfate), **C:** Mix of microcrystals and quasi crystals in condition 45 of JCSG Core IV screen (1.26 M Tri-sodium citrate, 0.09 M HEPES (pH 7.5) and 10 % Glycerol).

To further improve matriptase crystallization, crystalline material from the first high throughput nanodrop screen was used for seeding experiments following the guide “Microseed Matrix Screening (MMS) for optimization in protein crystallization” that was published in 2014 <sup>[193]</sup>. Introduced crystalline particles can act as a nucleation site for crystal growth in the new experiment. Since optimal conditions for crystal nucleation are hypothesized to often differ from the optimal conditions for crystal growth, the hit rate is thought to be increased with this method. Furthermore, *de novo* nucleation is one of the most (if not the most) critical point in crystallization. In typical seeding experiments the same conditions are used as the seeding crystals grew in with simply lowering supersaturation (e.g. by a lower protein concentration)

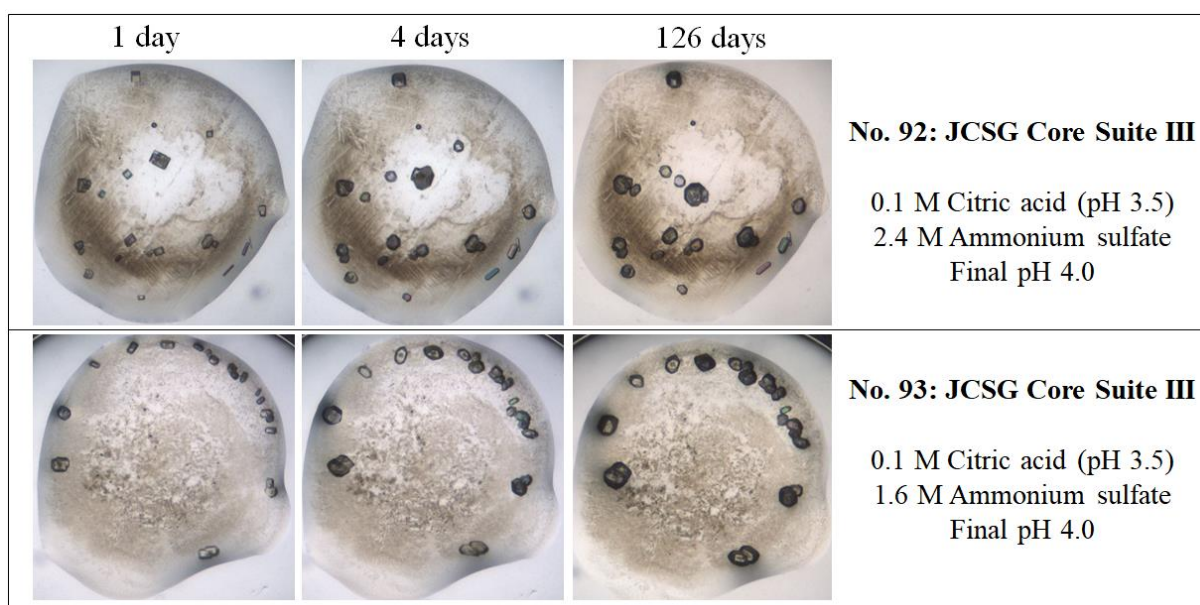
[194]. The MMS approach from D'ARCY is based on the introduction of crystalline material from an initial experiment into a variety of new and often unrelated crystallization conditions, as initially published in 2004 [195], using a robotic liquid-handling device to seed into sparse-matrix screens [196]. Two different seed stocks were prepared using crystalline C122S/benzamidine material that was not of sufficient quality for structure determination by X-ray diffraction. From one crystal out of 288 tested conditions, crystal quantity was thereby improved to 4 crystals out of 192 screened conditions by MMS. However, these C122S-matriptase crystals appeared again only after a long incubation period of several months.



**Figure 3-14: Improved C122S-matriptase/benzamidine crystal morphology by MMS.** 6.4 mg/ml C122S-matriptase /10 mM benzamidine solution was mixed 1:1 with a mixture of buffers No. 10 (JCSG Core Suite III screen) or No. 49 (JCSG Core Suite IV screen) and the respective seeding stock in a ratio of 2:1. Well plates were incubated in a sitting-drop experimental set up at 18 °C. Single protein crystals appeared between 126 and 225 days of incubation.

HT nanodrop screening is used to find suitable starting conditions for further optimizations in a larger drop scale, which is required since conditions are often not simply 1:1 transferrable in scale-up experiments. However, larger drops are favorable since the crystal quantity per drop, as well as the crystal size increases with the drop-size. Even though MMS led to nicely shaped single matriptase crystals, the long incubation time is not optimal for optimization experiments. In contrast to this very long incubation time, in two crystallization conditions appeared crystalline material in the MMS already after only 1 day (Figure 3-15). If these crystals are simply salt crystals or represent protein crystals has to be clarified. However, in case these drops contain protein crystals these conditions are very promising.





**Figure 3-15: Improved matriptase/benzamidine crystal growth by MMS.** 6.4 mg/ml C122S-matriptase/10 mM benzamidine was mixed 1:1 with a mixture of buffers No. 92 and 93 (JCSG Core Suite III screen) and a seeding stock in a ratio 2:1. Well plates were incubated in a sitting-drop experimental set up at 18 °C. Multiple differently shaped crystals appeared after 1 day of incubation and grew for the next 4 days.

### 3.3. Summary of matriptase preparation and characterization

In this chapter, the preparation of two active recombinant matriptase SPD variants from *E. coli* inclusion bodies using two different refolding approaches had been described. Wildtype matriptase was recovered successfully following an already established protocol including a rapid dilution refolding approach. A single site mutated C122S-matriptase variant, containing an EK cleavage site to allow directed activation, has been created and its preparation in *E. coli* could be established. For the refolding of C122S-matriptase a stepwise dialysis strategy had been applied. Both refolding approaches result in high quality and pure enzyme with equal catalytic performance after subsequent chromatographic steps. In terms of production efficiency, the rapid dilution refolding strategy applied for wt-matriptase resulted in a higher percental recovery yield. However, the total recovery yield was, with 2 % for C122S-matriptase and 4 % in the case of the wild-type variant, generally low and relatively inefficient for both enzymes. Besides the refolding method itself, the refolding buffer composition is critical for the success of protein recovery and various factors influence the recovery success of denaturized protein. Temperature, pH and different additives are known to inhibit protein aggregation or enhance protein folding and were discussed as explorable variables to further optimize the recovery yield of recombinant matriptase. Lower recovery yields aside, the C122S-

matriptase production strategy is more favorable over the wt-matriptase production in terms of production costs.

In the course of this thesis, adequate amounts of pure high-quality C122S-matriptase could be recovered and subjected to further experiments, one of which was protein crystallization. After initial unsuccessful in-house approaches based on already published conditions and intensive manual screening for suitable conditions, C122S-matriptase was subjected to high throughput nanodrop crystallization experiments using a robot device and commercially available buffer screens. After a long incubation time of one year a suitable crystal was available for structural analysis. By applying an MMS approach using a seeding stock (prepared from crystalline material from the first HP-nanodrop experiment that was not of sufficient quality for structure determination by X-ray diffraction) four more excellent looking matriptase crystals could be found. However, also these crystals again grew only after a long incubation time. In combination with seeding, two buffer conditions were identified which showed promising results for further optimization since crystalline material appeared already after 1 day. Nevertheless, both conditions are very similar and the crystals are of a different shape than others, which might indicate that these are salt, but not protein crystals. Even though matriptase crystallization requires optimization, in total 5 matriptase crystals were found during this thesis and can be used to determine the crystallographic binding mode of in-house inhibitors.

## **4. Computer assisted matriptase inhibitor design and their characterization**

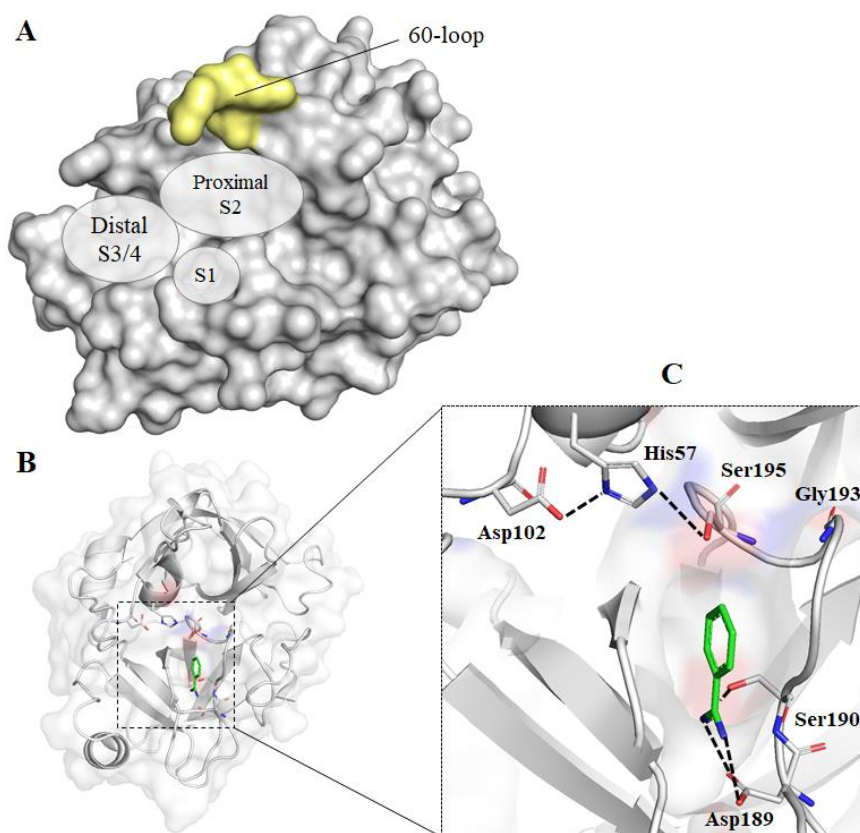
### **4.1. Introduction to matriptase inhibitor design**

A comprehensive insight into matriptase, including its multidomain architecture, physiology and its endogenous regulation as well as patho-physiology has been given in the previous chapter 3. The strong involvement in cancer but also in the infectious disease influenza shapes matriptase into an interesting therapeutic target. In general, precise structural information such as amino acid composition, surface charge, and architecture of the binding pockets at sufficient resolution are essential for the rational structure-based design of efficient inhibitors. The first crystal structures of matriptase SPD in complex with benzamidine or the bovine pancreatic trypsin inhibitor aprotinin were published in 2002 <sup>[162]</sup>. Since then, several crystal structures of the matriptase SPD in complex with synthetic inhibitors <sup>[159, 189, 191, 197]</sup>, fragment antigen-binding (Fab) derived inhibitors E2 and S4 <sup>[190, 198]</sup> or with its endogenous inhibitor HAI-I <sup>[163]</sup> have been solved and contributed to a thorough structural characterization of matriptase. However, as being able to recognize and cleave numerous strongly basic and therefore, relatively polar substrates, matriptase represents a challenging target for potential drug development. Efficient and selective inhibitors, that additionally exhibit appropriate pharmaceutical properties such as bioavailability and prodrug potential, are in demand. Conventional as well as computational approaches led to a large selection of very potent hits that lend themselves as candidates for further hit to lead optimization. This chapter concentrates on the identification of synthetic monobasic matriptase inhibitors of the 3-amidinophenylalanine type in order to obtain a fair selectivity profile against the serine proteases thrombin and fXa. In principle, such amidino compounds can be easily converted into nonbasic prodrugs, like hydroxyamidines, as known from the development of the orally available thrombin inhibitor ximelagatran <sup>[199]</sup> or the combined orally active uPA and matriptase inhibitor mesupron <sup>[200]</sup>. A computer assisted approach was applied to reasonably choose promising starting molecules that come with the best compromise of effort and costs.

#### **4.1.1. Structure of matriptase SPD**

Based on the similarities to other SPDs of trypsin-like serine proteases, the commonly used chymotrypsinogen amino acid numbering is used to describe and analyze the structure of matriptase. Matriptase shares typical characteristic features including the catalytic triad with, His57, Asp102 and Ser195, the oxyanion hole created by the main chain NH of Gly193 and

Ser195 and the S1 specificity pocket with the characteristic Asp189 at the bottom (Figure 4-1 B and C).



**Figure 4-1: Surface binding pocket representation (A) and secondary structure (B) of matriptase with a close-up of the S1 binding site including the catalytic triad (C).** Protein surface (A) and secondary structure in cartoon (B+C) are colored in grey with slightly visible surface. Important residues are given in sticks. Pictures were made with PyMol using PDB code 1EAX. The S1 binder benzamidine is shown in green. Dashed black lines in C indicate interactions between residues of the catalytic triad or between benzamidine and matriptase.

The matriptase S1 pocket is bordered by segments Val213-Cys220, Ser190-Ser195, Pro225-Tyr228 and the disulfide bridge between Cys191 and Cys220. Variations within this region between trypsin-like serine proteases were mainly found at position 217, 192 and 190, with an Asp217, Gln192 and Ser190 for matriptase. The hydrophobic S2 site is well suited for accommodating small to medium-sized P2-amino acids but has shown some flexibility to also accept larger amino acids. Although the small alanine was found to be a preferred P2 residue [128, 201], the potential space restricting Phe99 of the 90-loop possesses a rotatable side chain that allows matriptase to also efficiently cleave substrates with larger P2 amino acids such as Phe or Leu [103]. A unique characteristic of matriptase is the nine-residue 60-insertion loop on top of the S2 pocket (yellow in Figure 4-1 A). The averted orientation of this loop results in an accessible active site cleft and its conformational flexibility helps matriptase to adapt to different ligands by forming a number of additional favorable interactions [162, 163]. Another



characteristic of matriptase is the S3/4-binding site that is formed by the four amino acids Phe97 on the top, Phe99 on the right side, Trp215 at the bottom, and Gln175 on the left (as shown below in Figure 4-5), which can accommodate both, hydrophobic and cationic residues. Presumably cation/ $\pi$ -interactions,  $\pi/\pi$ -interactions and an overall negative potential formed by the carbonyl groups of amino acids 96-98 and the side chain carbonyl of Gln175 contribute to the preferred binding of basic amino acids to this binding site. In the case of non-substrate like analogues, these adjacent binding sites are also referred to as proximal (for S2) and distal (for S3/4) binding pockets. In the following S1, S2 and S3/4, as well as proximal and distal binding sites are used for matriptase areas as shown in Figure 4-1 A.

#### 4.1.2. Inhibitors of matriptase

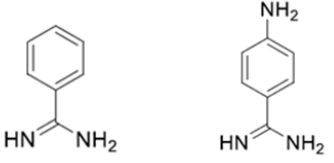
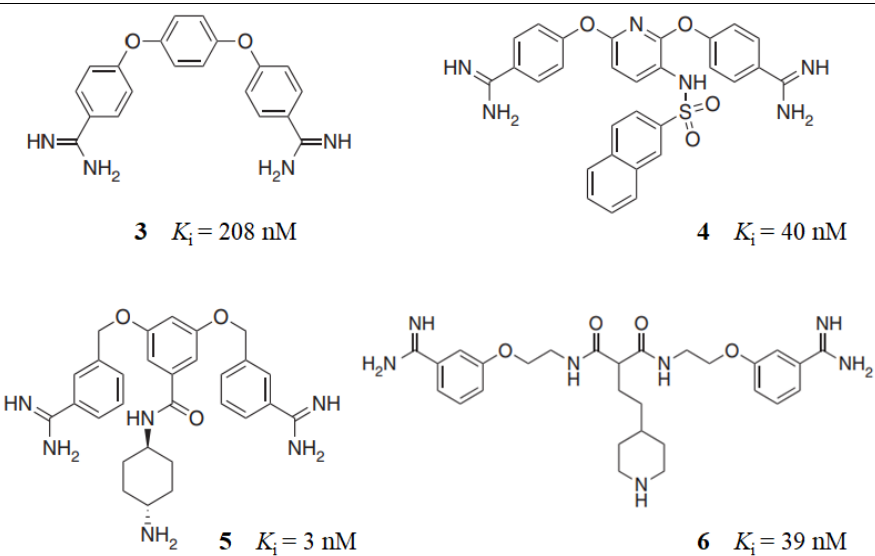
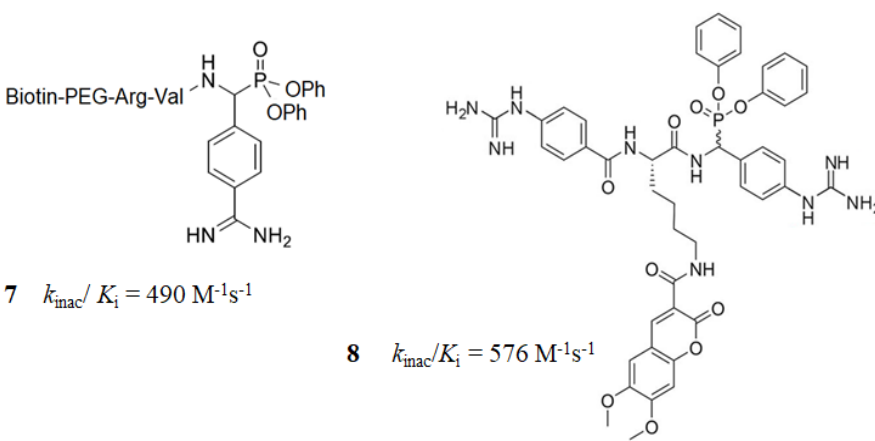
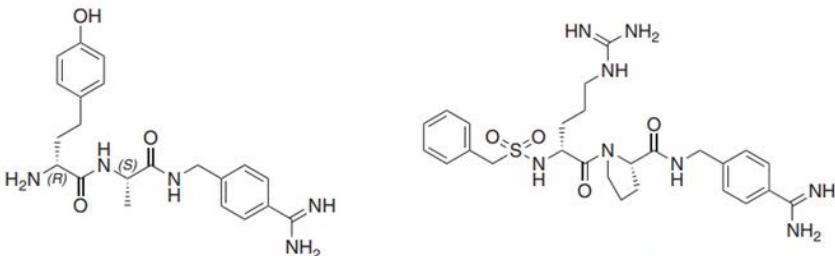
##### *Natural inhibitors*

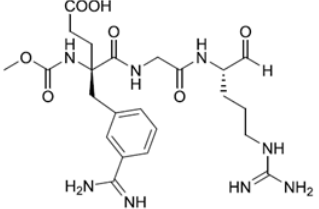
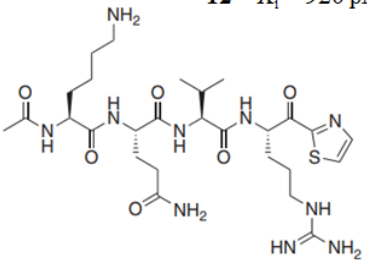
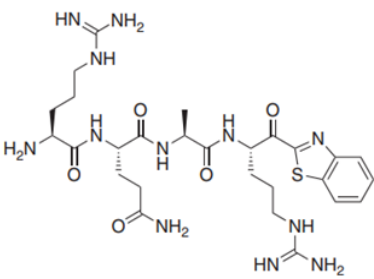
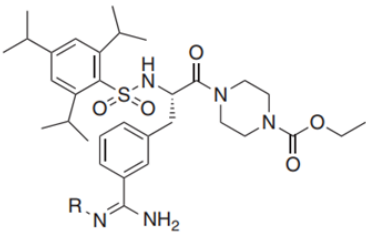
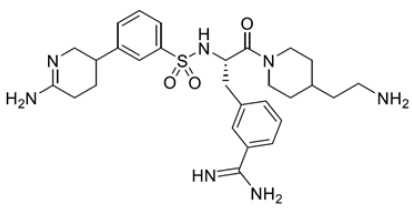
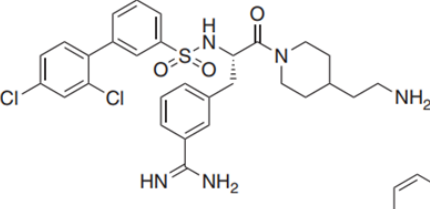
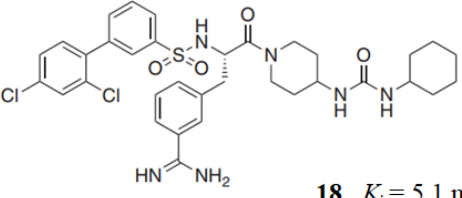
A wide variety of natural and synthetic matriptase inhibitors has been described in the literature. Besides the most discussed cognate and endogenous matriptase inhibitors HAI-1 and HAI-2 [129, 163, 202], other natural inhibitors include ecotin [160], SFTI-1 [203] and its derivatives [204-207] as well as mABs [190, 208]. Highly efficient molecules were observed with  $K_i$  values of 9.8 pM for a double mutated (M84R/M85R) ecotin variant [160] or 920 pM for SFTI-1 [203]. Starting with affinities of 50 pM, mABs, more precisely chimeric proteins of two linked single-chain variable region fragments (scFv) comprising of heavy and light antibody chains, have also shown a high affinity to matriptase [208]. Due to their high selectivity and low toxicology, antibodies are well suited as therapeutic reagents but are large of size and have a restricted oral bioavailability. More recently, very potent HAI-1 derivatives were published. A soluble inhibitor with a  $K_i$  value of 70 pM has been identified, where the inactive Kunitz domain (KD) 2 of HAI-1 was replaced by a chimeric variant of KD2/KD1 and additionally fused to the crystallizable fragment (Fc) domain of an antibody. However, in contrast to antibodies, the resulting KD1-KD2/1-Fc HAI-1 derivative lacks a strong selectivity profile as it also shows low nanomolar affinities to kallikrein 4 ( $K_i = 8.0$  nM; selectivity: 110-fold) and hepsin ( $K_i = 1.5$  nM; selectivity: 20-fold) [129].

##### *Synthetic small molecule matriptase inhibitors*

Different types of synthetic inhibitors have been developed, including non-peptidic molecules as well as substrate and non-substrate like peptidic inhibitors. Examples for different inhibitor types including their inhibition constant for matriptase are summarized in Table 4-1 but are not all discussed in detail.

**Table 4-1: Structures and inhibitory potency of different matriptase inhibitor types.**

Inhibitor type	Structure
Non-peptidic inhibitors	
Benzamidine, p-aminobenzamidine (S1 binders)	 <p><b>1</b> <math>K_i = 390 \mu\text{M}</math>      <b>2</b> <math>K_i = 90 \mu\text{M}</math></p>
Bis-benzamidine derivatives	 <p><b>3</b> <math>K_i = 208 \text{ nM}</math>      <b>4</b> <math>K_i = 40 \text{ nM}</math></p> <p><b>5</b> <math>K_i = 3 \text{ nM}</math>      <b>6</b> <math>K_i = 39 \text{ nM}</math></p>
Peptidic inhibitors	
Peptidic phosphonate- derivatives (irreversible)	 <p><b>7</b> <math>k_{\text{inac}}/K_i = 490 \text{ M}^{-1}\text{s}^{-1}</math>      <b>8</b> <math>k_{\text{inac}}/K_i = 576 \text{ M}^{-1}\text{s}^{-1}</math></p>
Substrate analogue inhibitors	 <p><b>9</b> <math>K_i = 26 \text{ nM}</math>      <b>10</b> <math>K_i = 55 \text{ nM}</math></p>

Inhibitor type	Structure
Covalent substrate analogue inhibitors	<p><b>11</b> <math>K_i = 3.3 \text{ nM}</math></p>  <p><b>12</b> <math>K_i = 920 \text{ pM}</math></p>  <p><b>13</b> <math>K_i = 11 \text{ pM}</math></p> 
	<p><b>14</b> R = H (WX-UK1) <math>K_i = 370 \text{ nM}</math>  <b>15</b> R = OH (Mesupron)</p>  <p><b>16</b> <math>K_i = 0.08 \text{ nM}</math></p>  <p><b>17</b> <math>K_i = 2 \text{ nM}</math></p>  <p><b>18</b> <math>K_i = 5.1 \text{ nM}</math></p> 
Tertiary amides of sulfonylated 3-amidinophenyl-alanine	

**3** [209]; **4** [191]; **5** [210]; **6** [211]; **7** [189]; **8** [166]; **9** [128]; **10** [212]; **11** [213]; **12** [214]; **13** [201]; **14** [215]; **15** [216]; **16** [217]; **17** and **18** [218]

Simply benzamidine (compound **1**) inhibits matriptase with a  $K_i$  value of  $390 \mu\text{M}$ . Introducing an amino function in *para*-position (compound **2**) results in a four times more potent compound ( $K_i = 90 \mu\text{M}$ ). Both compounds solely occupy the S1 binding site. Bis-benzamidines can further improve  $K_i$  values to the nanomolar range ( $208 \text{ nM}$ , compound **3**). Since benzamidine perfectly fits into the S1 pocket of trypsin-like serine proteases, it is often used as an arginine mimetic in inhibitor design. In case of matriptase, it is sandwiched between the parallel peptide segments Trp215-Cys220 and Ser190-Gln192 while the amidino group forms two salt bridges with the

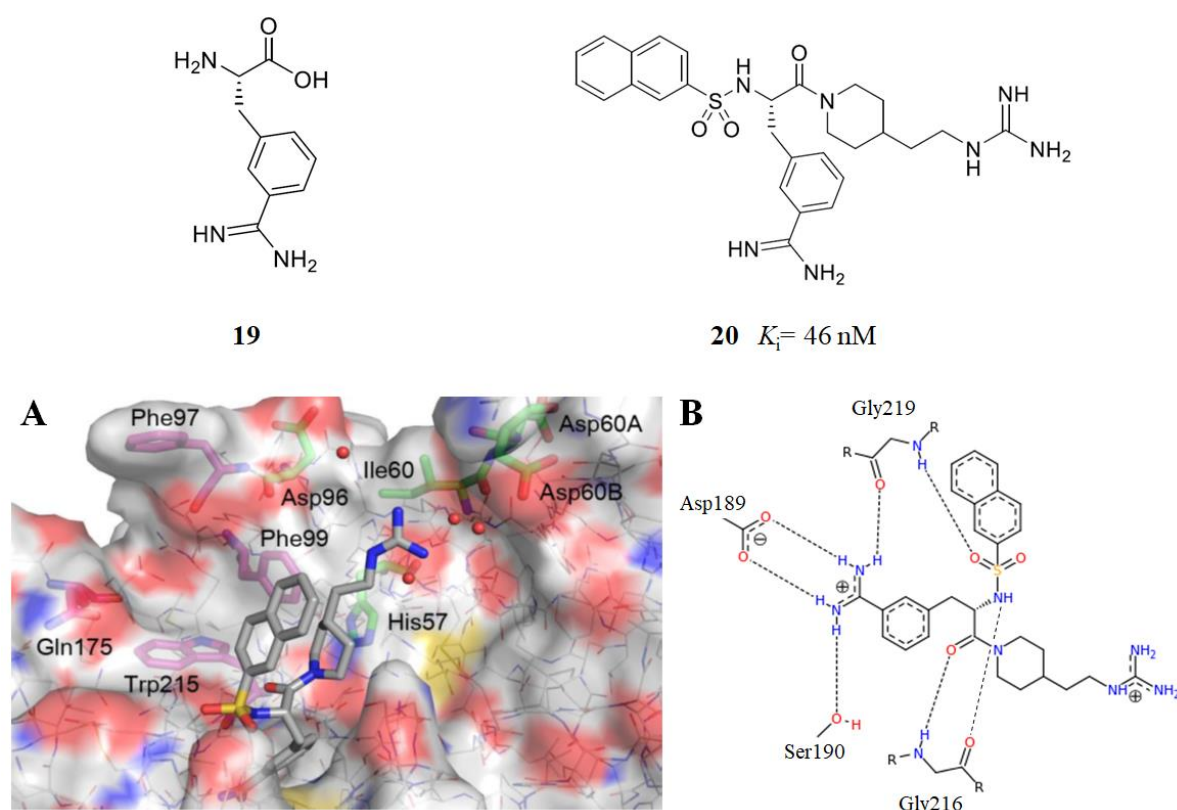
opposing Asp189 carboxylate and the side-chain OH of Ser190 (Figure 4-1 C) and Gly219 (not visible in Figure 4-1 C) side chain on both sides at the bottom of the S1 pocket. However, due to the similar architecture of the S1 pockets in most trypsin-like serine proteases, the design of selective inhibitors is quite challenging. Matriptase, in contrast to trypsin, does not indiscriminately cleave after lysine or arginine residues but requires recognition of additional residues surrounding the scissile P1-P1' peptide bond especially in the N-terminal direction for efficient cleavage. Potency, but also selectivity can be achieved for compounds that also address adjacent binding sites, including the S2 and S3/4 binding pocket above Trp215. Very potent inhibitors with a decent selectivity profile against other serine proteases were already developed and are discussed later on.

The most potent compound for matriptase so far is an irreversible covalent binding substrate analogue inhibitor with a  $K_i$  value of 11 pM (compound **13** in Table 4-1) possessing an arginyl-ketobenzothiazole group at the C-terminus<sup>[201]</sup>. A more detailed overview of various matriptase inhibitor types is given in a recently published review<sup>[219]</sup>.

#### 4.1.3. Synthetic inhibitors of the 3-amidinophenylalanine type

The amino acid 3-amidinophenylalanine (**19**, Figure 4-2) is a well-suited starting scaffold for the design of trypsin-like serine protease inhibitors. The benzamidine-function fits perfectly into the S1 pocket and the scaffold offers two different vectors for further optimization: the amino and carboxyl groups. A noteworthy member of this inhibitor type is Mesupron, an orally bioavailable hydroxyamidino prodrug of the 3-amidinophenylalanine-derived serine protease inhibitor of the second generation (compound **15** in Table 4-1, in development by RedHill Biopharma (Tel-Aviv)). Mesupron represents the first synthetic inhibitor of trypsin-like serine proteases targeting pancreatic cancer and inflammatory gastrointestinal diseases, which reached clinical phase II development. After oral administration, Mesupron is converted to the active inhibitor WX-UK1 (compound **14**), initially designed as an uPA inhibitor ( $K_i = 410$  nM). However, WX-UK1 also inhibits several other serine proteases in the nanomolar range including matriptase and thrombin ( $K_i$  value of 370 nM and 490 nM, respectively)<sup>[220]</sup>. By introducing a C-terminal basic group and replacement of the N-terminal sulfonyl group, the potency could be improved for matriptase to a  $K_i$  value of 46 nM in case of compound **20** (CJ-730 in Figure 4-2 from Curacyte Chemistry GmbH, Jena), with a poor inhibitory potency against uPA ( $K_i = 1.3$   $\mu$ M) at the same time<sup>[159]</sup>. The crystal structure of this compound in complex with matriptase (PDB code: 2GV6) revealed a compact Y-shaped inhibitor conformation (Figure 4-2 A) that is consistent with similar compounds in complex with thrombin or trypsin<sup>[221, 222]</sup>. The benzamidine moiety is typically sandwiched between

fragments of the S1 site and forms the earlier described interactions with Asp189, Ser190, and Gly219 at the bottom of the S1 pocket. Moreover, the backbone NH and carbonyl of the P1 residue form an antiparallel  $\beta$ -sheet interaction with the matriptase residue Gly216 (Figure 4-2 B). These interactions are typical for inhibitors of the 3-amidinophenylalanine type. The frequently used C-terminal piperidide ring extends to the proximal S2 binding site. The terminal guanidinoalkyl group is surrounded by several water molecules and does not form direct contacts with matriptase (Figure 4-2 A). Since the distal S3/4 binding pocket remains unoccupied by the N-terminal naphthylsulfonyl group of inhibitor **20** (CJ-730), further optimizations included its modification with aromatic and basic residues. The most potent matriptase inhibitor of the 3-amidinophenylalanine type to date is the tribasic inhibitor **16** (Table 4-1) with a  $K_i$  value of 80 pM <sup>[217]</sup>. The N-terminal basic moiety probably targets the distal S3/4 pocket and is most likely involved in cation- $\pi$  interactions to Trp215 and Phe99 of matriptase.



**Figure 4-2: Structures of 3-amidinophenylalanine (19) and the 3-amidinophenylalanine-derived inhibitor CJ-730 (20) including its binding mode in matriptase.** A was taken from literature <sup>[223]</sup>, B is showing polar contacts between the inhibitor and matriptase. Figure was taken and modified from PDB code: 2GV6.

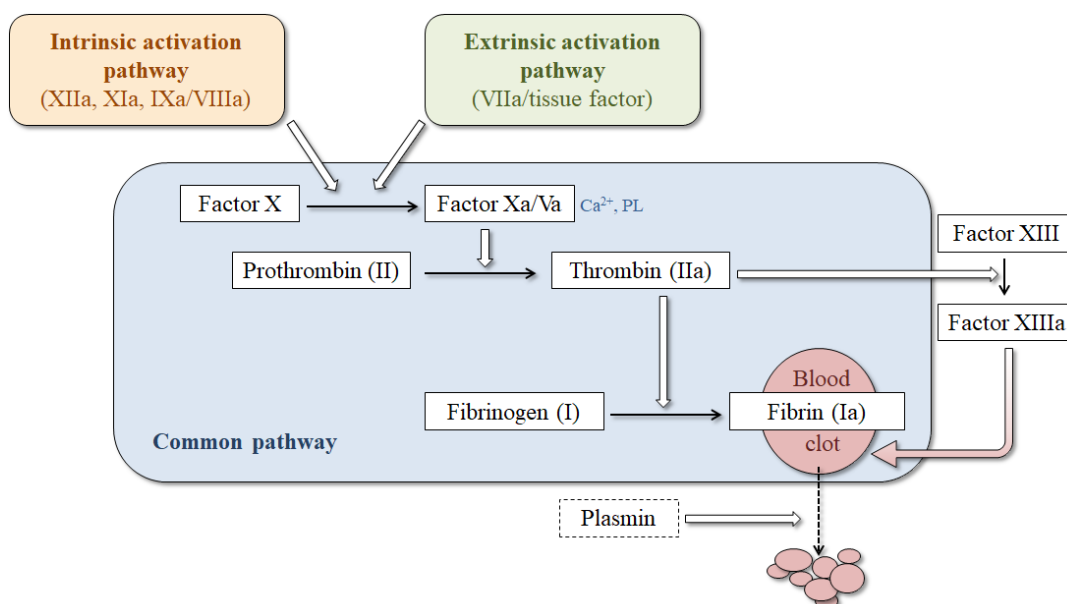
Between 2006 and 2009 numerous matriptase inhibitors based on sulfonlated 3-amidinophenylalanine with modified N- and C-termini were developed at Curacyte Chemistry GmbH. Some candidates showed improved potencies and decent selectivity towards other

serine proteases, including thrombin and factor Xa [159, 217, 224]. However, the multibasic and strongly polar character of most compounds limited their bioavailability. Therefore, in subsequent works it was focused on the exploration of potent derivatives with lower basicity [218]. Optimization of the substitution pattern on the N-terminal sulfonylated biphenyl moiety resulted in the potent, but still dibasic inhibitor **17** (Table 4-1, internal ID: MI-432) with a  $K_i$  of 2 nM. The monobasic analogue bearing a cyclohexyl urea as a replacement of the C-terminal amino group (compound **18**) retained high matriptase affinity ( $K_i$  value of 5.1 nM). Both, the dibasic compound **17** and the monobasic analogue **18** were tested in cell culture experiments and did not show any cytotoxic effects but effectively reduced the propagation of various LPAIV strains in infected cells at a concentration of 50  $\mu$ M [63].

Due to insufficient matriptase amounts available, the binding mode of inhibitor **17** was only determined by crystallization in complex with thrombin (PDB: 4E7R). As expected, the biphenyl moiety directs towards the distal S3/4 binding site. It interacts with side chains of amino acid Trp215 and Asn98 (*para*-chlorine). The *ortho*-chlorine points towards Glu217 but has a distance of 3.7 Å suggesting only a very weak possible interaction. The modeling of inhibitor **17** into matriptase suggested similar polar contacts of the P1 residue and a possible additional interaction of the *para*-chlorine to Phe99. Notably, all dibasic compounds with a C-terminal 4-(2-aminoethyl) piperidine possessed an approximately three to six-fold higher matriptase potency than their monobasic analogues containing a nonpolar C-terminal residue [218]. However, in contrast to monobasic amidino derivatives, which in principle can be easily converted into a single noncharged prodrug as described for Mesupron (**15**), this dibasic analogue would require a second prodrug approach to convert also the terminal amino group into an uncharged residue. Meanwhile, several strategies for the design of amine prodrugs are known [225] but have not been used for matriptase inhibitors so far.

#### 4.1.4. Selectivity of matriptase tailored inhibitors towards thrombin and fXa

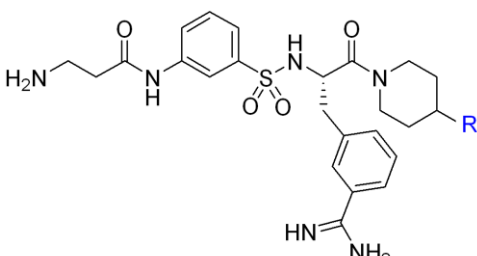


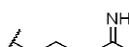
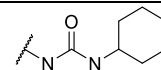
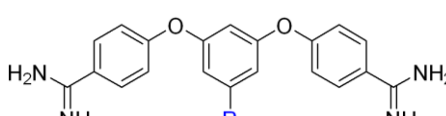
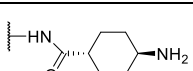
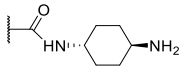
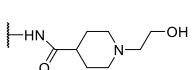
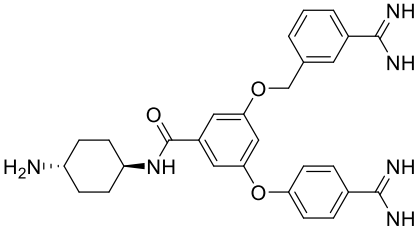
Due to high similarities between the active-site structures of various trypsin-like serine proteases, it is challenging to find inhibitors with a high potency towards only one target. Many matriptase inhibitors and also the uPA phase-II prodrug Mesupron (including the activated form WX-UK1) show similar affinities towards thrombin and fXa. These two soluble trypsin-like serine proteases play important roles in the blood coagulation cascade, where the fXa containing prothrombinase complex generates thrombin, which in turn activates fibrinogen. Fibrin is the insoluble clotting protein that finally closes wounds (Figure 4-3). Inhibiting one of these proteases would decelerate or prevent wound closure and therefore might lead to excessive bleeding with severe consequences even after small injuries.



**Figure 4-3: The role of thrombin and fXa in the blood coagulation cascade.** Factor Xa and thrombin are part of the common blood coagulation pathway. Factor X can be either activated to fXa by proteases of the intrinsic or extrinsic activation pathway (indicated by the orange and green box). Active fXa catalyzes the conversion of prothrombin to thrombin that in turn activates fibrinogen to the polymerizing fibrin (blue box, common pathway). Thrombin also generates the transglutaminase fXIIIa, which crosslinks and stabilizes fibrin clots. Clot degradation is mediated by plasmin.

For the drug development of matriptase inhibitors it is therefore necessary to not only focus on potency and bioavailability but also on selectivity against the named proteases. For all published matriptase tailored 3-amidinophenylalanine derivatives and also some bis-benzamidine derivatives the selectivity towards thrombin and factor Xa had been analyzed <sup>[159, 210, 217, 218, 224]</sup>. Exemplary compounds showing a good selectivity profile are summarized in Table 4-2 and include compounds with a selectivity of matriptase over thrombin that ranges from 210 to 6750-fold and over fXa from 52 to 5250-fold.

**Table 4-2: Selectivity profile of multibasic matriptase inhibitors.**

3-amidinophenylalanine derivatives					
	No.	R	K <sub>i</sub> [μM]		
			Mat	Thr	fXa
 Compounds <b>22</b> and <b>23</b> <sup>[159]</sup> ; <b>21</b> <sup>[217]</sup> and <b>24</b> <sup>[218]</sup>	<b>21</b>		0.004	27 [6750]	21 [5250]
	<b>22</b>		0.005	13 [2000]	15 [3000]
	<b>23</b>		0.002	6 [3000]	9.5 [4750]
	<b>24</b>		0.008	36 [4500]	1.2 [150]
Bis-benzamidine derivatives					
	No.	R	K <sub>i</sub> [μM]		
			Mat	Thr	fXa
	<b>25</b>		0.003	2.2 [730]	0.50 [170]
	<b>26</b>		0.01	2.1 [210]	0.52 [52]
	<b>27</b>		0.02	4.2 [210]	1.6 [80]
 Compounds <b>25-28</b> <sup>[210]</sup>	<b>28</b>	-	0.006	2.7 [450]	1.1 [180]

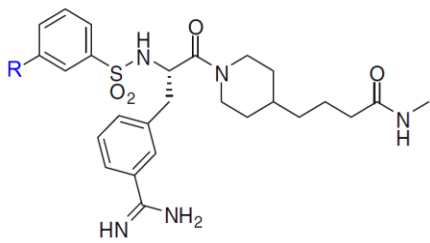
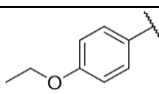
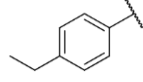
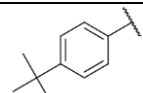

Mat: matriptase, Thr: thrombin, fXa: factor Xa; selectivity ratios ( $K_i$  Thr/ $K_i$  Mat or  $K_i$  fXa/ $K_i$  Mat) are given in box brackets.

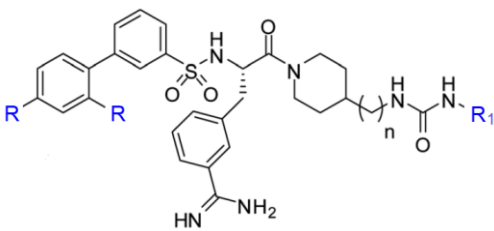
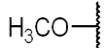



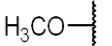



Although highly selective and likewise very potent matriptase inhibitors of the 3-amidinophenylalanine type have been identified, they share the drawback of a dibasic or even tribasic character that is connected to decreased bioavailability. Therefore, the development of monobasic amidine compounds, which are convertible into hydroxyamidino-prodrugs as shown for Mesupron, is highly intended. First monobasic inhibitors of the 3-amidinophenylalanine type with  $K_i$  values starting from 24 nM were published in 2009 <sup>[224]</sup> (Table 4-3 upper part). HAMMAMI and colleagues developed further monobasic derivatives with improved potency ( $K_i$  = 900 pM – 5 nM; Table 4-3 lower part) <sup>[218]</sup>. However, all monobasic analogues have a limited



selectivity profile towards thrombin and/or fXa with some of them being even more potent thrombin inhibitors (including compounds **29**, **30** and **32** - **34**). The highest selectivity (410-fold) towards thrombin was achieved by inhibitor **35** with a  $K_i$  of 1.23  $\mu\text{M}$  whilst retaining high matriptase affinity ( $K_i = 3$  nM). Anyhow, with a  $K_i$  value of 10 nM this compound is also a very potent fXa inhibitor. The highest matriptase selective over fXa shows compound **34** with a 32-fold selectivity. However, this compound is an even stronger thrombin inhibitor with an inhibition constant of 0.2 nM.

**Table 4-3: Potency and selectivity profile of existing monobasic 3-amidinophenylalanine derivatives.**

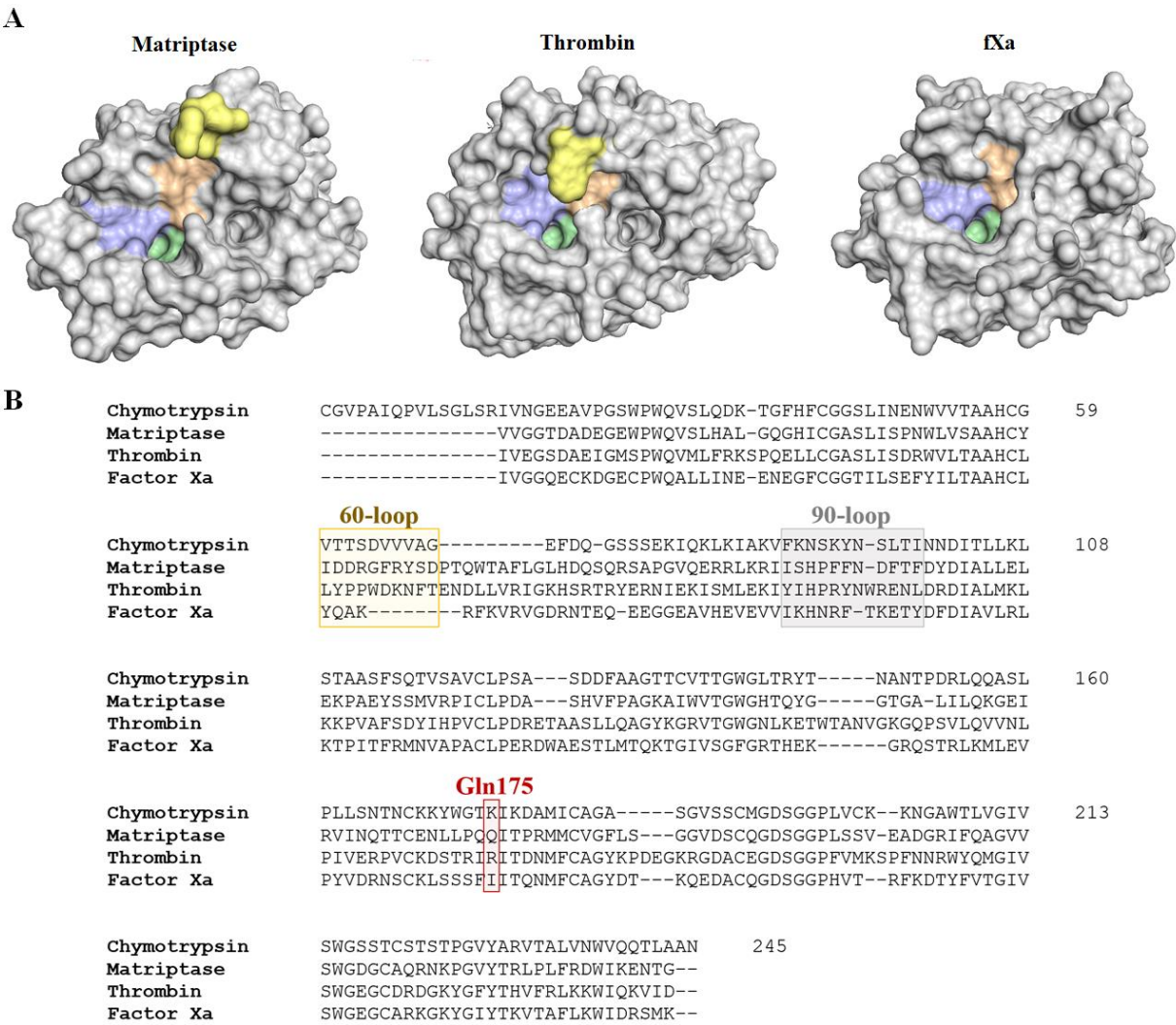
			K <sub>i</sub> [μM]		
Compound <b>29-32</b> : <sup>[224]</sup>	No.	R	Mat	Thr	fXa
	<b>29</b>		0.028	0.006 [0.02]	0.64 [0.64]
	<b>30</b>		0.024	0.004 [0.17]	0.37 [15]
	<b>31</b>		0.087	2.05 [24]	0.83 [9.5]
	<b>32</b>		0.03	0.001 [0.03]	0.70 [23]

			K <sub>i</sub> [μM]				
Compounds <b>18, 33-35</b> : <sup>[218]</sup>	No.	R	R <sub>1</sub>	n	Mat	Thr	fXa
	<b>33</b>			2	0.002	0.0008 [0.4]	0.05 [25]
	<b>34</b>			2	0.0009	0.0002 [0.22]	0.03 [33]
	<b>35</b>			0	0.003	1.23 [410]	0.01 [3]
	<b>18</b>			0	0.005	0.63 [126]	0.02 [4]

Mat: matriptase, Thr: thrombin, fXa: factor Xa; selectivity ratios ( $K_i$  Thr/ $K_i$  Mat or  $K_i$  fXa/ $K_i$  Mat) are given in box brackets.

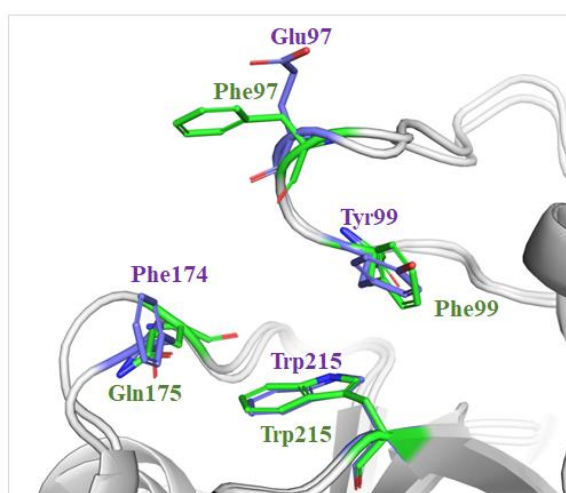
For all three enzymes, numerous crystal structures are available that enable the detailed comparison of the specific binding sites. Matriptase has a Ser190 adjacent to Asp189 at the bottom of the S1 pocket resulting in a more hydrophilic but narrower S1 pocket compared to proteases of the Ala190 type such as thrombin and factor Xa <sup>[226]</sup>. Thrombin and matriptase both possess a unique 60-insertion loop on top of the S2 binding site (highlighted yellow in

Figure 4-4). Whereas the segment Tyr60A-Pro60B-Pro60C-Trp60D of thrombin results in a well-formed but more restricted S2 binding site, the 60-insertion loop of matriptase has an averted orientation resulting in an open and accessible active site cleft (Figure 4-4 A). Therefore, the part of the compound addressing the proximal binding site (C-terminal residue of 3-amidinophenylalanine derivatives) is instrumental to achieve selectivity of matriptase inhibitors over thrombin. The proximal S2 binding site of fXa, which prefers Gly in P2 position, is rather unspecific and small due to the presence of the relatively bulky Tyr99 on the tip of the 90-hairpin loop [227].



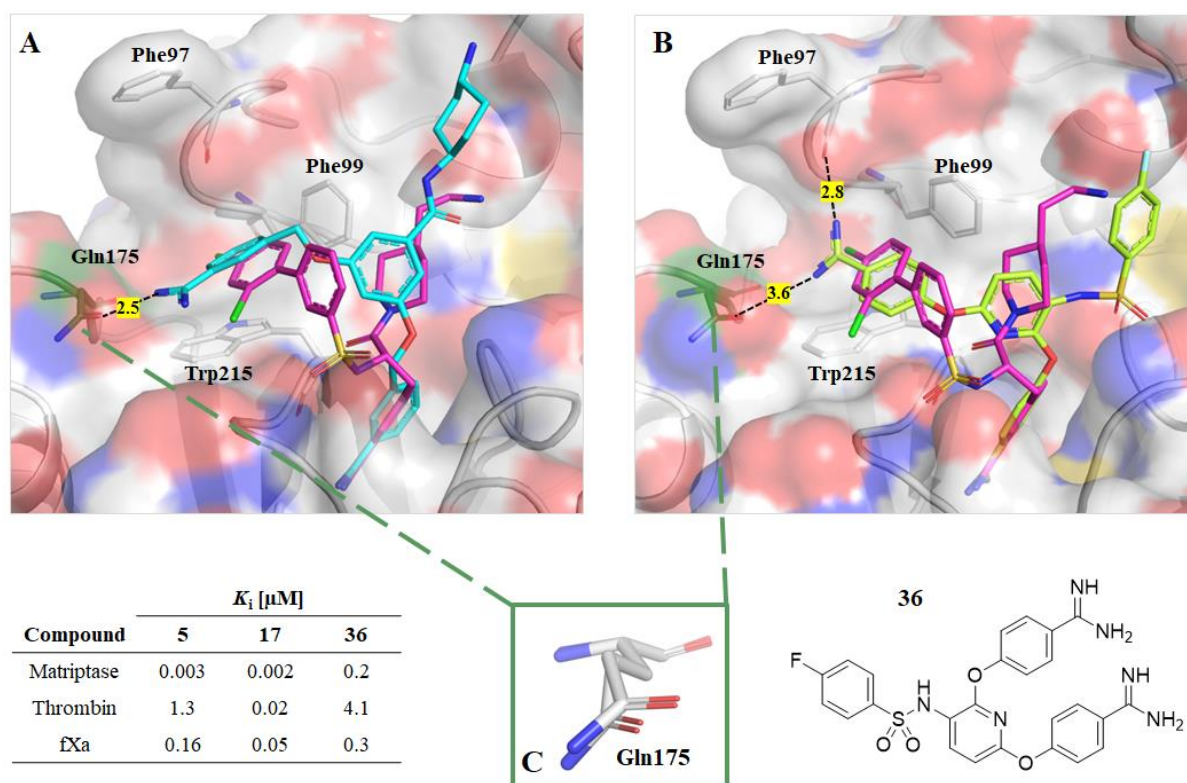
**Figure 4-4: Surface representation and multiple sequence alignment of different protease domains of the S1A-fold. A:** Surface representations of matriptase (PDB code: 2GV6), thrombin (PDB code: 4E7R) and fXa (PDB code: 5K0H) in grey with colored binding sites: S1 pocket in green, proximal S2 pocket in wheat and the distal S3/4 binding site in light blue. The characteristic 60 loops of matriptase and thrombin are highlighted in yellow. **B:** Multiple sequence alignment of chymotrypsin, matriptase, thrombin and fXa SPDs. Important regions for matriptase are highlighted including the 60-loop (yellow box), the 90-loop (grey box) and the Gln175 (red box). Alignment was done with Clustal Omega using the amino acid sequence from merops codes S01.001, S01.302, S01.217 and S01.216, respectively.

The distal S3/4 binding site of all three serine proteases has a hydrophobic character. Whereas this cavity is well defined in the case of thrombin and fXa, it is more open on the western side in the case of matriptase (Figure 4-4 A and Figure 4-5). In fXa this pocket has a box-shaped hydrophobic/aromatic cavity lined by aromatic side chains of Tyr99, Phe174 and Trp215 whereas in matriptase the distal binding site is formed by the three aromatic amino acids: Phe97 on the top, Trp215 at the bottom surrounded by Phe99 on the right and Gln175 on the left side (Figure 4-5). Both, fXa and matriptase, accept also basic residues and substitutions of the N-terminal aryl ring in *ortho*- and *para*-position and numerous inhibitors with such substitutions were identified to exert high affinities for both proteases [217, 218].



**Figure 4-5: Amino acid residues of the distal (S3/4) binding pocket of matriptase and fXa.** Protease structure is given in grey cartoon. Important amino acid residues forming the S3/4 binding sites are shown in sticks and colored in green for matriptase and purple for fXa. The picture was made with PyMol using PDB codes 2GV6 (matriptase) and 5K0H (fXa).

The 90-loop as well as the non-conserved Gln175 have been identified as reasonable structural targets to achieve selectivity of matriptase over fXa. GOSWAMI and colleagues successfully targeted the side chain carbonyl of Gln175 with a 4-aminocyclohexyl moiety or the amidines of bis-benzamidine inhibitor types, including compound **5** (cyan in Figure 4-6 A) and compound **36** (light green in Figure 4-6 B). Compared to compound **5** (structure given in Table 4-1), the amidino function in *para*-position of compound **36** is with 3.6 Å more distant and might only allow a very weak hydrogen bond interaction with the Gln175. However, a strong H-bond interaction to the carbonyl of Phe97 on top of the distal S3/4 binding pocket is suggested by the proximity of 2.8 Å. The Gln175 of both structures (Figure 4-6 C) exhibits some flexibility and can adapt to the respective compound.



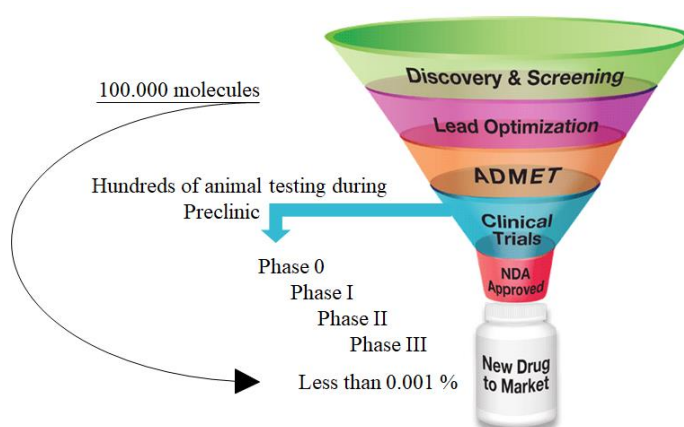
**Figure 4-6: Crystallographic binding mode of bis-benzamidines in matriptase superimposed with the model of inhibitor 17 of the 3-amidinophenylalanine type.** Compound **5** in cyan (A) and compound **36** <sup>[191]</sup> in light green (B) were crystallized in complex with matriptase (PDB codes 4O97 and 4JZ1, respectively). Black dashed lines indicate polar interactions in the S3/4 binding site. Their distances (in Å) are given in yellow boxes. Important matriptase residues Phe97, Phe99, Gln175 and Trp215 are labeled and shown as sticks. The Gln175 is highlighted in green with a zoom-in alignment from both structures in C. The structure of inhibitor **36** is shown in the lower right and the structures of compounds **5** and **17** in Table 4-1.  $K_i$  values for matriptase, thrombin and fXa are given in the lower left.

Binding mode comparison of compounds **5** and **36** with that of compound **17** (pink in Figure 4-6 A and B, respectively) suggests an amidino function in *meta*-position to be favorable in order to target the Gln175 side chain also in case of the 3-amidinophenylalanine inhibitor type. This additional interaction with the S3/4 binding site might be advantageous for matriptase potency and therefore increasing selectivity over fXa, where Phe174 is located in that position (Figure 4-5). To achieve an improved bioavailability, a non-charged hydroxy group could be used instead of amino or amidino groups as hydrogen bond donors in case of monobasic compounds.

#### 4.1.5. Computer assisted matriptase inhibitor design for drug discovery

The drug development process is cumbersome, time consuming and expensive. It typically takes more than 10 years to develop one drug and costs have been estimated to be > 2 billion USD. Figure 4-7 shows the different phases from the initial idea to a new drug on the market. Computer aided/assisted drug design (CADD) is one of the most effective new methods for facilitating and expediting this process, thereby saving time, money and resources <sup>[228]</sup>.





**Figure 4-7: Workflow of drug discovery.** The process of drug development can be divided in five major phases including the lead identification (green), its optimization (pink), the ADMET related optimization and testing (orange) and the clinical phase (blue) before the approval phase by the respective authorities (red). Figure was modified from [www.enzolifesciences.com](http://www.enzolifesciences.com).

Virtual screening is a key area in CADD and can be divided into two main categories: structure-based and ligand-based virtual screening. Ligand-based techniques are suitable for quickly retrieving analogues (i.e. similar molecules) of known active ligands. Thus, these approaches are very helpful when no structural information about the target is known. The most common structure-based approach is docking, which aims at predicting the binding mode of a ligand within the binding site of a protein. These predictions are carefully evaluated and then lead to hits, as well as an approximate estimate of the free energy of binding from the resulting complex (included in the docking score). CADD is vital in order to make rapid, yet reasonable predictions of the binding mode. However, docking has limits, as its purpose is to find the best compromise between the speed of screening and its accuracy. It is regarded as time-efficient, but the prediction of the binding affinity is only a rough estimation that is solely used for ranking (i.e. prioritize which compounds are more likely to bind favorably).

Before the first crystal structure of matriptase was available, docking was used to predict the binding of compounds based on matriptase models. In this way, ENYEDY and team discovered 15 new bis-benzamidine inhibitors of matriptase by docking against a comparative model built using thrombin as a template (e.g. compound **3** in Table 4-1) <sup>[209]</sup>. The binding of benzamidine to the S1 pocket was confirmed with the first matriptase crystal structure in 2002 <sup>[162]</sup>. In 2006, STEINMETZER and team published the two first and only structures with inhibitors of the 3-amindinophenylalanine type in complex with matriptase so far, including compound **20** (Figure 4-2). Based on these crystal structures, GOSWAMI and colleagues used docking for the structure-guided discovery of substituted benzenes and were able to confirm their docking results by crystallization <sup>[210]</sup> (e.g. compound **5** in Table 4-1 and Figure 4-6). In 2015, GOSWAMI made use of docking for applying a fragment-linking approach and discovered new P1 serine amides,

which were substituted with benzamidine on the Ser side chain <sup>[197]</sup>. Again, crystallization helped to evaluate these docking results. Docking was also used by COLOMBO <sup>[201]</sup> to discover new peptidomimetic matriptase inhibitors and by AVRUTINA <sup>[207]</sup> and FITTLER<sup>[206]</sup>, where new SFTI-1 variants have been designed. In general, docking of large molecules containing many rotatable bonds, such as inhibitors of the 3-amidinophenylalanine type is challenging and the outcome needs to be evaluated carefully. For every rotor, the number of possible conformers increases yielding a combinatorial explosion problem. The more rotatable bonds a molecule has, the less likely it is that the docking software finds the best conformer and reliably places it within the binding site. In the case of compounds with established binding modes, including 3-amidinophenylalanine derivatives, the docking results can be reasonably directed. Several docking applications allow to fix specific interactions, thereby reducing the present rotors of a molecule and significantly increasing the robustness of docking results.

In this thesis, docking was integrated in the SAR study to rapidly suggest compatible replacements of a given template inhibitor that might improve affinity but more importantly selectivity of our matriptase ligands with a known binding mode.

#### **4.1.6. Goals of computer assisted matriptase inhibitor design**

In collaboration with a computational and synthetic chemist as well as crystallographers the main goal of the project was to identify new potent monobasic matriptase inhibitors of the 3-amidinophenylalanine type that show an improved selectivity against thrombin and fXa. In detail, the goals of this project were:

- Computer assisted lead optimization with a focus on selectivity and bioavailability including the feasibility,
- a SAR analysis including compound synthesis and determination of inhibitory effect on the target proteases matriptase, thrombin and factor Xa,
- the determination of the crystallographic binding mode of selected compounds in complex with matriptase and/or related serine proteases,
- the testing of selected compounds in cell culture regarding cytotoxicity and inhibitory effect on H9N2 influenza virus propagation (collaboration with the group of PROF. DR. FRIEBERTSHÄUSER, Institute of Virology).

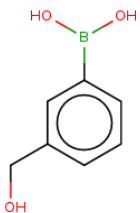
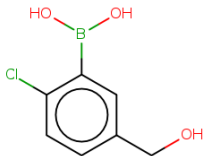
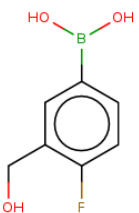
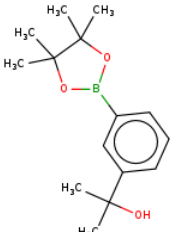
## 4.2. Results and discussion of matriptase inhibitor design and characterization

This project was part of an interdisciplinary collaboration between different colleagues of the pharmaceutical chemistry department of the Phillips University of Marburg. The computational part was carried out in collaboration with the computational chemist DR. FLORENT CHEVILLARD (group of PROF. DR. KOLB). The synthesis and analysis of the compounds were performed by OLIVER PILGRAM, who works on his doctoral thesis in the group of PROF. DR. STEINMETZER. DR. STEFAN MERKL from CrystalsFirst GmbH performed the inhibitor soaking and measuring of matriptase and trypsin crystals. The structure refinement was carried out by the scientists DR. JANIS MUELLER and DR. STEFAN MERKL of CrystalsFirst GmbH.

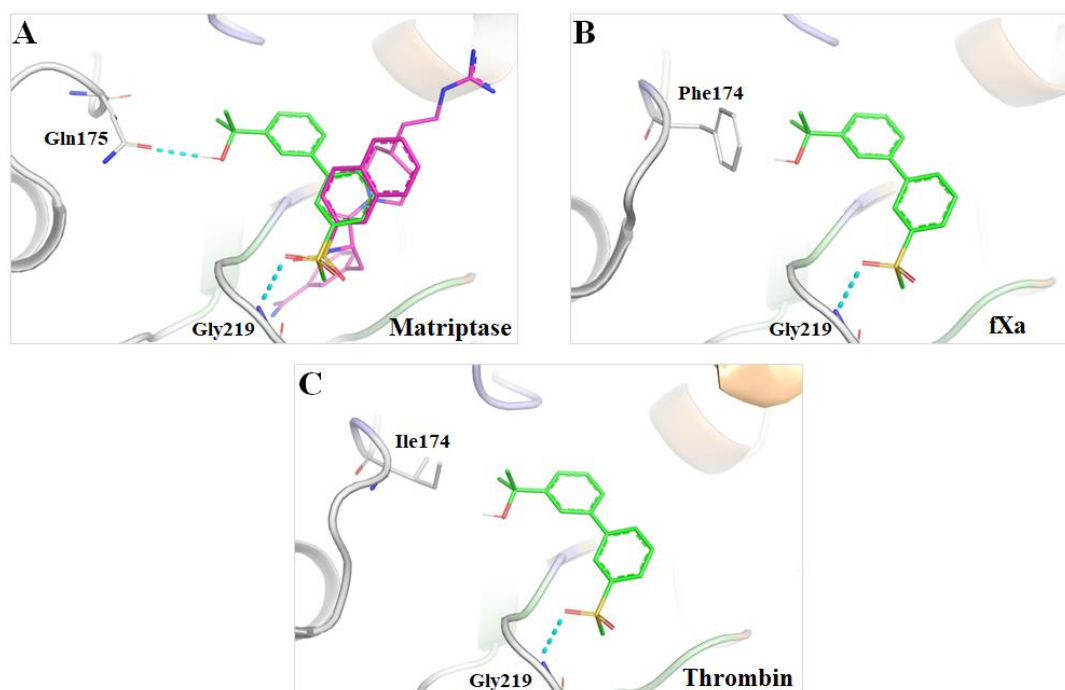
### 4.2.1. Selection of compatible building blocks

The crystal structure of matriptase in complex with compound **20** (PDB code 2GV6, Figure 4-2) was used as starting point for docking studies. In order to target specific amino acids in the distal matriptase S3/4 binding pocket the naphthyl moiety of compound **20** was replaced with a substituted biphenyl moiety, which can be obtained by SUZUKI coupling of an iodo-benzene core with various phenylboronic acids (see also compounds **16-18**, Table 4-1). As compound **20** is too complex for rapid docking simply *m*-iodobenzene methylsulfone was used as a surrogate compound. The benzene ring and the sulfone group were sufficient to have a shape reference for visual inspection of the docking poses. Using PINGUI, commercially available boronic acids were filtered and merged with the surrogate *m*-iodobenzene methylsulfone. All 2000 resulting virtual products were docked and the top 500 virtual product docking poses (based on the docking score) were visually inspected. For each pose, the main quality criterion was the overlap of the methylsulfone benzene group with the crystal structure of compound **20** in complex with matriptase (Figure 4-8 A). Upon fulfillment of this criterion, it was assumed that the remaining part of the compound in the S1 and proximal pocket would not significantly change. Compounds interacting with Gln175 were highly prioritized as this residue was the main target of the optimization regarding matriptase selectivity over fXa. Additionally, compounds containing an amine or amidine moiety were initially discarded due to limited bioavailability issues. The chosen surrogate compounds were docked with fXa and thrombin in order to assure an either unfavorable binding mode or a poor docking score. An unfavorable binding mode was thought to penalize the compound affinity, thus improving the selectivity towards matriptase. Under consideration of the used libraries, eight products were found to respect all the criteria. After evaluating the price and availability, four boronic acids retained (compounds **37-40**, Table 4-4) that were subsequently subjected to compound synthesis.

**Table 4-4: Boronic acids subjected to inhibitor synthesis (based on docking results).**

Selected boronic acids			
			
37	38	39	40

The docking pose of the virtual product containing boronic acid **40** is shown for each receptor in Figure 4-8. Docking proposes an interaction between the compound's terminal hydroxy group and the Gln175 side chain of matriptase (Figure 4-8 A). Under the giving orientation and shape, no polar interactions are predicted to occur in the case of fXa (Phe174, see Figure 4-8 B) and thrombin (Ile174, Figure 4-8 C). Indeed, the polar hydroxyl group is pointing towards the hydrophobic residues in case of fXa and thrombin, which is energetically unfavorable mostly due to the cost of desolvation.



**Figure 4-8: Predicted binding mode of the virtual biphenyl product containing the substituted phenyl ring of boronic acid **40** in matriptase, fXa and thrombin.** The virtual product (green C-atoms) was created using PINGUI and subsequently docked with the respective protease. **A:** Predicted binding mode of the virtual product in matriptase in superimposition with compound **20** (PDB code: 2GV6, carbons in pink). The overlap of the sulfone benzene (sulfur in yellow, oxygen in red, phenyl rings in pink and green, respectively) was the main quality criterion for compound selection. **B:** Docking of the virtual product with fXa (PDB code: 5K0H). **C:** Docking of the virtual product with thrombin (PDB code: 4E7R). Protease binding sites are shown in cartoon style and important amino acids as sticks. The cyan dashed lines indicate predicted polar interactions between the ligand and the proteases including the interaction of the compound with the NH of residue Gly219, which is present in all three proteases.



### 4.2.2. Compound synthesis

The synthesis of the 3-amidinophenylalanine derivatives has been established in our laboratories after methods as described before [159, 217, 218, 224]. In the first step, 3-iodophenylsulfonylated 3-cyanophenylalanine (Figure 4-9) is prepared and serves as the initial scaffold that is subsequently modified N- and C-terminally. In the first series, different inhibitors with a cyclohexyl urea portion at the C-terminus were synthesized. The bulky cyclohexyl urea has shown to prevent proper binding below the 60-insertion loop of thrombin and resulted in decent matriptase selectivity of a monobasic derivative (410-fold, see compound **35** in Table 4-3) against thrombin [218]. While keeping the C-terminus constant, the N-terminus was modified in order to achieve additional selectivity over fXa. As for earlier monobasic derivatives [218, 224] an additional variably substituted benzene ring was linked to the phenylsulfonyl moiety of the initial scaffold. The resulting substituted biphenyl-3-sulfonyl moiety was obtained by SUZUKI coupling of the halide core molecule with varying phenylboronic acids. All building blocks contained a hydroxymethyl group that was predicted to act as hydrogen bond donor and to interact with the side-chain carbonyl of Gln175 on the left side of the distal binding pocket of matriptase, whereas the inhibitors' terminal phenyl ring should bind to Phe99 on the right side via hydrophobic contacts.

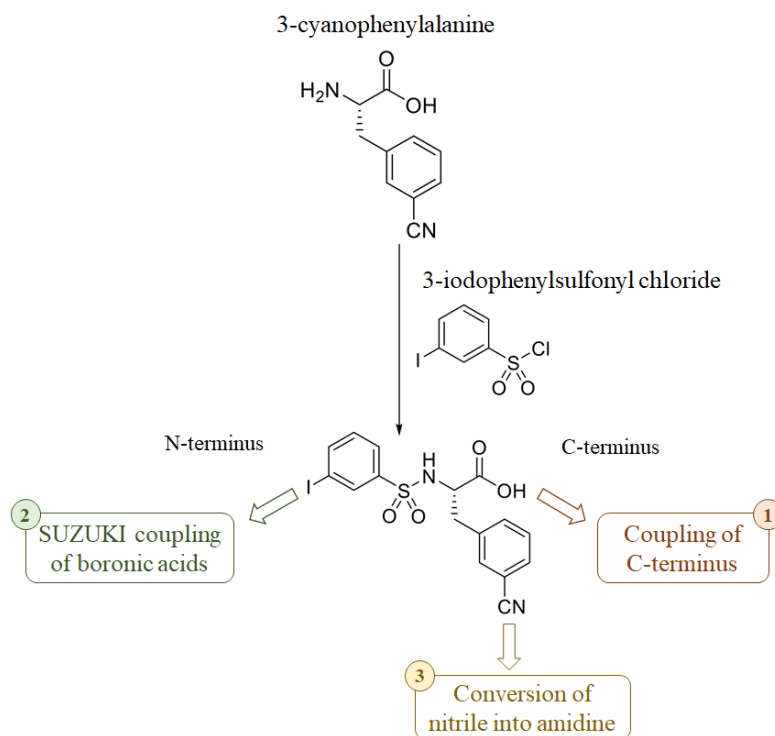


Figure 4-9: Simplified synthesis route of 3-amidinophenylalanine derivatives.

### 4.2.3. First compounds

Following the simplified synthesis workflow in Figure 4-9, four inhibitors (Table 4-5) were initially synthesized using the previously chosen boronic acids. All compounds showed only medium inhibition of matriptase with  $K_i$  values between 0.11  $\mu\text{M}$ -0.37  $\mu\text{M}$  and all four compounds are additionally more potent thrombin inhibitors ( $K_i$  values of 0.014  $\mu\text{M}$ -0.098  $\mu\text{M}$ ).

**Table 4-5: Inhibition of matriptase, thrombin and fXa by monobasic inhibitors with varying N-termini.**

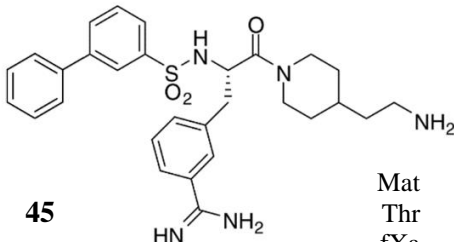
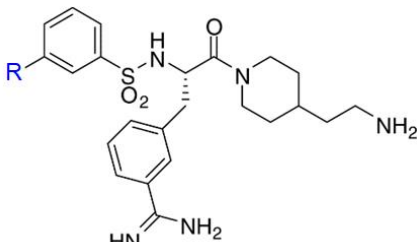


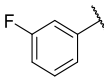
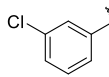
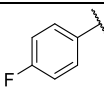
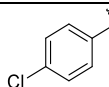
No.	R	$K_i$ [ $\mu\text{M}$ ]		
		Mat	Thr	fXa
<b>41</b> MI-479		0.21	0.042 [0.2]	1.31 [6.2]
<b>42</b> MI-482		0.29	0.097 [0.33]	1.88 [3.4]
<b>43</b> MI-481		0.11	0.098 [0.89]	1.02 [9.3]
<b>44</b> MI-480		0.37	0.014 [0.04]	2.66 [7.2]

Mat: matriptase, Thr: thrombin, fXa: factor Xa; selectivity ratios ( $K_i$  Thr/ $K_i$  Mat or  $K_i$  fXa/ $K_i$  Mat) are given in box brackets.

Low matriptase affinities probably result from an unfavorable combination of the additional ethylene spacer between the piperidine and the cyclohexylurea as well as suboptimal substituents on the aryl moiety. From these four compounds, inhibitor **43** containing the N-terminal 3-hydroxymethyl-4-fluorophenyl group, represents the most potent matriptase inhibitor ( $K_i = 0.11 \mu\text{M}$ ) and shows the highest selectivity over fXa (9.6-fold). Compared to compound **41**, compound **43** contains an additional fluoride in *para*-position of the terminal phenyl ring resulting in an affinity boost. The incorporation of fluorine atoms is a standard strategy to enhance affinity, metabolic stability and/or selectivity in pharmaceuticals [229, 230] and had been shown to also successfully improve inhibitory potency of matriptase tailored

compounds before <sup>[218]</sup>. Due to their high electronegativity, some halogens and especially fluorine atoms can interact with electropositive regions of receptor sites and therefore strengthen protein-ligand binding <sup>[231]</sup>. In the case of matriptase, substitutions in *ortho*- or *para*-position (compound **46**, **48**, **49** and **51**) with either fluoride <sup>[218]</sup> or chloride <sup>[217]</sup> resulted in increased matriptase but also fXa affinity compared to the non-substituted analogue inhibitor **45** (Table 4-6). Halogen substitutions in *meta*-position showed a slightly reduced matriptase affinity compared to the other substituted aryls but resulted in higher matriptase over fXa selectivity (circa 60-fold, each). Compared to fluorine, chlorine is slightly less electronegative and can form halogen/ $\pi$ -interactions due to a positive electrostatic potential located at the tip of the chlorine atom, the so-called  $\sigma$ -hole <sup>[232]</sup>. This  $\sigma$ -hole is absent in fluoride substituents.

**Table 4-6: Inhibition of matriptase, thrombin and fXa by known dibasic inhibitors containing N-terminal halogen substitutions.**

<div></div> <p><b>45</b></p> <p>Mat 0.110 <math>\mu</math>M Thr 0.056 <math>\mu</math>M fXa 5.4 <math>\mu</math>M</p>					<div></div>				
<i>K<sub>i</sub></i> [ $\mu$ M]									
No.	<b>R</b>	Mat	Thr	fXa	No.	<b>R</b>	Mat	Thr	fXa
<b>46</b> MI-403		0.023	0.04	0.081	<b>49</b>		0.03	0.01	0.70
<b>47</b> MI-404		0.052	0.04	3.2	<b>50</b>		0.09	0.05	5.6
<b>48</b> MI-405		0.033	0.005	1.12	<b>51</b>		0.03	0.01	0.91

Compounds **45,49-51** from <sup>[217]</sup> and compounds **46-48** from <sup>[218]</sup>. Mat: matriptase, Thr: thrombin, fXa: factor Xa

The present effect of the fluoride on the slight gain in affinity in the case of compound **43** compared to compound **41** cannot be clearly determined at this point. However, since fluoride substituents are known to be able to form a wide array of interactions with carbonyl carbons or CH groups <sup>[231]</sup>, its *para*-position probably facilitates such interactions by moving the fluoride close towards the protein surface. In contrast, the additional chloride in *ortho*-position of compound **42** does not improve affinity compared to the halogen-free analogous compound **41** (Table 4-5). As the cyclohexyl urea with the additional ethyl linker of compounds **41-44** is not restricting thrombin affinity, the C-terminus was subsequently optimized (Table 4-7).

Compound **35** (Table 4-3) with 2,4-dimethoxy substituted phenyl on the N-terminus served as a reference compound.

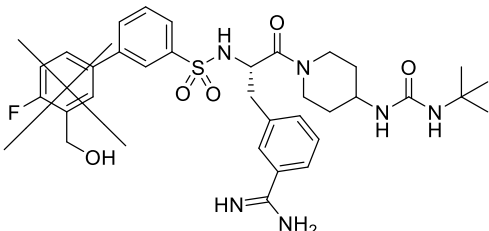
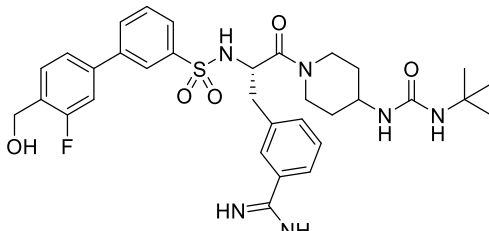
**Table 4-7: Inhibition of matriptase, thrombin and fXa by monobasic inhibitors with variable C-termini.**

		$K_i$ [ $\mu$ M]		
No.	R	Mat	Thr	fXa
<b>52</b>		0.020	0.085	0.017
MI-483			[4]	[0.9]
<b>53</b>		0.003	0.273	0.003
MI-484			[91]	[1]
<b>54</b>		0.004	2.3	0.002
MI-485			[575]	[0.5]

Mat: matriptase, Thr: thrombin, fXa: factor Xa; selectivity ratios ( $K_i$  Thr/ $K_i$  Mat or  $K_i$  fXa/ $K_i$  Mat) are given in box brackets.

Enzyme kinetic data revealed 4-*tert*-butylureido-piperidide of compound **54** as a suitable C-terminus. In combination with 2,4-dimethoxyphenyl at the N-terminus it shows a 575-fold higher matriptase affinity compared to thrombin and is even more selective than the analogous compound **35** ( $K_i$  Thr/ $K_i$  Mat: 410-fold). Even though matriptase ( $K_i$  = 4 nM) and fXa ( $K_i$  = 2 nM) are inhibited equally by compound **54** the 4-*tert*-butylureido-piperidide was chosen for further optimizations of the N-terminus. Since 3-hydroxymethyl-4-fluoridophenyl, as also used in compound **43**, was the most promising N-terminus of the initial four compounds containing the docking based selected boronic acids, it was now submitted to synthesis in combination with the aforementioned C-terminal 4-*tert*-butylureido-piperidide. However, based on a mix up of boronic acids by the vending company the following compound **55** (Table 4-8) had been identified to contain the hydroxymethyl substitution in *para*- but not as intended in *meta*-position. This mix up was detected after synthesis by NMR characterization and by enzyme kinetic measurements with a resynthesized reference compound as well as during structure refinement following crystallization of the respective compound in complex with matriptase (shown and discussed in 4.2.4).

**Table 4-8: Inhibitory effect including the selectivity profile of compound 55.**

Intended structure	Actual structure of compound <b>55</b>
	
	<b><i>K</i><sub>i</sub> [μM]</b>
	<b>No.</b> Mat                      Thr                      fXa                      Try
	1 <sup>st</sup> synthesis                      -                      0.015 (± 0.002)                      23 (± 3.031)                      0.19 (± 0.029)                      0.0006 (N.D.)
	2 <sup>nd</sup> synthesis                      -                      0.014 (± 0.0001)                      22 (± 1.8)                      0.16 (± 0.011)                      N.D.
	Mean <b>55</b> (MI-1904)                      0.015                      23 [1533]                      0.18 [12]                      N.D.

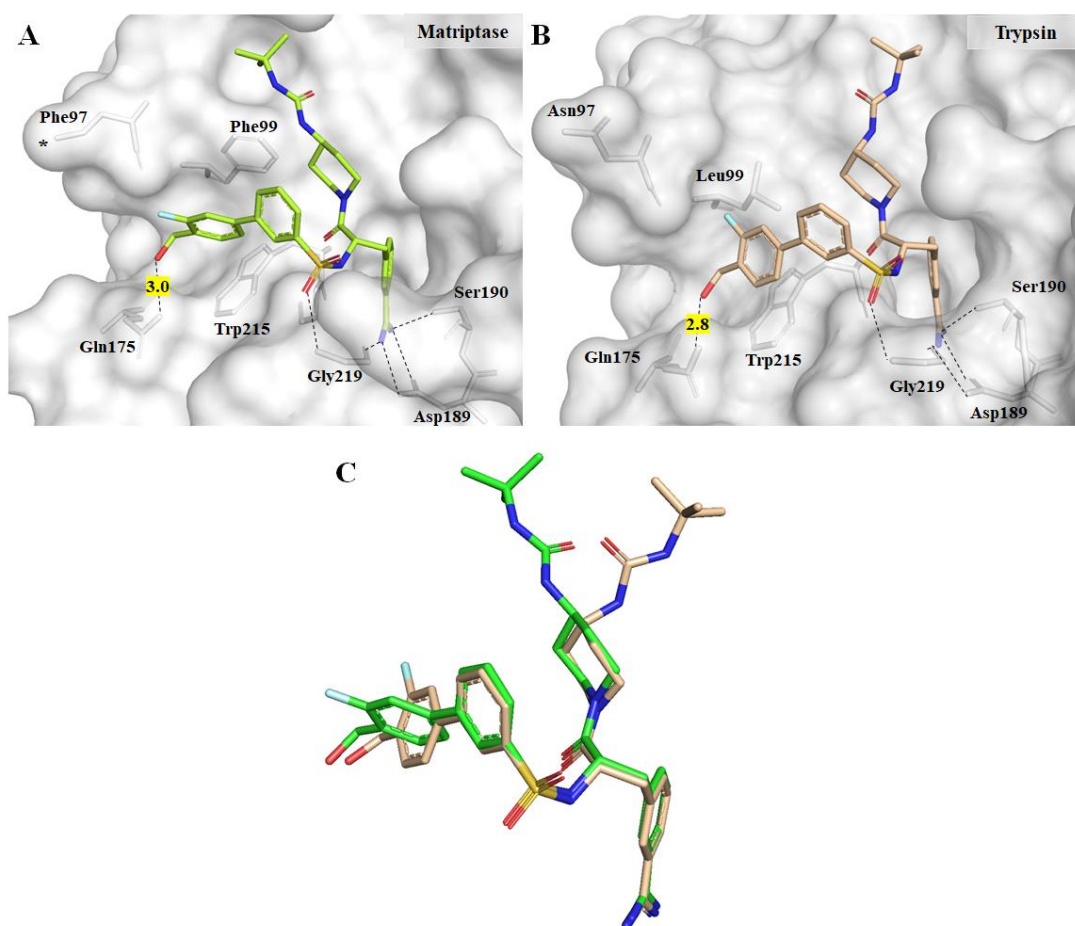
Mat: matriptase, Thr: thrombin, fXa: factor Xa, Try: trypsin; standard deviations are given in round brackets; selectivity ratios ( $K_i$  Thr/ $K_i$  Mat or  $K_i$  fXa/ $K_i$  Mat) are given in box brackets

After repeated synthesis of compound **55** using a 3-fluoro-4-(hydroxymethyl)benzene boronic acid from a different vendor and detailed comparative analysis (data not shown) the structure of compound **55** as shown in Table 4-8 was further confirmed. With only a slight decrease in matriptase affinity ( $K_i = 15$  nM) compared to compound **54** and significantly decreased thrombin affinity by a factor of 10 ( $K_i = 23$   $\mu$ M) compound **55** has an excellent matriptase selectivity over thrombin (1533-fold). In addition, also fXa affinity could be decreased from a  $K_i$  value of 2 nM (compound **54**) to 180 nM in the case of compound **55**, resulting in a 12-times higher matriptase affinity compared to fXa.

#### 4.2.4. Binding mode of inhibitor **55** in complex with matriptase, thrombin and trypsin

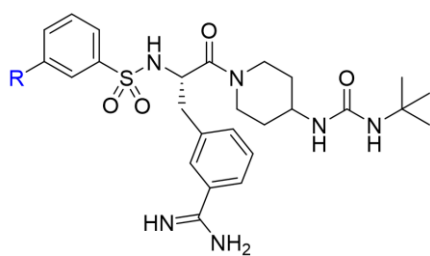
In the following the binding mode of the monobasic compound **55** was determined in complex with matriptase, thrombin and trypsin protein crystals. Whereas trypsin was directly co-crystallized with the inhibitor, matriptase and thrombin (thrombin experiments conducted by the PhD student ANNA SANDNER) were co-crystallized with benzamidine each before being soaked with compound **55**. Matriptase and trypsin structures were refined based on molecular replacement. Final structures are available under PDB codes 6T9T for matriptase and 6T9V for trypsin. The binding mode of compound **55** in thrombin could not be determined based on the given electron density since only the benzamidine moiety was clearly placeable in the S1 pocket. In matriptase and trypsin, compound **55** shows the typical Y-shape binding mode of

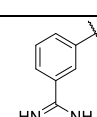
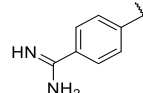
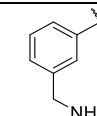







arylsulfonylated 3-amidinophenylalanine derivatives containing a C-terminal cyclic tertiary amide residue. Moreover, the central 3-amidinophenylalanine forms the typical interactions in the S1 pocket including the polar contacts of the amidino function to Asp189, Ser190 and Gly219 as well as the interaction between the H-bond of the sulfonyl oxygen to the NH of Gly219 (Figure 4-10 A and B). The antiparallel  $\beta$ -sheet interaction between the 3-amidinophenyl backbone to Gly216 is also present but not highlighted in these figures. The C-terminal 4-*tert*-butyl urea moiety shows a distinct orientation that differs between the matriptase (Figure 4-10 A) and the trypsin (Figure 4-10 B) crystal structure. Whereas trypsin prefers the axial position, matriptase has a higher affinity to the equatorial position (Figure 4-10 C), which is expected for sterically demanding substituents. Hence, in trypsin, the urea substituent is positioned perpendicular to the piperidine ring, while in matriptase, it points towards the same direction as the piperidine ring. As a result of this parallel orientation relative to the piperidine ring, the terminal 4-*tert*-butyl group of compound **55** points towards matriptase's cation cleft that is formed by the carbonyl groups of Ile60 and Asp60B from the 60-loop above of His57 and the carbonyl of Asp96. In contrast to matriptase, trypsin lacks a specific 60-loop resulting in an open non-well-defined area above His57. Even though electron density might suggest a partial axial position of the 4-*tert*-butyl urea moiety in the case of matriptase it could not be reasonably confirmed during refinement. However, both isomers are always present in solution (in DMSO- $D_6$ ), as proven by double signals in the NMR spectra of the inhibitor (data not shown). The N-terminal *para*-hydroxymethyl group of compound **55** interacts with the side chain of Gln175 of both proteases, as the 3-hydroxymethyl-4-(fluoro)benzene was predicted to do by docking (Figure 4-8). In addition, the matriptase structure (Figure 4-10 A) suggests a preference of the *para*-hydroxymethyl group to interact with the Gln175 on the left side of the distal binding pocket instead of interacting with the carbonyl of Phe97 on the top, which was suggested by the modeled binding mode of the related inhibitor **17** in Figure 4-6 B. The overlay of compound **55** from the matriptase and trypsin structures in Figure 4-10 C also shows a varying orientation of the N-terminal aryl ring, resulting in a slightly different exposition of the fluoride. However, the 4-hydroxymethyl has a similar distance to the Gln175 with 2.8 Å and 3.0 Å for trypsin and matriptase, respectively. Since the availability of matriptase crystals is the limiting factor for detailed analysis of compound binding modes, similar proteases such as thrombin<sup>[218]</sup> and trypsin<sup>[128]</sup> have been used so far to investigate the binding mode of novel compounds in our group. With the focus of matriptase selective inhibitors against thrombin, only trypsin represents an alternative to matriptase for compounds with very weak thrombin affinity. In addition, trypsin also harbors the non-conserved Gln175 (Figure 4-10 B) that was targeted in the inhibitor design of this study.



**Figure 4-10: Crystallographic binding mode of compound 55 in complex with matriptase (A, green carbon atoms) and trypsin (B, wheat carbon atoms) including an overlay of both ligand structures (C).** The protease surfaces in A+B are shown in light grey, important amino acids are labeled and shown as sticks. Black dashed lines implicate interactions between the compound and the respective protease. The distances (in Å) of the polar interactions between the hydroxymethyl group to the Gln175 side chain are given in yellow boxes. \* indicates the missing phenyl ring in A of the matriptase Phe97 (not visible in the electron density).

Based on the findings so far, a second set of compounds was synthesized. To retain a weak inhibitory potency against thrombin these compounds had the C-terminal 4-*tert*-butylureido-piperidide in common. Varying N-termini were used in order to improve fXa selectivity and to investigate the influence of the variously substituted terminal aryl rings. As basic groups, such as amidine moieties, are known to suit well into the distal S3/4 binding pocket of matriptase, four dibasic compounds were included to compare the influence of aryl substitutions in *para*- and *meta*-position. The results are summarized in Table 4-9.

**Table 4-9: Inhibition of selected serine proteases by monobasic inhibitors with varying N-termini.**


No.	R	$K_i$ [ $\mu$ M]			
		Mat	Thr	fXa	Try
<b>56</b> MI-490		0.0007 ( $\pm$ 0.0001)	2.6 ( $\pm$ 0.64) [3700]	0.009 ( $\pm$ 0.001) [12]	0.005 (N.D.)
<b>57</b> MI-495		0.0008 ( $\pm$ 0.0001)	0.45 ( $\pm$ 0.07) [560]	0.084 ( $\pm$ 0.003) [100]	0.0007 (N.D.)
<b>58</b> MI-493		0.001 ( $\pm$ 0.0001)	1.3 ( $\pm$ 0.23) [1300]	0.0001* ( $\pm$ 0.00001) [0.1]	N.D.
<b>59</b> MI-494		0.004 ( $\pm$ 0.0006)	0.55 ( $\pm$ 0.11) [137]	0.010 ( $\pm$ 0.0009) [2.5]	0.009 (N.D.)
<b>60</b> MI-1902		0.059 ( $\pm$ 0.006)	0.49 ( $\pm$ 0.049) [8.3]	0.033 ( $\pm$ 0.004) [0.6]	N.D.
<b>61</b> MI-1903		0.011 ( $\pm$ 0.0004)	12 ( $\pm$ 1.4) [1090]	0.060 ( $\pm$ 0.010) [5.5]	N.D.
<b>62</b> MI-491		0.041 (0.005)	0.90 (0.023) [22]	0.024 (0.0004) [0.59]	N.D.
<b>63</b> MI-1905		0.020 ( $\pm$ 0.0008)	0.009 ( $\pm$ 0.0002) [0.45]	0.10 ( $\pm$ 0.008) [5]	N.D.
<b>64</b> MI-1906		0.008 ( $\pm$ 0.00007)	0.26 ( $\pm$ 0.020) [27]	0.005 ( $\pm$ 0.0003) [0.5]	N.D.
<b>65</b> MI-1907		0.014 ( $\pm$ 0.0005)	0.012 ( $\pm$ 0.06) [0.7]	0.79 ( $\pm$ 0.0003) [56]	N.D.

Mat: matriptase, Thr: thrombin, fXa: factor Xa, Try: trypsin; standard deviations are given in round brackets, selectivity ratios ( $K_i$  Thr/ $K_i$  Mat or  $K_i$  fXa/ $K_i$  Mat) in box brackets.

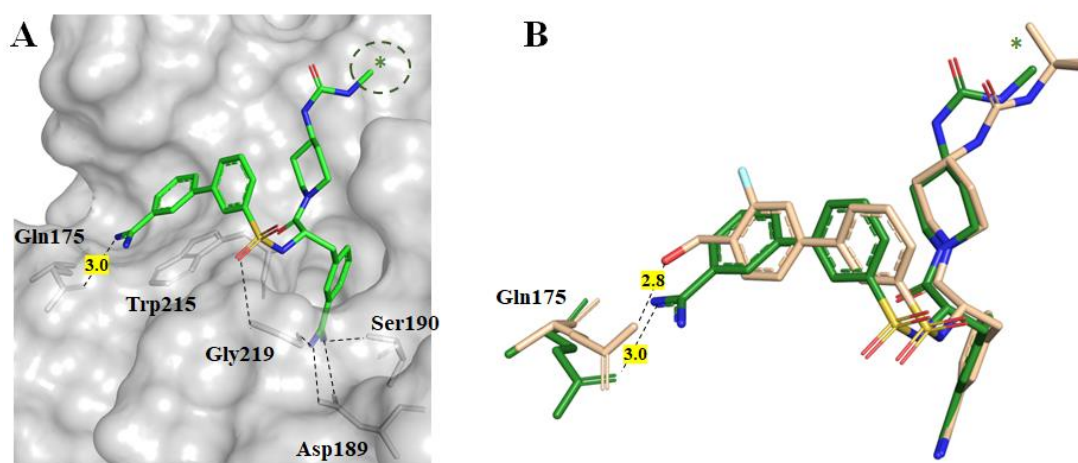
\*: slow binder.

a: dehydroxylated byproduct of compound 63 synthesis.

b: exact constitution of the hydroxymethyl and methyl substituent not known.



As expected, the four dibasic compounds **56-59** possess strong affinities to matriptase with the two compounds containing an amidine moiety (compounds **56** and **57**) inhibiting matriptase in the picomolar range (700 and 800 pM respectively). With a 3700-fold higher matriptase affinity, compound **56** represents the most matriptase selective inhibitor over thrombin of this thesis but has only moderate selectivity over fXa (12-fold). Compound **57** (amidine in *para*-position) shows the highest matriptase selectivity over fXa (100-fold). In parallel to compound **55** (Figure 4-10), the binding mode of compound **56** with the amidino substitution in *meta*-position on the terminal aryl ring was crystallographically determined in complex with trypsin (Figure 4-11).



**Figure 4-11: Crystallographic binding mode of compound **56** in complex with trypsin (A, PDB code: 6T9U) and overlay with compound **55** from PDB structure 6T9V (B).** Compound **56** is shown with green carbon atoms, compound **55** in wheat. The green star (circled in A) implies the three missing methyl groups of the C-terminal 4-*tert*-butyl moiety of compound **56** (structure shown in Table 4-9) that were not visible in the electron density during structure refinement. Selected polar interactions are given as black dashed lines. The distances (in Å) of the polar interactions between the amidine (compound **56**) or hydroxymethyl group (compound **55**) to the Gln175 side chain are given in yellow boxes. The protease surface (in A) is shown in light grey and important amino acids are labeled and highlighted in sticks.

Compound **56** shows the typical interactions with the residues Asp189, Ser190, Gly219, and Gly216 (the latter not shown in Figure 4-11 A). In addition, the 3-amidine group on the biphenyl moiety interacts with the side chain carbonyl of Gln175, as did the 4-hydroxymethyl substituent of compound **55** (PDB code: 6T9V) (Figure 4-11 B). This suggests that H-bond donors in *meta*- as well as *para*-position of the N-terminal aryl ring prefer to form a polar interaction with the Gln175 side chain on the left side instead of the carbonyl of residue Phe97 on top of the distal binding site. At the C-terminus, the three methyl groups of the *tert*-butyl group could not be refined by the given electron density (indicated by the star in Figure 4-11).

The replacement of the amidines in compounds **56** and **57** by aminomethyl moieties in compounds **58** and **59** caused a slight affinity reduction for matriptase ( $K_i$  values of 1 nM and

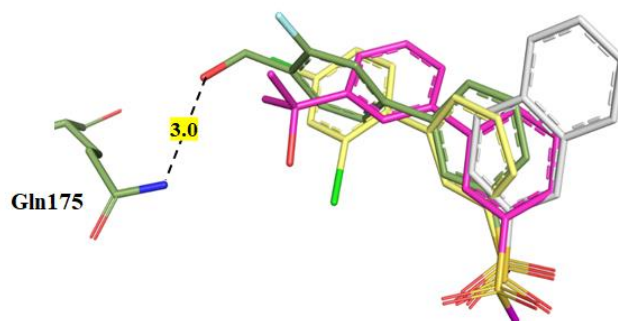
4 nM, respectively). The analogous monobasic compounds **60** and **61** with a hydroxymethyl replacing the aminomethyl are both even less potent matriptase inhibitors. With a matriptase  $K_i$  value of 59 nM compound **60** containing a hydroxymethyl substituent in *meta*-position shows a significant drop in affinity. Based on docking, the 3-hydroxymethyl was also predicted to interact with the Gln175 side chain. However, SAR results suggest only a preference of matriptase for *meta*-substituted dibasic compounds, but a clear preference for compounds with a hydroxymethyl moiety in *para*- instead of the *meta*-position. In general, matriptase potency of compounds **56** - **61** can be ranked by their  $K_i$  values depending on the terminal moieties as amidine (picomolar) < aminomethyl (one-digit nanomolar) < hydroxymethyl (two-digit nanomolar). Thrombin  $K_i$  values of these compounds vary between 0.45  $\mu$ M – 12  $\mu$ M, however, the inhibitory potency for thrombin is relatively weak due to the presence of the C-terminal *tert*-butyl urea group which is strongly disfavored by thrombin in most cases. Dibasic compounds containing *meta*-substituted amidine or aminomethyl biphenyls (**56** and **58**) are less favored by thrombin, instead, it prefers a hydroxymethyl group in *meta*-position as in compounds **60** and **62** in case of monobasic compounds. Compound **61** with the hydroxymethyl in *para*-position ( $K_i = 12 \mu$ M) is a 24-times less potent thrombin inhibitor than the *meta*-substituted compound **60** ( $K_i = 0.49 \mu$ M). The addition of a fluorine atom to hydroxymethyl substituted phenyls (as in compound **55** and **62**) further decreased thrombin affinities compared to the analogous compounds (**61** and **60**) regardless of their substitution pattern. Since the additional fluorine does not (or only marginally) affect matriptase affinities, fluorine-containing compounds have an increased matriptase selectivity against thrombin. The introduction of a second fluorine and the influences on thrombin affinity and matriptase selectivity are currently under investigation. As for matriptase, compounds **56-61** are also strong to medium fXa inhibitors ( $K_i$  values between 0.1 and 84 nM) with *meta*-substituted derivatives **58** and **60** showing an even higher affinity to fXa than to matriptase. For the mono- as well as dibasic compounds of this series, fXa has a clear preference for N-terminal biphenyl residues with *meta*-substitutions. Whereas the introduction of a fluorine in *meta*-position (compound **55** in Table 4-8) resulted in decreased fXa affinity compared to compound **61** ( $K_i$  values of 60 nM (**61**) and 180 nM (**55**)), no such effect could be observed for compound **62** (containing the fluoride substituent in *para*- and the hydroxymethyl in *meta*-position) compared to the halogen-free analogous compound **60** (Table 4-9). Compound **63** was designed with two hydroxymethyl groups (in *para*- and *meta*-position) and therefore thought to show improved matriptase affinity by interacting with Gln175 and Phe97. With a  $K_i$  value of 20 nM matriptase affinity was moderate and only a 5-fold selectivity over fXa could be determined. More surprising was the fact that compound **63** is also a relatively strong thrombin inhibitor. With a  $K_i$  value 9 nM for

thrombin, the earlier determined restricting effect of the C-terminal 4-*tert*-butylureido-piperidine was completely abolished by the introduction of a second hydroxymethyl group on the N-terminal aryl ring. Compounds **64** and **65** were isolated as byproducts during the final hydrogenation for the synthesis of compound **63**. Compound **65** possesses a 56-fold matriptase selectivity over fXa and represents the most selective monobasic inhibitor of the 3-amidinophenylalanine type so far (regarding fXa). However, matriptase ( $K_i = 14$  nM) and thrombin ( $K_i = 12$  nM) are inhibited equally strong.

Enzyme kinetic measurements have identified the 3-hydroxymethyl substitution at the terminal aryl ring as unsuitable for the design of matriptase inhibitors of the given 3-amidinophenylalanine type containing N-terminal biphenyl-3-sulfonyl groups. Reduced matriptase affinity and selectivity against thrombin was observed. Also, these compounds were more affine to fXa than to matriptase. In contrast, the *para*-position hydroxyl group results in monobasic compounds with excellent matriptase selectivity over thrombin, which is further increased by the addition of a *meta*-fluorine to the terminal phenyl ring. CRAMER and colleagues' recent review discusses the effect of a single hydroxy group addition to compounds [233]. The hydrogen bond including this hydroxy group can be either highly advantageous in terms of binding affinity and selectivity or deleterious if geometric requirements of the interactions are not met. In general, the addition, removal or modification of a single hydroxyl group can cause an affinity shift by several folds. In this thesis, no beneficial effect and only a slightly negative effect on matriptase affinity could be observed but selectivity against thrombin was strongly improved by the 4-hydroxymethyl group. For matriptase, the crystal structure in complex with compound **55** shows an interaction of the hydroxymethyl and the Gln175. Based on the alignment of the crystallographic binding mode of compound **55** and **56** (crystallized in complex with trypsin) also a 3-hydroxymethyl should be in the right distance for the H-bond formation with the side chain of Gln175 (Figure 4-11). To investigate such specific ligand-protein interactions the crystallization of the target protein in complex with the compound of interest is vital. Also, as more structures of similar compounds become available, the quality of further modeling studies will improve.

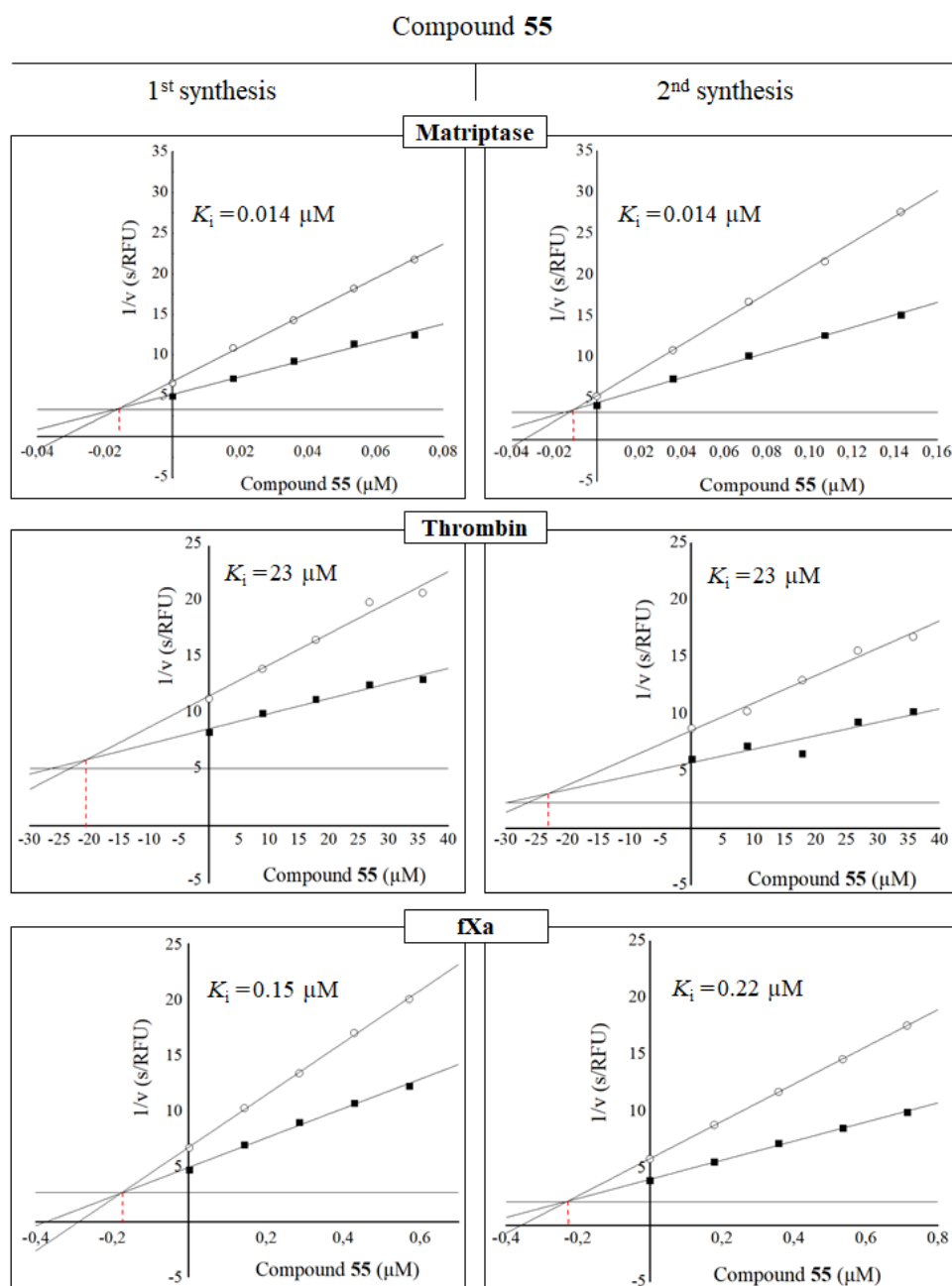
Figure 4-12 shows the alignment of N-termini from compound **17** (yellow), **20** (white) and **55** (yellow), as well as the virtual biphenyl product (pink) that was generated by PINGUI. Whereas the binding mode of compound **17**, **20** and **55** was crystallographically determined (**20** and **55** in matriptase and **17** in thrombin), the positioning of the virtual product was predicted by docking, where the naphthylsulfonyl of compound **20** was used as reference. The 4-hydroxy biphenylsulfonyl of compound **20** and the 2,4-dichloro substituted biphenyl sulfonyl (of compound **17**) have an aligned position in the matriptase S3/4 binding site. However, the

terminal aryl ring shows a slightly varying orientation by what the *meta*-fluorine in case of compound **55** points away from the matriptase binding site. The new matriptase crystal structure from this thesis (6T9T) can serve as a starting point to further improve predictions made by docking with regard to the optimal distance and angle of the hydroxyl group to the protein. Two points that were discussed by CRAMER to be crucial for the affinity and selectivity boost caused by the addition of a single hydroxyl group <sup>[233]</sup>.



**Figure 4-12: Crystallographic binding mode alignment of different N-termini from 3-amidinophenylalanine derivatives in comparison to the virtual compound 40.** Compound **20** (white C-atoms, PDB code 2GV6) and compound **55** (green C-atoms, PDB code 6T9T) were crystallized in matriptase. Compound **17** (yellow C-atoms, PDB code: 4E7R) was crystallized in thrombin. The virtual product **40** was docked into matriptase from PDB code 2GV6. The Gln175 from the matriptase structure of (PDB: 6T9T) is shown in green sticks. The interaction between compound **55** and the Gln175 is shown in a black dashed line and the distance (in Å) is given in a yellow box.

The crystal structures of inhibitors **55** and **56** in complex with matriptase and/or trypsin (Figure 4-10 and Figure 4-11) confirmed the competitive reversible binding mode of these 3-amidinophenylalanine-derived serine protease inhibitors. This binding mode was already determined before during the analysis of the enzyme kinetic measurements. Typical Dixon-plots, indicating a competitive reversible inhibition mechanism, with all three proteases are shown in Figure 4-13.

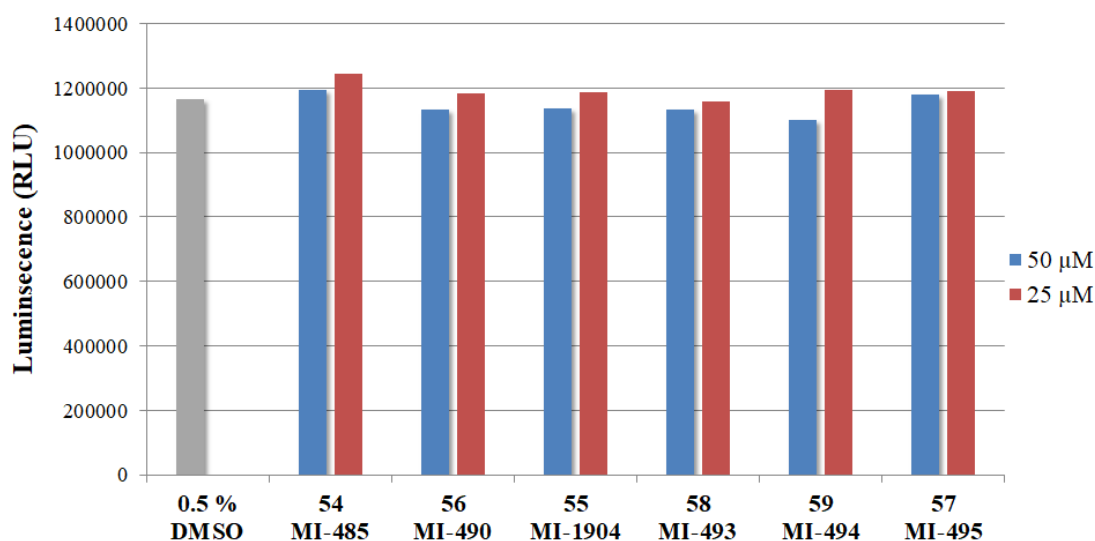


**Figure 4-13: Representative Dixon-plots of the inhibition of matriptase, thrombin and fXa by compound 55.** Measurements for matriptase and fXa were performed with the substrate Mes-dArg-Pro-Arg-AMC (MI-507) at concentrations of 14.3  $\mu\text{M}$  (■) and 7.2  $\mu\text{M}$  (○) for matriptase and 50  $\mu\text{M}$  (■) and 25  $\mu\text{M}$  (○) for fXa. Measurements with thrombin were performed with 10  $\mu\text{M}$  (■) and 5  $\mu\text{M}$  (○) of substrate Tos-Gly-Pro-Arg-AMC (MI-109, Table 7-1). The parallel line to the X-axis represents the constant value for  $1/V_{\text{max}}$  that was determined in an independent measurement of the kinetic constants.

#### 4.2.5. Inhibition of H9N2 influenza virus propagation by matriptase inhibitors

As highlighted in chapter 2, infectivity and replication of influenza virus A subtypes depend on the proper cleavage of its surface glycoprotein HA that mediates binding to sialic acid-containing cell surface receptors and fusion of the viral envelope with the endosomal membrane. Since matriptase is known to cleave hemagglutinin subtype H9<sup>[63]</sup> the more potent dibasic matriptase inhibitors **56-59** (MI-490, MI-493, MI-494 and MI-495) as well as the

monobasic compounds **54** (MI-485) and **55** (MI-1904) were examined for their inhibitory effect on the propagation of H9N2 influenza virus with the R-S-S-R↓ cleavage site in MDCK II cells. As shown in Figure 4-14, no or only a minimal cytotoxic effect of any tested inhibitor on MDCK-II was detected when these cells were treated at a concentration of 50  $\mu$ M. A further reduced, negligible cytotoxicity on MDCK-II cells was observed for all analogues at an inhibitor concentration of 25  $\mu$ M. Therefore, all inhibitors could be used to investigate their inhibitory effect on virus propagation (Figure 4-15).

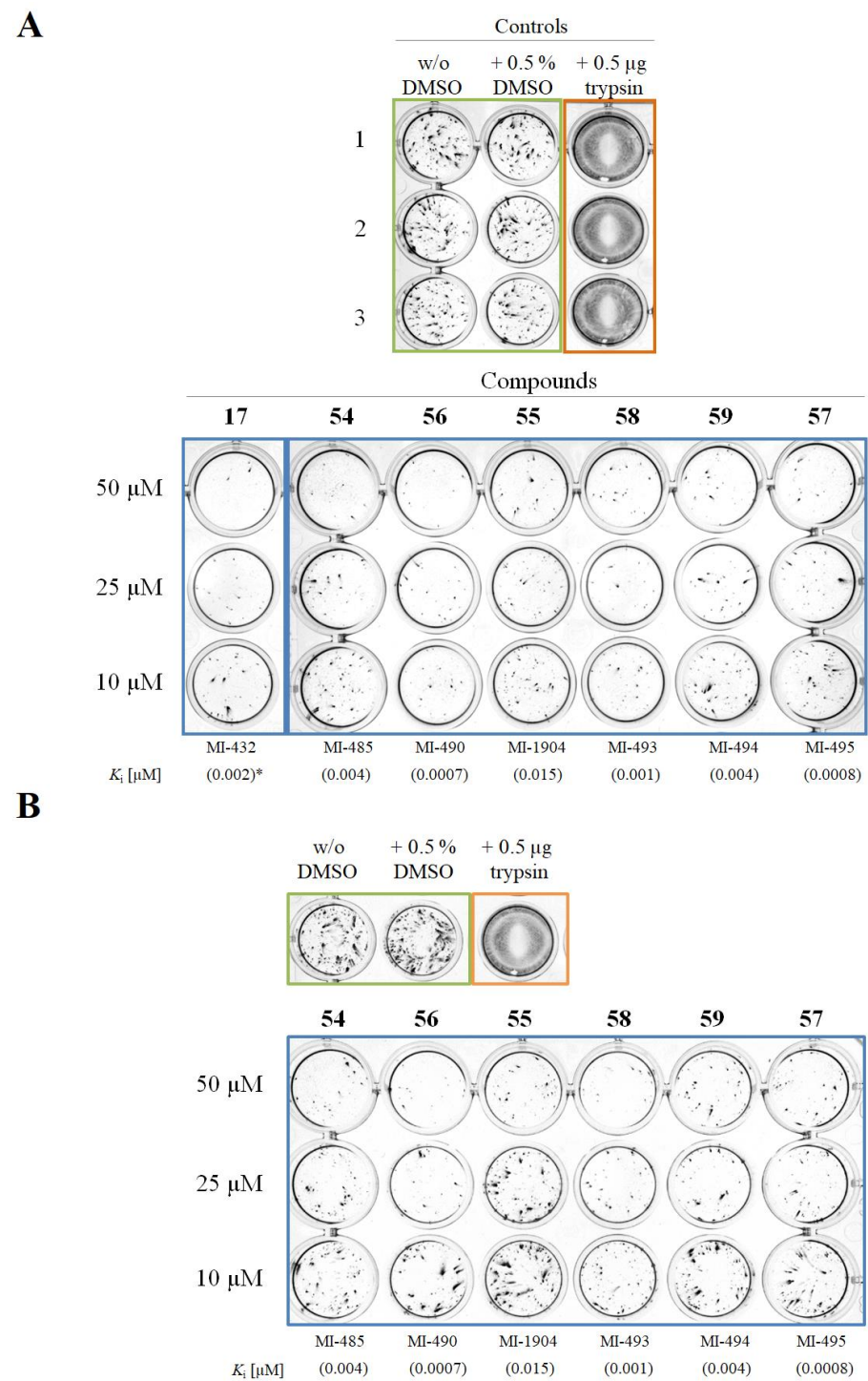


**Figure 4-14: Determination of cytotoxic effects of selected inhibitors on MDCK-II cells.** Cells were incubated 48 h with inhibitor concentrations of 50  $\mu$ M (blue) or 25  $\mu$ M (red). Final DMSO concentration in wells after inhibitor treatment was 0.5 %. As control, cells were treated equally with 0.5 % DMSO only. Absolute luminescence is given on the Y-axis and was determined photometrically using the CellTiter Glo 2.0 kit (Promega) following manufacturer's instructions.

MDCK-II cells were infected with H9N2 (A/quail/Shantou/782/00) at a MOI of 0.005 resulting in moderate virus propagation of approximately 50 % (green box) and 100 % infection after the addition of 0.5  $\mu$ g trypsin (orange box). DMSO at a concentration of 0.5 % has no inhibitory effect on virus propagation in contrast to all inhibitors, which all show a significant inhibition on virus propagation at all three applied concentrations (10, 25 and 50  $\mu$ M). At an inhibitor concentration of 10  $\mu$ M compound **56** with a *meta*-amidine substitution and the *meta*-aminomethyl derivative **58** show the strongest inhibitory effect of the new compounds (Figure 4-14 A), in consistency with their low  $K_i$  values of 700 pM and 1 nM for matriptase, respectively. In wells containing compounds **54**, **55**, **57** and **59** the virus propagation was inhibited to a lesser extent, but no clear differences are determinable even though compound **55** with a  $K_i$  of 15 nM has been identified as an 18-times less potent matriptase inhibitor compared to compound **57** ( $K_i$  = 800 pM) in enzyme kinetic experiments. In an independent repetition of the experiment including the same inhibitors (**54-59**, see Figure 4-15 B) inhibitor

**55** indeed showed lower inhibitory effect than the remaining tested inhibitors including a dose-dependent effect. Whereas the concentration of 50  $\mu\text{M}$  leads to circa 90 % inhibition, the concentration of 25  $\mu\text{M}$  approximately only halves virus propagation (compared to control with or without DMSO (green box)). At 10  $\mu\text{M}$  of inhibitor **55**, only a slight inhibition of virus propagation is visible. The *meta*-aminomethyl substituted compound **58** results again in the strongest virus inhibition at all tested concentrations. More surprisingly, inhibitor **57** seems to only show medium inhibitory effect at 10  $\mu\text{M}$  in both experimental set-ups even though a matriptase  $K_i$  value of 800 pM has been determined in enzyme kinetic experiments (Table 4-9). Also, compound **56** as the most potent matriptase inhibitor ( $K_i = 700$  pM) did not completely inhibit virus propagation at 10  $\mu\text{M}$  in the second set-up (Figure 4-15 B). However, all compounds have an inhibitory effect on virus propagation without showing cytotoxic effect on MDCK-II cells. Compound **17** (structure shown in Table 4-1) served as a positive control in Figure 4-15 A, since this inhibitor has been shown to inhibit H9N2 earlier <sup>[218]</sup>.





**Figure 4-15: Inhibition of H9N2 influenza virus propagation by treatment with matriptase inhibitors in MDCK-II cells.** (A) MDCK II cells were infected with the virus A/quail/Shantou/782/00 (H9N2) at a MOI of 0.005 and incubated with the respective inhibitors at 37 °C. After 24 h the infected cells and comet-like spread of infection were fixed with 4 % paraformaldehyde in minimum essential medium. These cells were incubated with a specific primary antibody against the viral nucleoprotein (NP) of H9N2, followed by conjugation with an HRP-coupled secondary antibody. Infected cells were immunostained using the TrueBlue™ Peroxidase substrate (KPL International Limited, India). \*: Compound including the  $K_i$  value adopted from <sup>[218]</sup> (B) Independent repetition of experiment A. The experiments were conducted by HANNAH LIMBURG in the Institute of Virology (University Marburg).



### 4.3. Summary of matriptase inhibitor design and characterization

In the course of this thesis, new inhibitors of the 3-amidinophenylalanine type were produced and investigated regarding their affinity to matriptase, thrombin and fXa in order to obtain potentially bioavailable affine matriptase inhibitors with negligible affinities to the enzymes fXa and thrombin. For the initial virtual screening of promising compounds, the experimental crystal structure of matriptase in complex with compound **20** (Figure 4-2) was used as a template. Based on experience, thrombin selectivity was achieved by using rather bulky C-terminal residues that target the proximal S2 thrombin binding site, which accessibility is restricted by the thrombin specific 60-loop. With docking in combination with close analysis of the distal S3/4 binding sites of matriptase and fXa, a hydroxymethyl in *meta*-position on the terminal biphenyl-3-sulfonyl group was predicted to be in optimal distance and orientation to interact with the Gln175 side chain on the left side of the distal binding pocket of matriptase. This additional interaction in matriptase, was thought to result in an improved affinity in matriptase but not in fXa containing Phe174 in this position. Hence an increased matriptase selectivity over fXa was suggested.

In total, 18 novels monobasic and 4 dibasic compounds, were subjected to SAR analysis. After identifying 4-*tert*-butylureido-piperidide as suitable C-terminal residue to achieve selectivity against thrombin (Table 4-7) 11 different substitution patterns on the N-terminal phenyl ring were investigated (Table 4-8 and Table 4-9). Thereby, *meta*- and *para*- phenyl-substitutions were compared, including also double substituted N-terminal phenyl rings. The main findings of this analysis were for

#### Matriptase affinity:

- Matriptase prefers dibasic inhibitors over monobasic compounds.
- Matriptase prefers *meta*-positioned amidino- and aminomethyl-substitutions (dibasic compounds **56** and **58**) on the N-terminal ring of its biphenyl-3-sulfonyl group.
- Matriptase prefers *para*-positioned hydroxymethyl-groups (**55** and **61**) compared to *meta*-hydroxymethyl substitutions (**60** and **62**) in the case of monobasic compounds.
- No improvement in matriptase affinity with double substituted aryl rings (compounds **62-65**) was found.
- The bis-amidino compounds **56** and **57** show the highest matriptase affinity (700 – 800 pM).

Matriptase selectivity over thrombin:

- Higher selectivity with *meta*-positioned amidino- and aminomethyl-groups (dibasic compounds **56** and **58**).
- Higher selectivity with *para*-positioned hydroxymethyl-groups (monobasic compound **61**) that is further increased by an additional *meta*-positioned fluorine (compound **55**).
- Complete abolishment of selectivity with the introduction of a second hydroxymethyl group (compound **63**) or two methyl groups in *meta*- and *para*-position (compound **65**).
- Dibasic compound **56** shows highest selectivity against thrombin (3700-fold).
- Compound **55** is the most thrombin selective monobasic compound of this series (1533-fold stronger matriptase inhibition).

Matriptase selectivity over fXa:

- Higher selectivity in the case of *para*-substituted aryl rings of dibasic as well as monobasic compounds (**55**, **57**, **59** and **61**).
- Only low till moderate fXa selectivity was found with monobasic as well as dibasic compounds.
- Dibasic compound **57** shows highest selectivity against fXa (100-fold).
- Compound **65** is the most fXa selective monobasic compound (56-fold stronger matriptase affinity).

The binding mode of two inhibitors (compounds **55** and **56**) was crystallographically determined in matriptase and/or trypsin and showed the inhibitor typical interactions in both proteases. In addition, the amidine in *meta*- (**56**) and the hydroxymethyl in *para*-position (**55**) on the N-terminal biphenyl-3-sulfonyl group interact with the side chain of Gln175 that is present in matriptase and trypsin. However, the additional interaction in the distal binding site did not result in the predicted selectivity increase of matriptase against fXa.

With compound **55** being investigated in complex with matriptase and trypsin, the comparison of the binding mode revealed that trypsin is a suitable substitute for matriptase when interactions of the compound within the distal S3/4 binding site of matriptase are in focus. In contrast, the C-terminal substituent on the piperidide moiety of compound **55** showed a slightly different orientation in the proximal binding site of the two proteases. The matriptase crystal structure of compound **55** is the first structure in complex with a compound of the 3-amidinophenylalanine type since 2006 and the first in general with this inhibitor type where the S3/4 binding pocket is occupied by the compound's biphenyl-3-sulfonyl moiety. So far, the orientation of the biphenyl-3-sulfonyl including possible interactions with amino acids of the distal matriptase binding site was predicted based on crystal structures in complex with the

related proteases thrombin or trypsin and subsequent modeling with matriptase. These models are reliable predictions with high accuracy, however as also shown in this thesis (comparison of binding mode of compound **55** in matriptase and trypsin) slight differences are present. Based on the present new matriptase crystal structure the binding modes of similar inhibitors can now be even better predicted and this new structure should improve the quality and precision of future docking studies.

In addition, selected monobasic as well as dibasic compounds demonstrated strong suppression of H9N2 influenza virus replication in a matriptase-expressing MDCK II cell model. Especially the monobasic compounds are suitable for the application in further cell culture and animal studies as a potential prodrug, as it is well established for various benzamidino-drugs.



## 5. Characterization of truncated and mutated TMPRSS2 variants

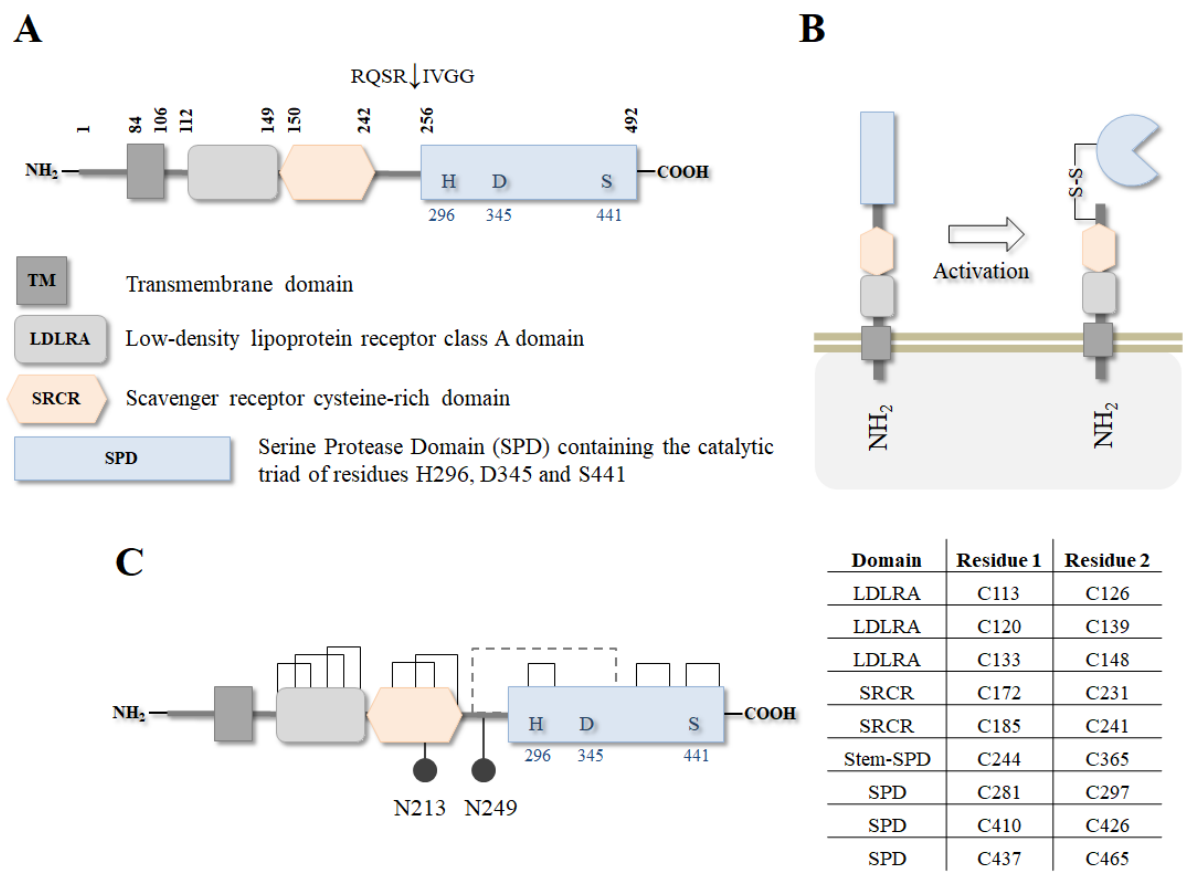
### 5.1. Introduction to TMPRSS2 characterization

The type II transmembrane protease TMPRSS2 (Transmembrane protease serine S1 member 2) is a member of the Hepsin/TMPRSS subfamily (Figure 2-7). The mouse analogue is known as epitheliasin, which is also recommended as the official name by the peptide database MEROPS (S01.247). In the following, TMPRSS2 is used to describe the gene (*italic letters*) as well as the protein (regular letters).

#### 5.1.1. *TMPRSS2* gene organization and maturation

The *TMPRSS2* gene was first described in 1997 and is located on the smallest human chromosome number 21 <sup>[234]</sup>. The protein is synthesized as a single-chain 70 kDa inactive precursor protein of 492 aa with a short N-terminal cytoplasmic tail (residues 1-83), a transmembrane domain (aa 84-106) and a C-terminal extracellular portion (aa 107 - 492) (Figure 5-1 A). Besides the SPD (residues 256-492) with the catalytic triad of His296, Asp345 and Ser441, the extracellular portion comprises of one LDLRA domain (aa 112-149) and one SRCR domain (aa 150-242). These two latter domains form the stem region. The process of TMPRSS2 maturation involves several posttranslational modifications and in order to be active, TMPRSS2 needs to be cleaved after the highly conserved arginine in position 255. Based on similarity, 18 cysteines have been identified to form nine disulfide bonds, three of which are located within the SPD and one that keeps the SPD connected to the stem region post cleavage after Arg255 (Figure 5-1 B and C). TMPRSS2 has an additional cysteine residue within the SPD at position 379. In hepsin, two additional cysteines were found, one which is discussed to form a fourth disulfide bond and another unpaired cysteine that was suggested to assist with protein or ligand interactions <sup>[235]</sup>. The function of the additional Cys379 of TMPRSS2 is unclear. Based on sequence analysis, TMPRSS2 harbors at least two *N*-glycosylation sites at residues N213 and N249 (Uniprot: O15393). The activation cleavage after residue R255 is a prerequisite for proteolytic TMPRSS2 activity and occurs to be autocatalytic since the single site mutagenesis of the active site serine to alanine (S441A) as well as the arginine 255 to glutamine (R255Q) in the cleavage site result both in the loss of cleavage capacity <sup>[76]</sup>. AFAR and colleagues investigated TMPRSS2 cleavage in western blot experiments using polyclonal  $\alpha$ -TMPRSS2 primary antibodies. The wild-type TMPRSS2 showed an approximately 32 kDa signal in addition to the 70 kDa precursor but the S441A and R255Q mutated TMPRSS2 samples did not show this 32 kDa protein signal, which was referred to as protease domain.

More recent data suggest a slightly smaller size of the protease domain of approximately 27 kDa [236, 237] or 24 kDa [238]. The calculated size of the TMPRSS2 SPD is about approximately 27 kDa (calculated with ExPASyProtParam). Slight differences might result based on the usage of different antibodies or protein size markers. So far, no experimental structural data for TMPRSS2 is available.



**Figure 5-1: Schematic overview of the mosaic-like TMPRSS2 zymogen structure and its maturation. A:** TMPRSS2 comprises of an N-terminal cytoplasmic tail, a transmembrane domain and a C-terminal extracellular portion including one LDLRA and one SRCR domain as well as the SPD. TMPRSS2 is expressed as inactive pro-enzyme that requires cleavage for activation. The arrow (↓) indicates the autocatalytic cleavage site after arginine R255. **B:** Cleavage of TMPRSS2 zymogen. After the cleavage, the SPD remains linked to the stem region through a conserved disulfide bond. **C:** Disulfide pattern of TMPRSS2, which harbors 18 cysteine residues forming 9 disulfide bonds including one linking the SPD to the stem region post cleavage (dashed line). Three disulfide bonds were identified by similarities within the LDLRA and SPD each and two within the SRCR. In addition, two N-glycosylation sites were found by sequencing analysis (asparagine residues 213 and 249 (●)).

Human *TMPRSS2* mRNA is mainly expressed in epithelial cells of several tissues including mainly the prostate and colon but also breast, kidney, small intestine, pancreas, ovary, salivary gland, stomach, and lung [77, 239-242]. The physiological function of TMPRSS2 is still unclear. Protease deficient mice do not show any phenotypical alterations in comparison to their wild type littermates, suggesting that TMPRSS2 functionality is not essential for mice vitality [243]. Functional expression studies in *Xenopus* oocytes demonstrated that TMPRSS2 might

contribute to the regulation of the airway surface liquid (ASL) volume by decreasing the levels of epithelial sodium channels (ENaCs) in the human airway epithelium through proteolytic cleavage [244]. In addition, cardiac myocytes have shown to express *TMPRSS2*, which might imply that *TMPRSS2* contributes to influenza-associated myocarditis [245]. LUCAS and colleagues identified *TMPRSS2* as a potential participant in proteolytic cascades with relevance in the physiological function of the prostate [246].

### 5.1.2. *TMPRSS2* and cancer

*TMPRSS2* expression had been detected at higher levels in prostate cancer cells compared to healthy prostate cells. Unlike other TTSPs, the *TMPRSS2* transcription in prostate cells is regulated by androgenic ligands and the androgen receptor (AR). In most cases, the AR-regulated *TMPRSS2* upstream region was fused with the oncogenic *ETS-related gene (ERG)*, a member of the erythroblast transformation-specific (ETS) gene family [76, 247]. *ERG* overexpression, as an effect of gene fusion, was discussed to impact the AR-signaling and possibly contribute to the development of an androgen-independency in prostate cancer cells [248]. Based on this high expression in prostate cancer, *TMPRSS2* was suggested to serve as a potential marker or therapeutic target for prostate cancer [249]. The *TMPRSS2:ERG* fusion is unlikely to result in proteolytic function as the product does not encode the SPD of *TMPRSS2* [250]. However, *TMPRSS2* activity in soluble or membrane-anchored form can be linked to diseases, including cancer and others, by processing disease-related substrates. Cell-based expression studies revealed PAR-2 as potential *TMPRSS2* substrate, although it might be also activated by other trypsin-like serine proteases including matriptase. PAR-2 is found to be up-regulated in many tumor tissues including prostate [251]. Therefore, also *TMPRSS2* activity is discussed as a potential candidate to modulate aspects of tumor cell growth and metastasis. CHUN-JUNG KO and colleagues discussed *TMPRSS2* to promote prostate cancer growth and metastasis via matriptase activation and degradation of the extracellular matrix (ECM) proteins nidogen-1 and laminin  $\beta$ -1 [237]. In their study, *TMPRSS2* expression correlated with higher levels of matriptase activation in prostate cancer tissues. *TMPRSS2* was thereby shown to directly activate matriptase but also to possibly shed the endogenous matriptase inhibitor HAI-1, and thus further enhancing matriptase activity. Since HAI-1 inhibits various serine proteases including, for example, hepsin and prostasin [252], the *TMPRSS2* induced shedding may also cause activation of these pericellular proteases. In general, matriptase activity itself is linked to many forms of cancer and other diseases as already discussed in section 3.1 of this thesis. Alike matriptase, also *TMPRSS2* had been identified as a potent activator of the hepatocyte growth factor zymogen pro-HGF as a result of positional scanning substrate combinatorial library (PS-

SCL)<sup>[246]</sup>. HGF, also named as scatter factor, interacts with the c-Met receptor tyrosine kinase and therefore influences well-described developmental programs that are linked to pathological processes including carcinogenesis and metastasis<sup>[253]</sup>. In addition, the analysis revealed the tissue-type plasminogen activator (tPA) and kallikrein-2 (KLK-2) as potential TMPRSS2 substrates<sup>[246]</sup>. Furthermore, TMPRSS2 activity had been connected to certain infectious diseases caused by influenza viruses<sup>[58, 62, 254]</sup>, corona viruses<sup>[245, 255, 256]</sup>, the parainfluenza virus<sup>[257]</sup>, the human metapneumovirus<sup>[258]</sup> and the hepatitis C virus<sup>[259]</sup>. TMPRSS2, such as other host proteases, is involved in virus spread by activating the respective fusion proteins on the virus surface, which is an essential step in the virus replication cycle.

### 5.1.3. The Role of TMPRSS2 in influenza virus activation

Cleavage of the influenza-A-virus (IAV) surface glycoprotein HA by TMPRSS2 has been shown in several *in vitro*<sup>[58, 151, 236, 254, 260]</sup> and *in vivo*<sup>[60-62]</sup> studies. GALLOWAY and colleagues investigated TMPRSS2 cleavage activity regarding all 16 avian HA subtypes *in vitro* and found a broad cleavage activity for almost every HA subtype except for H8 and H12<sup>[151]</sup>. This included also the HPAIV subtypes H5 and H7 that contain multibasic cleavage sites and are cleaved *in vivo* by furin or furin-like proprotein convertases within the TGN<sup>[153, 261]</sup>. Similar to furin, TMPRSS2 was found to cleave IAV HA exclusively in the intracellular space *in vitro*<sup>[262, 263]</sup>. *In vivo* studies including *tmprss2* <sup>-/-</sup> knockout mice showed a TMPRSS2 dependency of influenza A subtype H1N1 virus spread and pathogenesis. Due to the lack of HA cleavage, *tmprss2* <sup>-/-</sup> knockout mice were protected from viral spread and pathogenesis upon H1N1 infection. Further studies confirmed these findings and reported also a TMPRSS2 dependency of H7N9 viruses<sup>[60-62]</sup>. In contrast to H1N1 and H3N2 that circulate in humans, the avian H7N9 is only transmitted after direct contact with infected animals. However, over 1500 confirmed cases of severe influenza infections with a mortality rate of 39 % were reported<sup>[264]</sup>. Infection of *tmprss2* <sup>-/-</sup> knockout mice with the H3N2 virus still resulted in an efficient virus spread into the pulmonary tissue implying the involvement of another protease in H3-HA activation of the mouse lung. Those mice succumbed to infection. In a recent study including H3N2 IAV, LIMBURG and colleagues identified TMPRSS2 as a major HA-activating protease in primary human airway cells. Efficient knockdown of TMPRSS2 prevented proteolytic activation and multiplication of H1N1pdm, H7N9 and H3N2 IAV in primary human bronchial epithelial cells (HBEC) and primary type II alveolar epithelial cells (AECII) suggesting that TMPRSS2 is essential for activation and multiplication of these IAV in the human lower respiratory tract<sup>[254]</sup>. Currently also H10N8 IAV is under surveillance by the WHO based on its potential to infect humans<sup>[265]</sup>. Since the first detection in humans 2013<sup>[266]</sup>, only three cases have been



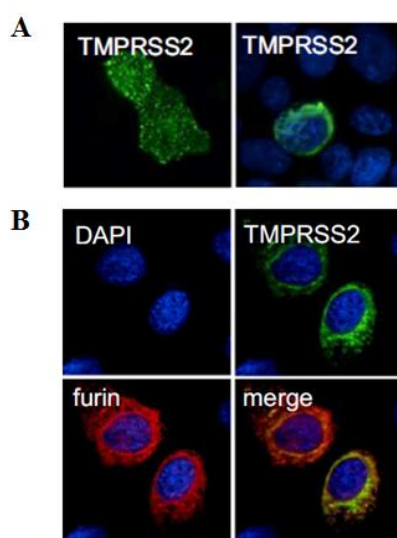
reported, of which two had a lethal outcome. In the *in vitro* study of STRAUS and WHITTAKER, the H10-HA subtype was not or only poorly cleaved by the tested human respiratory tract proteases. Minor cleavage was only observed by matriptase in addition to full cleavage of trypsin <sup>[267]</sup>. TMPRSS2 was not included in this study. However, very recent investigations with *tmprss2* <sup>-/-</sup> deficient mice implied that TMPRSS2 activity is also essential for *in vivo* replication and pathogenesis of H10 IAV, since knockout mice did not show viral replication and also reduced inflammatory lesions in the lung compared to wild-type mice <sup>[152]</sup>. H10-HA from H10N7 was investigated in the *in vitro* study of GALLOWAY and colleagues and showed an aberrant TMPRSS2-dependent cleavage profile compared to the cleavage by trypsin. Slightly smaller HA1 and HA2 subunits than expected and multiple cleavage products were observed in western blots. The authors discussed that TMPRSS2 might cleave at multiple sites or only a specific glycosylated H10-HA form <sup>[151]</sup>. TMPRSS2 is one of the most discussed TTSPs in IAV infections and therefore, represents an interesting target for anti-influenza drug development. However, besides TMPRSS2, a broad range of serine proteases represent potential candidates for activation of the IAV HA and cleavage of the precursor protein HA<sub>0</sub> can take place at different stages of the replication cycle or even extracellularly <sup>[268, 269]</sup>.

Among many other serine proteases, also TMPRSS2 is likely to be involved in different activation cascades. For example, it has been shown that TMPRSS2 can activate matriptase <sup>[237]</sup>. As discussed in section 3.1.6, matriptase cleaves IAV H9N2 HA, resulting in efficient virus spread <sup>[63]</sup>. The involvement of TMPRSS2 in these protease cascades remains unclear since the availability of data on TMPRSS2 compared to other TTSPs is still rather low. Nevertheless, the cleavage activity towards matriptase implies also a possible indirect role in IAV activation.

#### 5.1.4. Subcellular TMPRSS2 localization

Since *TMPRSS2* mRNA transcription and gene expression is increased in prostate cancer <sup>[76, 240, 242, 270]</sup>, localization of TMPRSS2 has been mainly studied in normal and cancerogenic prostate cells. Based on immunohistochemistry (IHC) studies, LUCAS and colleagues reported TMPRSS2 mislocalization in high-grade cancer; TMPRSS2 loses membrane polarity and was prominently detected in the cytoplasm in contrast to normal prostate cells, where TMPRSS2 is expressed along the apical membrane of luminal secretory epithelial cells <sup>[270]</sup>. Nevertheless, these studies did not reveal a clear subcellular localization and the authors discussed that TMPRSS2 may be localized in intracellular vesicles or organelles of the secretory pathway. AFAR and colleagues detected shedded soluble TMPRSS2 in cell culture medium of androgen-stimulated prostate cancer cells (LNCaP cells), in the prostate lumen and in seminal fluid <sup>[76]</sup>.

BÖTCHER-FRIEBERTSHÄUSER and colleagues studied subcellular localization of *TMPRSS2* expressing MDCK-H cells <sup>[262]</sup>. The *TMPRSS2* zymogen, as well as the mature membrane-bound *TMPRSS2* were found on the surface of the plasma membrane and marginally as soluble form in the supernatant. However, only *TMPRSS2* within the cell showed proteolytic activation of HA, suggesting that *TMPRSS2* only supports the cleavage of newly synthesized HA. Indirect immunofluorescence (IF) with protease specific primary antibodies gave further evidence that *TMPRSS2* is located on the cell surface of non-permeabilized MDCK cells (Figure 5-2 A, left panel) and is localized in the TGN of permeabilized MDCK cells (Figure 5-2 B, right panel). In addition, *TMPRSS2* co-localization with furin, a TGN resident protease, was shown in Huh-7 cells (Figure 5-2 B) <sup>[271]</sup>.



**Figure 5-2: Determination of subcellular localization by indirect immunofluorescence.** Proteases were stained using protease-specific primary antibodies and cell nuclei were counterstained with DAPI. **A:** Subcellular *TMPRSS2* localization on the cell surface (non-permeabilized cells; left panel) and within MDCK cells (permeabilized cells; right panels). *TMPRSS2* (in green) was stained using a *TMPRSS2*-specific primary antibody and FITC-conjugated secondary antibody. **B:** Co-localization of *TMPRSS2* and furin (in red) after transient expression in Huh-7 cells. Cells were permeabilized and stained using *TMPRSS2*- and furin-specific primary antibodies, respectively and FITC-(*TMPRSS2*) or TRITC-(furin) conjugated secondary antibodies. Figure is taken and modified from literature <sup>[271]</sup>.

### 5.1.5. Starting approach for isolation of active *TMPRSS2* from mammalian cells

*TMPRSS2* is anchored in the plasma membrane even after proteolytic activation cleavage, due to a conserved disulfide bond that connects the SPD with the SRCR domain (Figure 5-1 B). Although shedding may occur, *TMPRSS2* activity is thought to be limited to the intracellular space in the case of non-prostate cancer cells. The isolation of membrane-bound proteins from mammalian cells, more precisely from the TGN in case of *TMPRSS2*, is usually accomplished by adding detergent that solubilizes the biomembrane and forms a soluble complex with the membrane. The solubilization is one of the most critical steps and has to be carefully optimized

to avoid protein loss and inactivation. Protein denaturation and/or aggregation are frequently encountered. Earlier isolation attempts of TMPRSS2 from MDCK-TMPRSS2 cells were not successful in FRIEBERTSHÄUSER lab; only inactive TMPRSS2 could be purified. To circumvent the harsh treatment during solubilization, in this thesis truncated TMPRSS2 constructs without the transmembrane anchoring domain were designed in order to obtain soluble TMPRSS2. The transmembrane domain of Type-II transmembrane proteins also harbors a specific signal peptide, regulating the translocation of the newly transcribed protein into the secretory pathway, where essential post-translational modifications take place. Therefore, the removal of the transmembrane domain, including the signal peptide requires the introduction of an artificial signal peptide to cover the functional loss and mediate the translocation into the secretory pathway post translation.

KNAPPE and colleagues designed soluble variants of the related TTSP corin, by cloning its extracellular portion into the pSecTag2 HygroB plasmid <sup>[84]</sup>. This plasmid possesses the murine Ig kappa secretion peptide (Igκ-SP) upstream of the multiple cloning site to facilitate the secretion of the fused protein <sup>[272]</sup>. Soluble and active recombinant 6xHis-tagged corin had been isolated from conditioned cell culture medium of transfected 293T cells. In addition to soluble wt-corin, they created a corin variant where the autocatalytic cleavage site was substituted with an EK cleavage site (DDDDK) to purify soluble EK-corin in a less sensible inactive form. Functional soluble corin could be recovered following an activation step under controlled conditions by adding recombinant EK. To study TMPRSS2 in prostate cancer, CHEN and colleagues generated a specific monoclonal TMPRSS2 antibody using a transmembrane domain truncated inactive Igκ-SP containing R255Q mutated TMPRSS2 variant. The *TMPPRSS2* gene was fused downstream to the Igκ-SP in the MCS of the pSecTag2 HygroB plasmid. Stably transfected 293T clones were cultivated and FLAG-R255Q-TMPRSS2-6xHIS was isolated from conditioned medium and purified via IMAC and gel filtration <sup>[273]</sup>. These data suggest that the Igκ-SP is generally suitable to produce soluble TTSPs from 293T cells and that proteins can be secreted to the cell culture medium. In this thesis, TMPRSS2 without its anchoring domain should be produced and tested for the applicability in large expression experiments.

The stem region of TTSPs is continually discussed to play an important role in proteolytic activity. For example, removing the TMPRSS2 LDLRA domain using peptide-conjugated phosphorodiamidate morpholino oligomers (PPMOs) that interfere with *TMPPRSS2* pre-mRNA resulted in an inactive 63 kDa TMPRSS2 zymogen after expression in A549 or Calu-3 cells. Compared to the full-length TMPRSS2, no SPD signal at approximately 32 kDa was detected implying that the LDLRA domain is essential for the enzymatic activity of TMPRSS2 <sup>[236]</sup>.

Based on these data, various *TMPRSS2* constructs with N-terminally truncated extracellular portions and the Igk-SP at the N-terminus should be designed, produced and characterized. For purification purposes, a 6xHis-tag or a FLAG tag should be introduced C-terminally. Another approach included the single-site mutation of the cysteine residue C365 (C122 based on chymotrypsinogen numbering) within the full-length *TMPRSS2* gene. This highly conserved cysteine among TTSPs is involved in the disulfide bond formation that keeps the SPD connected to the stem region after the autocatalytic activation cleavage (Figure 5-1 B). By removing this cysteine, hence hampering the disulfide bond formation, the proteolytic active SPD is thought to be isolable from cell lysates. Both strategies follow the goal to obtain membrane-unbound correctly processed soluble and active *TMPRSS2* that can be purified from the cell interior or the cell culture medium. In parallel, a third approach focused on the isolation of the membrane-bound full-length *TMPRSS2* from MDCK-*TMPRSS2* cells by immunoprecipitation.

#### **5.1.6. Goals of *TMPRSS2* characterization**

Since its discovery in 1997, *TMPPRS2* has been identified as a potential marker for prostate cancer. In addition, *TMPRSS2* is an interesting target of host factor directed drug discovery for influenza treatment, since *TMPRSS2* is involved or even responsible for HA activation of several HA subtypes. Nevertheless, compared to other well-characterized TTSPs such as EK or matriptase, still little is known about *TMPRSS2* including its physiological function. Up to date, no *TMPPRS2* crystal structure is available, nor large substrate/ inhibitor studies were conducted by comparing enzyme kinetic parameters. To use the potential of *TMPRSS2* as a target for drug discovery, a profound functional and structural characterization of the protease is vital. The main goal of this part was to obtain new details on *TMPRSS2* functionality in order to identify and develop a system to isolate proper amounts of active recombinant *TMPRSS2* for functional (and structural) studies. In detail, these included the following steps:

- the design and production of truncated and/or mutated *TMPRSS2* constructs (plasmids),
- the transient expression of these *TMPRSS2* variants in mammalian cell culture including their functional characterization,
- the subcellular localization of *TMPRSS2* wild type compared to truncated variants,
- and the isolation of functional *TMPRSS2* for further studies.

## 5.2. Results and discussion of TMPRSS2 characterization

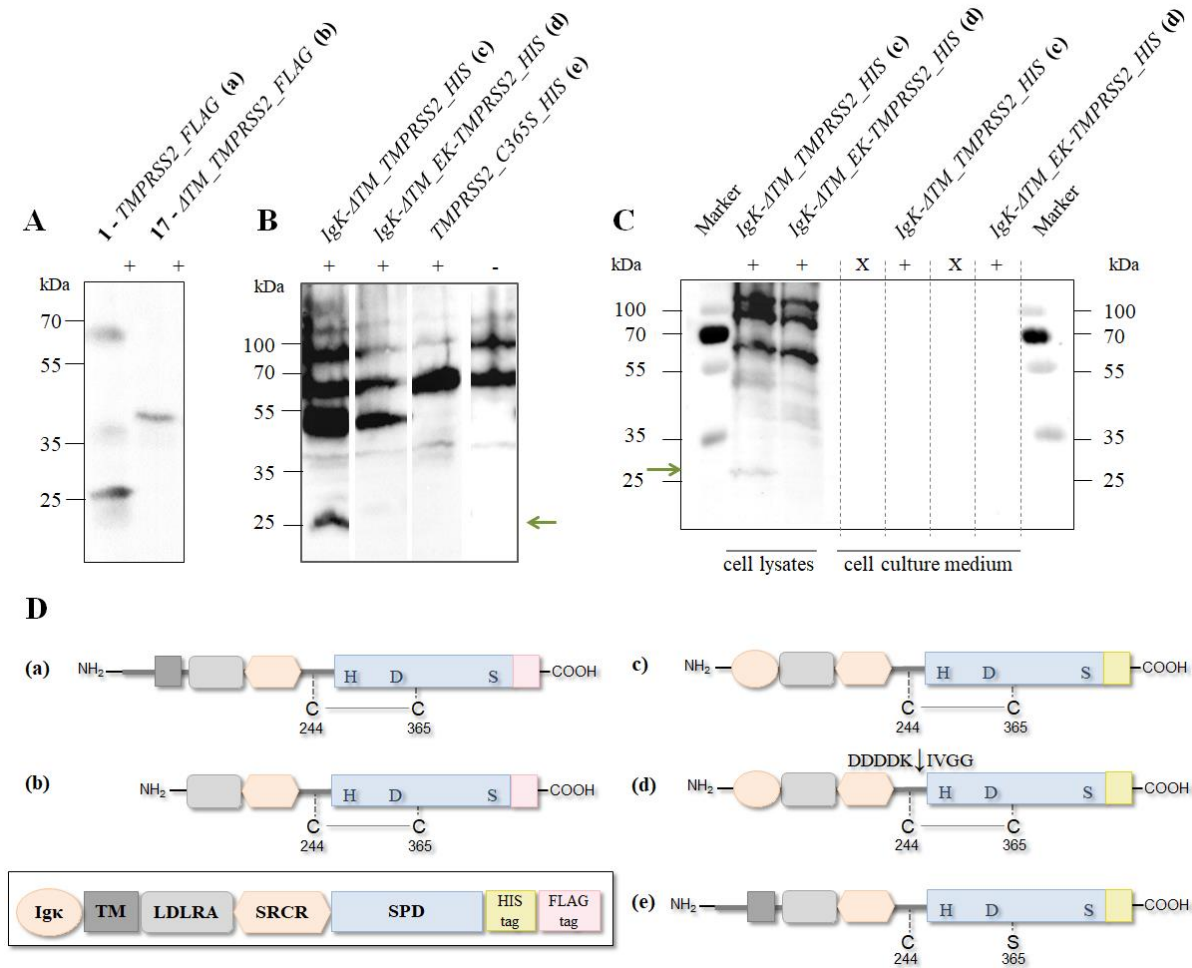
Mammalian cell culture-based experiments were performed in the laboratories of PROF. FRIEBERTSHÄUSER in the institute of virology (department of medicine, University Marburg). The cloning and bacterial based work were performed in the laboratories of PROF. KLEBE as part of the institute of the pharmaceutical chemistry (department of pharmacy).

### 5.2.1. *TMPRSS2* expression in *E. coli*

In parallel to C122S-matriptase (see chapter 3), a C122S-TMPRSS2 variant (numbering based on chymotrypsinogen) containing an EK cleavage site upstream of amino acid residue Ile16 had been created and deployed in recombinant expression experiments using *E. coli*. In the case of TMPRSS2, the free cysteine residue C136 was additionally mutated to a serine (C136S). Although TMPRSS2 could generally be produced using *E. coli* cells, no soluble active TMPRSS2 could be recovered after the activation step with recombinant EK. Optimization attempts focused mainly on the refolding conditions, such as the refolding method, temperature, pH and buffer compositions, but also on the use of different *TMPRSS2* constructs but did not result in active protein. Whereas for other TTSPs cloning of solely the SPD suffices to obtain functional protein, TMPRSS2 appears to be more challenging. Therefore, *TMPRSS2* constructs containing one or two domains upstream the SPD were tested additionally (in collaboration with the group of VIKTOR MAGDOLEN, TU Munich) without success. No active TMPRSS2 could be recovered from *E. coli* during this thesis (data not shown). Consequently, mammalian cell culture systems were chosen for further experiments and different TMPRSS2 variants were created and tested. In the following, chymotrypsinogen numbering will not apply, and amino acids will be numbered from 1-492 based on full-length TMPRSS2 (Figure 5-1).

### 5.2.2. Cloning and transient expression of a first set of *TMPRSS2* constructs

In a first round, different truncated or mutated *TMPRSS2* constructs were produced. The truncated *TMPRSS2* gene was therefore amplified by PCR and subcloned into the pSecTag2 Hygro B vector. In a second PCR-step, the final construct containing the Igk-SP was amplified using the pSecTag2 HygroB-*TMPRSS2* plasmid as a template. Finally, the respective Igk-*TMPRSS2* gene was subcloned into the mammalian expression plasmid pCAGGS. Site-directed mutagenesis was applied to generate single residue substitutions in the *TMPRSS2* gene. Confirmed by sequencing, positive constructs were then transiently transfected into 293T cells, and cell lysates were analyzed by SDS-PAGE and western blotting using specific primary and HRP-conjugated secondary antibodies 48 h post transfection. The results including the construct structure are given in Figure 5-3 A-D.



**Figure 5-3: Autocatalytic activation of truncated or mutated TMPRSS2 variants.** Different *TMPRSS2* constructs were expressed in 293T cells. Cell lysates (A+B) and cell culture medium (C) were analyzed via SDS-PAGE and subsequent blotting to a PVDF membrane. Proteins were visualized with protein or tag specific primary antibodies and HRP-conjugated secondary antibodies. Samples showing a 27 kDa signal are considered as "active" (indicated by a green arrow). Igκ: murine Ig-kappa secretion peptide, TM: transmembrane domain, EK: enterokinase cleavage site (DDDDK↓), C365S: site-directed mutation of cysteine 365 to serine. "+" implies transfection of cells with *TMPRSS2* plasmids and "-" indicates non-transfected cells, x indicates empty gel slots. A: *TMPRSS2* constructs containing a C-terminal FLAG-tag. *TMPRSS2*\_FLAG represents the full-length *TMPRSS2* without any modification in the *TMPRSS2* gene and is considered as a positive control. Proteins were visualized using a polyclonal α-*TMPRSS2* primary antibody and the respective secondary antibodies. B+C: *TMPRSS2* constructs containing a C-terminal 6xHis-tag (HIS) were immunostained with a primary α-His antibody. D: Schematic construct overview.

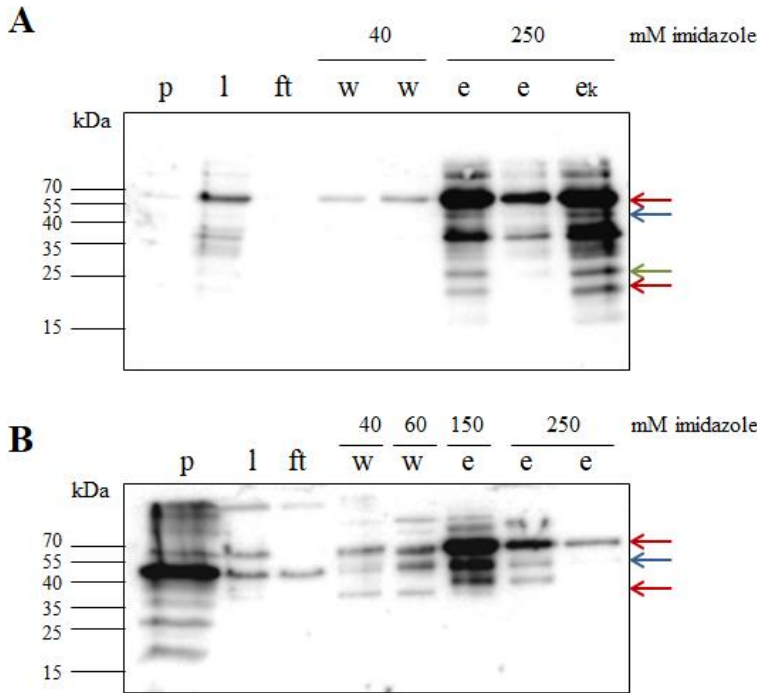
As shown in Figure 5-3 A-C, *TMPRSS2* variants were successfully expressed in 293T cells after transient transfection with the respective plasmids. No *TMPRSS2* was detectable in conditioned cell culture medium (Figure 5-3 C). The full-length *TMPRSS2* with a C-terminal FLAG-tag ((a) in Figure 5-3 A and D) served as a positive control since it has shown proteolytic cleavage of HA in co-expression experiments *in vitro* [58, 236, 262]. The appearance of a 27 kDa signal in addition to the zymogen form (65 kDa - 70 kDa in case of the full-length *TMPRSS2*) is linked to enzyme activity. Previous studies have shown, that intrinsic autocatalytic activity is required

for TMPRSS2 activity in expression experiments with 293T cells. A S441A mutated TMPRSS2 did neither show a 27 kDa signal in western blot experiments nor a detectable cleavage of HA in co-expression experiments [58, 76]. In the following, the presence of a 27 kDa signal will be used as a marker to evaluate the different TMPRSS2 variants regarding their potential proteolytic activity.

Besides the positive control, only the sample of *Igκ\_ΔTM\_TMPRSS2\_HIS* shows a 27 kDa protein signal ((c) in Figure 5-3 B and C). No autocatalytic cleavage activity could be observed in the case of the constructs *ΔTM\_TMPRSS2\_FLAG* (b), *Igκ\_ΔTM\_EK\_TMPRSS2\_HIS* (d) and *TMPRSS2\_C365S\_HIS* (e). The presence of the 27 kDa signal in the case of construct *Igκ\_ΔTM\_TMPRSS2\_HIS* paired with its absence in case of construct *ΔTM\_TMPRSS2\_FLAG* suggest that the entry into the secretory pathway is a decisive factor for TMPRSS2 activity. Without the Igκ-SP, TMPRSS2 maturation of truncated variants including the autocatalytic activation is not possible in 293T cells. Interestingly, the single site mutation of residue cysteine 365 to serine results in the loss of TMPRSS2 autocatalytic activity (construct (e) in Figure 5-3). Based on sequence alignment of TTSPs, this cysteine residue C365 is the conserved cysteine (C122 after chymotrypsinogen numbering) forming the disulfide bond between the SPD and the stem region of TMPRSS2. The loss of activity by replacing the cysteine residue with serine and hence inhibiting the formation of this disulfide bond might imply that this disulfide bond has a crucial structural function for TMPRSS2 activity. Nevertheless, one prominent background signal, caused by the used primary α-His antibody, overlaps with the calculated size of full-length TMPRSS2 and the general expressibility of the construct has to be further confirmed. In the case of *Igκ\_ΔTM\_EK\_TMPRSS2\_HIS* (d) no TMPRSS2 activity was expected. The autocatalytic cleavage site was altered to an EK cleavage site to create a secreted TMPRSS2 variant that can be activated upon EK addition under defined conditions.

In order to isolate substantial TMPRSS2 amounts, scale-up experiments were performed with constructs *Igκ\_ΔTM\_TMPRSS2\_HIS* (c) and *Igκ\_ΔTM\_EK\_TMPRSS2\_HIS* (d) (Figure 5-4 A and B). The respective plasmids were transiently expressed in 293T cells and protein was isolated via IMAC in batch purification experiments from crude cell lysates. Following standard purification protocols, proteins were eluted with imidazole. Figure 5-4 A shows that the purification of truncated TMPRSS2 from construct *Igκ\_ΔTM\_TMPRSS2\_HIS* (c) was not successful. Even though a signal at 27 kDa is visible (green arrow), the elution fraction 'e' shows multiple bands with two non-TMPRSS2 related background signals at 60 kDa and 37 kDa (red arrows) being the most prominent ones. The calculated size of both truncated TMPRSS2 variants is about 50 kDa, which is indicated by the blue arrow. Also, inactive TMPRSS2 with the EK cleavage site (Figure 5-4 B) could not be purified properly from the

cell lysate ‘l’. Although a strong TMPRSS2 signal at the expected size is visible in the cell pellet ‘p’, mainly the 60 kDa background protein was present in the elution fractions. This unspecific protein seems to interact stronger with the Ni-NTA resin than the 6xHis-tagged truncated TMPRSS2.



**Figure 5-4: Medium-scale expression and purification of truncated 6xHis-tagged TMPRSS2 variants.** TMPRSS2 variants were transiently expressed in 293T cells and purified via IMAC using Nickel-NTA. Several samples were taken during the purification process and separated in SDS-PAGE. Proteins were blotted to a PVDF membrane and visualized with an  $\alpha$ -His-tag primary and an HRP-conjugated secondary antibody. p: cell pellet, l: cell lysate, ft: flow through, w: washing step, e: elution step,  $e_k$ : concentrated eluate. Red arrow: background 60 kDa and 37 kDa signals, blue arrow: TMPRSS2 zymogen, green arrow: 27 kDa matriptase SPD. **A:** *Ig $\kappa$ \_ATM\_TMPRSS2\_HIS*. **B:** *Ig $\kappa$ \_ATM\_EK\_TMPRSS2\_HIS*.

The 293T cell line is a popular expression tool that has been extensively used in the last decades for recombinant protein production. The cells are easy to cultivate, show a high transfection efficiency and have a biochemical machinery providing most of the post-translational folding and processing that is often required to generate functional and mature proteins. However, the cell biochemical machinery also strongly produces non-target proteins, which obviously interfered with the Ni-NTA resin in our batch purification experiments. Certainly, optimizations including the transfection and the purification procedure are possible but the results of these first purification experiments at a medium-scale were not very promising. Additionally, the elution fraction also did not show proteolytic activity towards the pNA substrate Mes-DCha-Gly-Arg-pNA (described in chapter 3.2). Besides, the used transfection reagent Lipofectamine<sup>®</sup> 2000 is highly cost-intensive and not practicable for large scale experiments to produce required TMPRSS2 quantities. However, the purification experiments support the results that truncated TMPRSS2 from the construct *Ig $\kappa$ \_ATM\_TMPRSS2\_HIS* ((c)



in Figure 5-3 D) might result in functional TMPRSS2 since autocatalytic cleavage took place (27 kDa signal, Figure 5-3 B and Figure 5-4 A (green arrow)). With these first expression experiments it has been shown that TMPRSS2 does not need its transmembrane domain, but the Ig $\kappa$ -signal to mature properly. In contrast to the similar experiments with the TTSP corin<sup>[274]</sup>, no secreted soluble TMPRSS2 was detected in conditioned cell culture medium of 293T cells also at a medium expression scale (T75 cell culture flask, Figure 5-3 C). The experiments of CHEN and colleagues, where inactive truncated R255Q\_TMPRSS2 was isolated from culture media of transfected 293T cells, could not be reproduced with the new TMPRSS2 variants during this thesis<sup>[273]</sup>. However, they used stably transfected cells allowing them to perform experiments at a much larger scale. Eventually, secreted TMPRSS2 concentration in the screened cell culture media was below the detection limit of the primary antibody and much larger cell numbers are required. Nevertheless, since TMPRSS2 was perfectly detectable in crude cell lysates it is suggested that truncated TMPRSS2 remains at least partly located somewhere in the cells despite the introduction of the artificial secretion peptide.

### **5.2.3. Further preparation of TMPRSS2 variants with a C-terminal FLAG-tag**

Since the purification of 6xHis-tagged TMPRSS2 variants using IMAC at the present scale did not turn out as an appropriate method, further truncated and/or mutated *TMPRSS2* constructs were designed containing a C-terminal FLAG-tag for purification and detection. All truncated TMPRSS2 variants have the Ig $\kappa$ -SP added to the N-terminus. Table 5-1 gives an overview of the produced and analyzed *TMPRSS2* constructs including their expected calculated molecular weight. Molecular weights were determined solely based on the amino acid sequence without consideration of eventual PTMs using the ExPASy ProtParam tool (SIB Bioinformatics Resource Portal). In small scale expression experiments, TMPRSS2 variants were investigated regarding their autocatalytic cleavage activity upon transient expression in 293T cells by SDS-PAGE and in western blot experiments using specific primary and HRP conjugated secondary antibodies. As specified before, samples showing a 27 kDa signal are considered as potentially 'active'.

Table 5-1: List and schematic overview of TMPRSS2 constructs containing a C-terminal FLAG tag.

<div><div><div><div><div>1</div><div>NH<sub>2</sub></div><div><div><div></div><div></div><div></div></div></div><div><div><div>C</div><div>244</div></div><div><div>C</div><div>365</div></div></div><div><div>H</div><div>D</div><div>S</div></div><div><div>COOH</div></div></div><div>2</div><div>NH<sub>2</sub></div><div><div><div></div><div></div><div></div></div></div><div><div><div>C</div><div>244</div></div><div><div>C</div><div>365</div></div></div><div><div>H</div><div>D</div><div>A</div></div><div><div>COOH</div></div></div><div>3</div><div>NH<sub>2</sub></div><div><div><div></div><div></div><div></div></div></div><div><div><div>S</div><div>244</div></div><div><div>C</div><div>365</div></div></div><div><div>H</div><div>D</div><div>S</div></div><div><div>COOH</div></div></div><div>4</div><div>NH<sub>2</sub></div><div><div><div></div><div></div><div></div></div></div><div><div><div>C</div><div>244</div></div><div><div>S</div><div>365</div></div></div><div><div>H</div><div>D</div><div>S</div></div><div><div>COOH</div></div></div> <div>5</div> <div>NH<sub>2</sub></div> <div><div><div></div><div></div><div></div></div></div> <div><div><div>S</div><div>244</div></div><div><div>S</div><div>365</div></div></div> <div><div>H</div><div>D</div><div>S</div></div> <div><div>COOH</div></div>
--

6

NH<sub>2</sub>

↑

149

C

244

C

365

H

D

S

COOH

7

NH<sub>2</sub>

C

244

C

365

↑

479

H

D

S

COOH

Igκ

TM

LDLRA

SRCR

SPD

FLAG tag

8

NH<sub>2</sub>

C

244

C

365

H

D

S

COOH

9

NH<sub>2</sub>

C

244

C

365

H

D

A

COOH

10

NH<sub>2</sub>

C

244

S

365

H

D

S

COOH

11

NH<sub>2</sub>

C

244

C

365

H

D

S

COOH

12

NH<sub>2</sub>

C

244

S

365

H

D

S

COOH

13

NH<sub>2</sub>

C

244

C

365

H

D

S

COOH

14

NH<sub>2</sub>

C

244

C

365

H

D

S

COOH

15

NH<sub>2</sub>

C

244

C

365

H

D

S

COOH

16

NH<sub>2</sub>

C

244

C

365

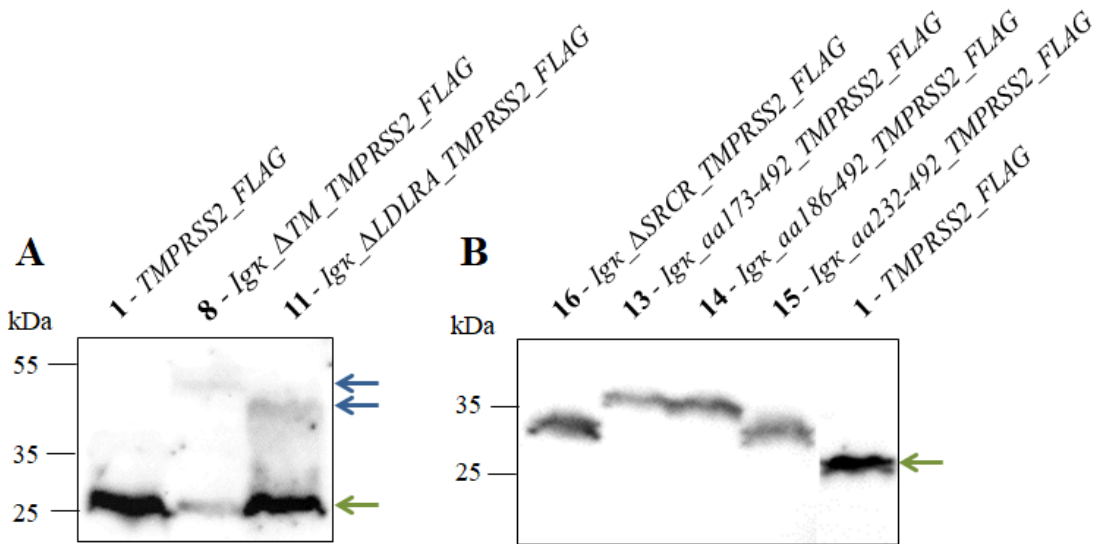
H

D

S

COOH

12	<i>Igκ_ALDLRA_TMPRSS2_C365S_FLAG</i>	149 - 492	Truncated by cytosolic tail, TM domain and LDLRA domain, mutation of cysteine 365 to serine.	45.4
13	<i>Igκ_aa173-492_TMPRSS2_FLAG</i>	173-492	Truncated by amino acid 1-172 (including the cytosolic tail, TM & LDLRA domains and the first 22 residues of SRCR domain)	42.6
14	<i>Igκ_aa186-492_TMPRSS2_FLAG</i>	186-492	Truncated by amino acid 1-185 (including the cytosolic tail, TM & LDLRA domains and the first 35 residues of SRCR domain)	41.1
15	<i>Igκ_aa231-492_TMPRSS2_FLAG</i>	231-492	Truncated by amino acid 1-230 (including the cytosolic tail, TM & LDLRA domains and 80 residues of SRCR domain)	36.0
16	<i>Igκ_ΔSRCR_TMPRSS2_FLAG</i>	243 - 492	Truncated by cytosolic tail, TM, LDLRA and SRCR domain.	34.8
17	<i>ΔTM_TMPRSS2_FLAG<sup>b</sup></i>	106 - 492	Truncated by cytosolic tail and TM domain. Without artificial signal peptide N-terminally	43.8
<sup>a</sup> calculated protein size using ExPASyProtParam ( <a href="https://web.expasy.org/protparam">https://web.expasy.org/protparam</a> ). PTMs are not considered.				
<sup>b</sup> truncated construct 17 without <i>Igκ</i> -SP ((d) in Figure 5-3 D)				
* indicate single-site mutations				



**Figure 5-5: Autocatalytic activation of truncated TMPRSS2 variants with C-terminal FLAG-tag.** TMPRSS2 variants were transiently expressed in 293T cells. Crude cell lysates were separated in SDS-PAGE and blotted to a PVDF membrane. TMPRSS2 was visualized using a specific  $\alpha$ -TMPRSS2 primary antibody and an HRP-conjugated secondary antibody. Blue arrows indicate the TMPRSS2 zymogens (if visible), the green arrow indicates the 27 kDa SPD. **A:** Constructs 1, 8 and 11. **B:** Constructs 13-16. The full-length construct (construct 1) containing a FLAG-tag was used as positive control.

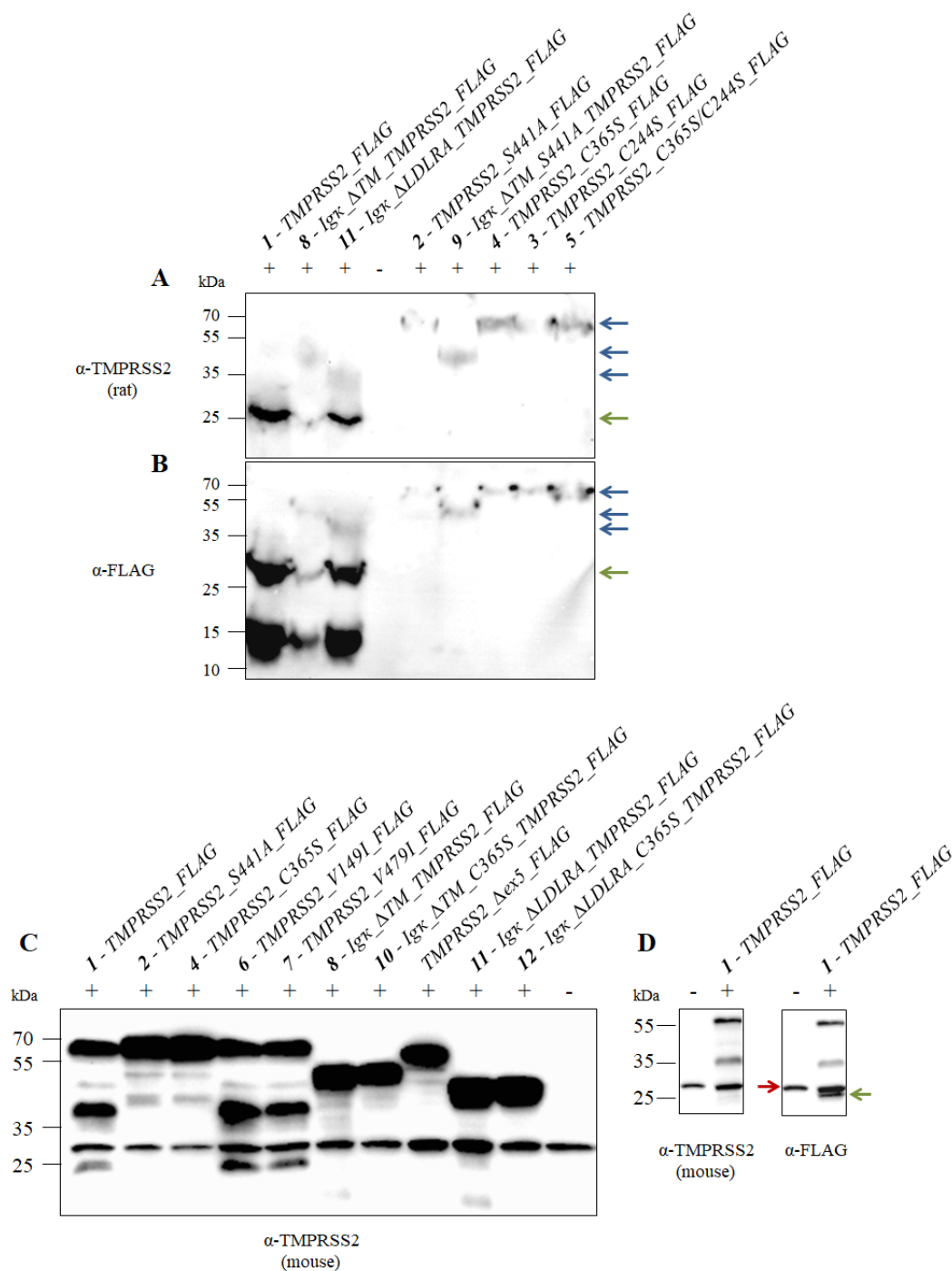
As already shown before in Figure 5-3, TMPRSS2 without the transmembrane domain can be autocatalytically processed in 293T cells also with a C-terminal FLAG tag (construct 8, Figure 5-5 A) instead of a 6xHis-tag ((e) in Figure 5-3). In addition, the TMPRSS2 variant, lacking the transmembrane and the LDLRA domain (construct 11) also harbors autocatalytic cleavage activity. Besides the 27 kDa signal, signals of 50 kDa (*Igκ\_ΔTM\_TMPRSS2\_FLAG*) and 45 kDa (*Igκ\_ΔLDLRA\_TMPRSS2\_FLAG*) are visible and match the expected calculated size of the truncated TMPRSS2 variants, respectively. Even further truncated TMPRSS2 variants do

not result in properly matured protein. Samples from constructs 13-16 only show signals of the zymogens but no SPD related signal at 27 kDa (Figure 5-5 B). These results suggest that the SRCR domain, but not the LDLRA nor the transmembrane domain are important for protease activity, measured on the availability of the freshly synthesized TMPRSS2 macromolecule to undergo autocatalytic activation. The SRCR domain comprises of amino acid 150 - 242 and contains two disulfide bridges, one between cysteines 172 and 231 and another between cysteines 185 and 241 (UniProtKB - O15393). Not only the removal of the complete SRCR domain (construct 16), but already removing the first cysteine residue downstream of the LDLRA domain (C172; construct 13) results in the loss of activity. Whether hampering a crucial disulfide bond formation or removing another crucial residue by truncating for further 22 amino acid compared to the construct 11 causes the loss of autocatalytic activity remains to be further investigated. In the case of hepsin, a crystal structure has shown that the SRCR domain was rigidly bound to the back of the SPD through the conserved linking disulfide bridge and a network of non-covalent interactions <sup>[87]</sup>. It could be well possible that also in case of TMPRSS2 the SRCR domain interacts with the SPD thereby contributing to its active conformation as discussed for other TTSPs <sup>[88]</sup>.

In a previous study, the deletion of nucleotides nt382 - nt501 ( $\Delta$ ex5) resulted in a TMPRSS2 variant without LDLRA domain, but still containing the transmembrane domain <sup>[236]</sup>. The deletion was in frame but caused a single site mutation of valine residue 149 to isoleucine (V149I). *In vitro* studies revealed that the  $\Delta$ ex5\_TMPRSS2 variant did not display autocatalytic cleavage activity and was also not able to cleave HA in co-expression experiments. The authors concluded that the LDLRA domain is essential for enzymatic activity. However, the new results from this work imply that the LDLRA domain is not a limiting factor for at least TMPRSS2 autocatalytic cleavage activity. The main differences between the LDLRA depleted TMPRSS2 variants include the V149I mutation in the case of the exon 5 silenced inactive variant ( $\Delta$ ex5) and the transmembrane domain that is removed additionally in case of the 'active' *Igκ\_ΔLDLRA\_TMPRSS2\_FLAG* variant (construct 11) from this work. A negative impact of the V149I mutation on TMPRSS2 activity can be excluded since a full-length TMPRSS2 variant harboring this exact amino acid substitution (construct 6, *TMPRSS2\_V149I\_FLAG*) shows autocatalytic cleavage activity (Figure 5-6 C). If activity loss of the  $\Delta$ ex5\_TMPRSS2 occurs due to possible interactions between the remaining extracellular portion and the transmembrane domain or a general structural rearrangement occurs remains highly speculative. Here, only a crystal structure of this TMPRSS2 variant could reveal more detail. The LDLRA domain itself is in any event not a limiting factor for TMPRSS2 activity (see construct 11 in Figure 5-6 A and B).

In Figure 5-6, truncated and/or mutated TMPRSS2 variants were visualized using three different primary antibodies: two polyclonal  $\alpha$ -TMPRSS2 (isolated from rat or mouse) and one  $\alpha$ -FLAG primary antibody. In consistency with the previous results truncated TMPRSS2 variants (construct 11 and weaker also in construct 8) yet again showed autocatalytic cleavage activity when detected with the rat  $\alpha$ -TMPRSS2 primary antibody (Figure 5-6 A) and  $\alpha$ -FLAG-tag antibody (Figure 5-6 B). As expected, no autocatalytic cleavage activity was observed in the case of the negative control (construct 2). In parallel, the S441A mutation results also in activity loss of the transmembrane truncated TMPRSS2 (construct 9 in Figure 5-6 A and B). The results shown in Figure 5-6 yet again support that the disulfide bond connecting the SPD to the stem region post activation cleavage is crucial for TMPRSS2 activity. The mutation of either one of the involved cysteine residues (C244 or C365) or both (constructs 3,4 and 5 respectively) caused the loss of autocatalytic cleavage activity. In addition, the lacking 27 kDa signal for construct 4 containing the FLAG-tag confirms the previous results, where no cleavage activity was observed for TMPRSS2\_C365S with a 6xHis-tag C-terminally ((e) in Figure 5-3). Both membranes in Figure 5-6 A and B reveal difficulties in the detection of precursor TMPRSS2 forms. Precursor proteins of all constructs are only vaguely visible and required the loading of high amounts of 293T cell lysates. Due to gel overloading, smeary signals are the consequence. Nevertheless, the protease domain is clearly visible, if present.

In Figure 5-6 C, TMPRSS2 variants were detected by using a different polyclonal  $\alpha$ -TMPRSS2 primary antibody (produced in mice, kindly provided by CHENG-YONG LIN (Georgetown University) to the group of PROF. FRIEBERTSHÄUSER) than in Figure 5-6 A. This antibody has a higher sensitivity towards the precursor proteins and shows several cleavage products, but is not able to detect the 27 kDa SPD signal as highlighted in Figure 5-6 D. Unfavorably, the more sensitive mouse-derived antibody results in a background signal of approximately 28-29 kDa that is slightly above the potential 27 kDa SPD signal. Only the extra incubation step with the  $\alpha$ -FLAG primary antibody (additional to incubation with the mouse-derived  $\alpha$ -TMPRSS2 antibody) shows the SPD-signal in the respective samples (Figure 5-6 D, right).



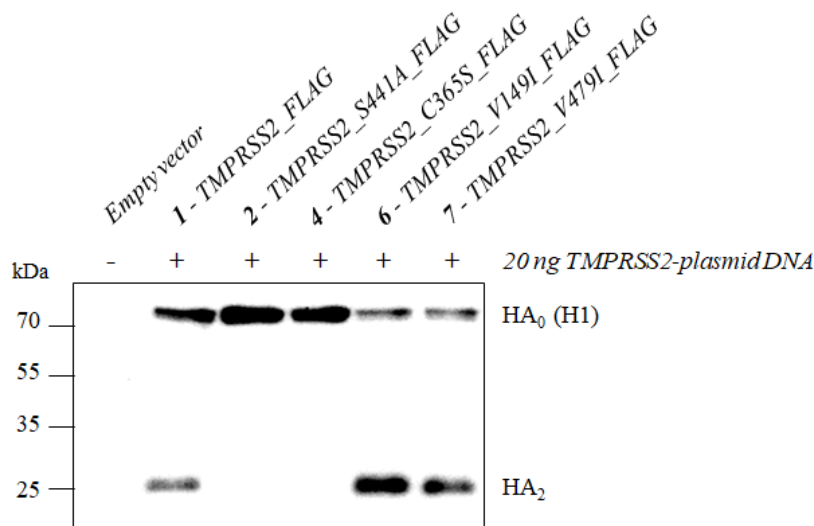
**Figure 5-6: Autocatalytic activation of truncated and/or mutated TMPRSS2 variants with C-terminal FLAG-tag.** TMPRSS2 constructs were transiently expressed in 293T cells. Crude cell lysates were separated via SDS-PAGE and immunostained after blotting to a PVDF membrane. +: 293T cell lysates were transfected with respective TMPRSS2 constructs, -: transfection with empty pCAGGs vector. In **A** and **B** the same samples were visualized using two different primary antibodies in combination with the respective HRP-conjugated secondary antibody. **A**: Rat-derived  $\alpha$ -TMPRSS2 primary antibody, **B**:  $\alpha$ -FLAG primary antibody. Blue arrows indicate the respective TMPRSS2 pro-enzymes and the green arrows indicate the 27 kDa SPD. **C**: Proteins in cell lysates containing constructs 1-3, 6-8 and 10-12 were visualized using a second mouse-derived  $\alpha$ -TMPRSS2 primary antibody (kindly provided by CHENG-YONG LIN to the lab of PROF. DR. BÖTTCHER-FRIEBERTSHÄUSER) which recognizes the precursor proteins and additional cleavage products but not the 27 kDa SPD. **D**: The identical PVDF membrane containing respective proteins was first incubated with the primary mouse  $\alpha$ -TMPRSS2 antibody (left) and then with the primary  $\alpha$ -FLAG primary antibody as well as species-specific HRP-conjugated secondary antibodies (right). A background signal of approximately 29 kDa (red arrow) is present in all cell lysates including the empty vector control. Only after incubation with the  $\alpha$ -FLAG antibody, the 27 kDa SPD appears (green arrow).

Yet, different TMPRSS2 variants can be compared and evaluated for autocatalytic cleavage activity based on their cleavage pattern in reference to the positive (construct 1) and negative control (construct 2). The positive control shows two strong bands: one of approximately 40 kDa and a weaker signal at approximately 24 kDa. The S441A mutant only reveals two weak signals at circa 45 kDa and 42 kDa. With the knowledge that the positive control possesses proteolytic activity, the 40 kDa and 24 kDa signal can be considered as a marker for autocatalytic TMPRSS2 cleavage activity when using the mouse-derived primary  $\alpha$ -TMPRSS2 antibody. Whereas C365S\_TMPRSS2 has the same cleavage pattern as the negative control, the V149I and V479I mutated TMPRSS2 variants are matching the cleavage pattern of the positive control suggesting autocatalytic cleavage activity of V149I\_TMPRSS2 and V479I\_TMPRSS2. Therefore, the applied method to create the single site mutated TMPRSS2 variants has no negative impact on TMPRSS2 activity itself. Again, results support a C365S mutation dependent TMPRSS2 activity loss. Since the same result was observed when mutating residue C244 to serine, it indicates that not the residue C365 itself is essential, but rather the formation of the corresponding disulfide bond. Besides linking the SPD to the stem region after the activation cleavage, this disulfide bond seems to impact TMPRSS2 maturation in general. The proofed autocatalytic cleavage activity of V149I\_TMPRSS2 (from construct 6) also excludes a V149I-related activity loss of the LDLRA lacking TMPRSS2 variant (construct *TMPRSS2\_Δex5\_FLAG* in Figure 5-6 C) from the PPMO-studies <sup>[236]</sup>. No clear conclusions can be drawn for the truncated TMPRSS2 constructs, since they show an aberrant cleavage pattern. Instead of the antibody-dependent activity related 24 kDa signal of the full-length TMPRSS2 variants, truncated TMPRSS2 from construct 8 and 11 (*Igκ\_ΔTM\_TMPRSS2\_FLAG* and *Igκ\_ΔLDLRA\_TMPRSS2\_FLAG*, respectively) possess a very weak signal clearly below 20 kDa (Figure 5-6 C) which might represent cleavage products from autocatalytic cleavage. Ultimately, the detection of different TMPRSS2 variants in western blot experiments was not straightforwardly possible with only one antibody, but the combination of three different primary antibodies led to conclusive results. The more sensitive mouse  $\alpha$ -TMPRSS2 antibody thereby allows detection of precursor proteins but is not suitable to evaluate activation cleavage of truncated TMPRSS2. In contrast, the other two used antibodies can recognize the 27 kDa SPD signal of full-length and truncated TMPRSS2 variants, if present.

#### 5.2.4. Co-expression experiments of TMPRSS2 constructs with H1-HA or H3-HA

So far, different TMPRSS2 variants were investigated for autocatalytic cleavage activity in western blot experiments. Depending on the used primary antibody, the appearance of the 27 kDa SPD signal or a specific cleavage pattern had been associated with a proteolytic active

protease. To further evaluate protease activity, *in vitro* TMPRSS2 cleavage activity towards H1-HA or H3-HA was investigated in co-expression experiments. Proteolytic active TMPRSS2 cleaves the precursor HA<sub>0</sub> into the subunits HA<sub>1</sub> and HA<sub>2</sub> [58, 236, 262]. Different TMPRSS2 variants were therefore co-transfected with H1- or H3-HA in 293T cells. The respective cell lysates were separated by SDS-PAGE, blotted to membranes and immunostained using primary  $\alpha$ -H1N1 or  $\alpha$ -H3N2 antibodies, respectively. The appearance of an HA<sub>1</sub>- (55 kDa) or HA<sub>2</sub>- (25 kDa) related signals suggest HA<sub>0</sub> (75 kDa) cleavage, thus TMPRSS2 activity. Results are shown in Figure 5-7, Figure 5-8 and Figure 5-9.

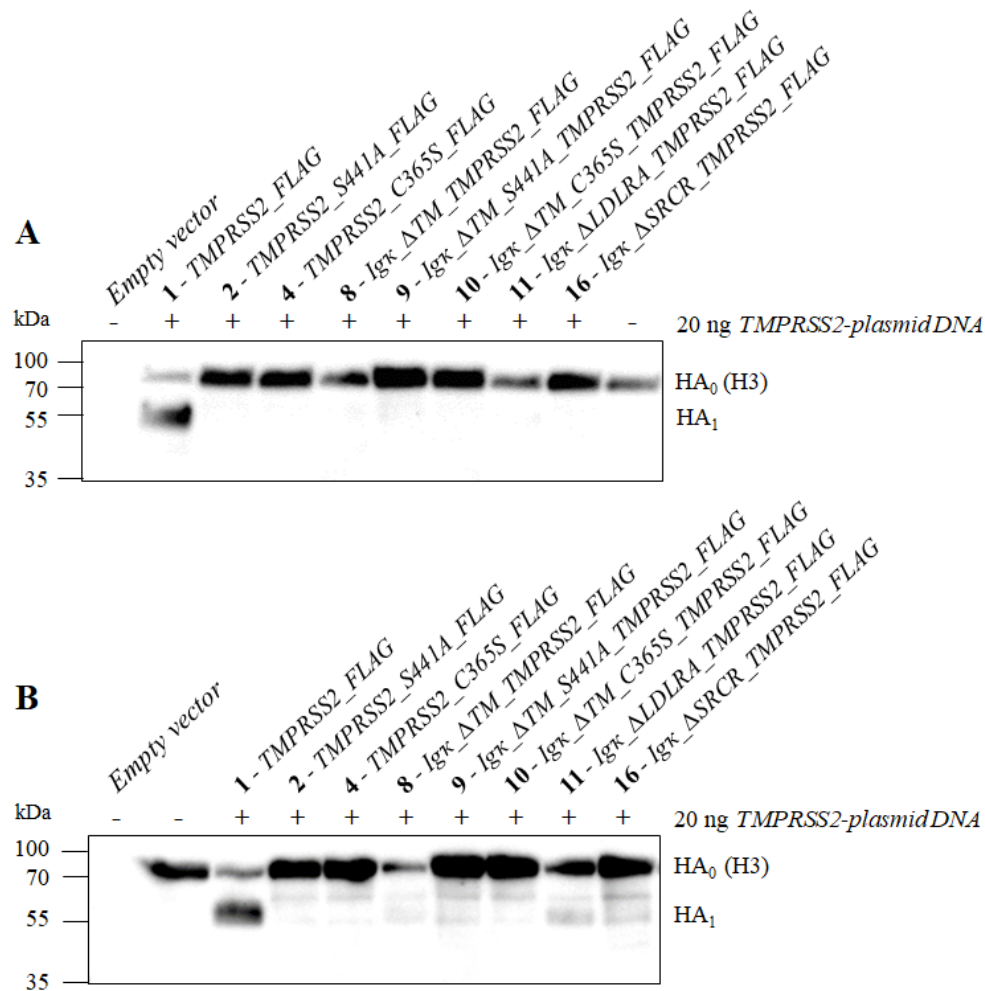


**Figure 5-7: Proteolytic cleavage activity of mutated full-length TMPRSS2 variants towards H1-HA.** 293T cells were transiently co-transfected with plasmids encoding H1-HA and different TMPRSS2 variants (+) or an empty vector as a negative control. Cleavage of H1-HA<sub>0</sub> was detected in crude cell lysates 48 h post transfection by western blot analysis using a primary  $\alpha$ -H1N1 and an HRP-conjugated secondary antibody. TMPRSS2 activity is indicated by the appearance of the HA<sub>2</sub> signal at 25 kDa.

Figure 5-7 supports proteolytic activity of the V149I and V479I mutated full-length TMPRSS2. Both TMPRSS2 variants were able to cleave the precursor HA<sub>0</sub> of the H1-HA subtype. As for the positive control, a clear signal for the HA<sub>2</sub> subunit at 25 kDa could be detected. In contrast, no cleavage activity towards HA<sub>0</sub> was observed in the case of the TMPRSS2\_C365S variant in consistency with the negative control (TMPRSS2\_S441A). These results confirm that the appearance of the 27 kDa SPD signal correlates with proteolytic activity in the case of full-length TMPRSS2 variants. Also, by implication, the lack of the SPD signal is associated with TMPRSS2 inactivity. The disulfide bond including the cysteine residue C365 has been identified as a crucial factor for TMPRSS2 maturation. Disulfide cross-links are known to have a substantial increase in the stability of proteins. In the case of TMPRSS2 the loss of this stabilizing disulfide bond might hamper structural proximity, required for autocatalytic activation as a prerequisite for protease activity and /or simply destabilize the molecule during



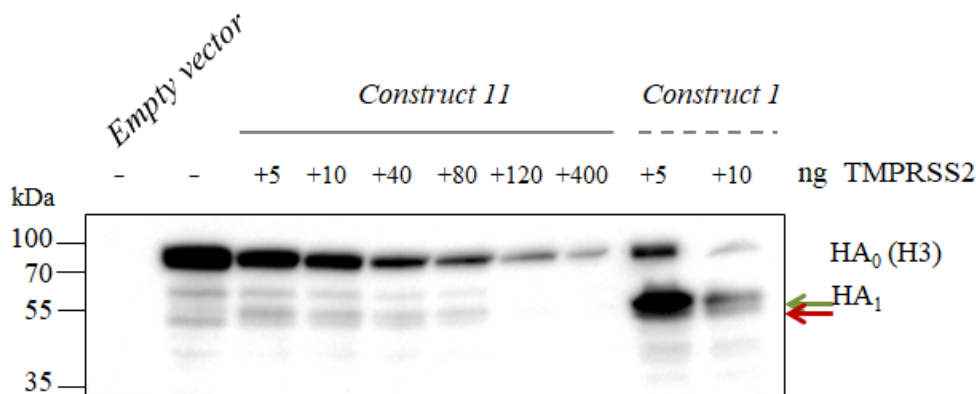
maturation. What exactly causes the activity loss of the TMPRSS2\_C365S variant remains unclear at this point.



**Figure 5-8: Proteolytic cleavage activity of mutated full-length and truncated TMPRSS2 variants towards H3-HA.** 293T cells were transiently co-transfected with plasmids encoding H3-HA and different TMPRSS2 variants (+) or with an empty vector as negative control. Cells solely transfected with H3-HA plasmid (-) served as additional control. Cleavage of H3-HA<sub>0</sub> was detected in crude cell lysates 48 h post transfection by western blotting using two different primary  $\alpha$ -H3N2 antibodies and the respective HRP-conjugated secondary antibody. **A:**  $\alpha$ -H3N2 primary antibody (isolated from goat), **B:**  $\alpha$ -H3N2 primary antibody (isolated from rabbit). The appearance of the 55 kDa HA<sub>1</sub> signal implies TMPRSS2 activity.

In Figure 5-8 A and Figure 5-8 B the same constructs were investigated for TMPRSS2 catalyzed H3-HA cleavage using two different  $\alpha$ -H3N2 primary antibodies. Only the positive control shows clear cleavage activity towards HA<sub>0</sub> (HA<sub>1</sub> associated 55 kDa cleavage product) in both cases. In consistency with the lacking 27 kDa SPD signal in the previous experiments, no cleavage activity was observed in case of H3-HA co-expression with constructs 2 and 4 as well as the constructs 9, 10 and 16. Results for constructs 8 (Ig $\kappa$  $\Delta$ TM\_TMPRSS2\_FLAG) and 11 (Ig $\kappa$  $\Delta$ LDLRA\_TMPRSS2\_FLAG) are rather distinct. No clear cleavage product of H3-HA as for the positive control is visible. However, as shown in Figure 5-8 B the co-expression with construct 11 in particular results in a slightly more intense and broader 55 kDa signal compared

to the negative control sample, where only H3-HA was transfected. In addition, the precursor HA<sub>0</sub> signals are less intense for the samples of constructs 8 and 11 compared to those of the other constructs (2, 4, 9, 10 and 16). As shown for the positive control, HA cleavage results in a less distinct signal for the precursor protein HA<sub>0</sub>. Therefore, already weaker precursor signals might imply marginal cleavage activity by the two discussed truncated TMPRSS2 forms. Qualitative western blot experiments, as performed in this thesis, are generally prone to different band intensities caused by experimental parameters, such as differences in sample loading, transfer of proteins to the membrane or the time-sensitive enhanced chemiluminescence (ECL) detection method. Quantification of proteins is generally possible but requires certain preconditions including antibody titration and internal loading controls. These parameters were not fulfilled in the given experiments. However, without clear HA<sub>1</sub> cleavage product, H3-HA cleavage by these two TMPRSS2 variants can not be ultimately evaluated. In the following, 293T cells were co-transfected with H3-HA and various amounts of plasmid DNA harboring the construct 11 (Figure 5-9).



**Figure 5-9: Proteolytic cleavage of H3-HA in a dose-dependent manner by truncated and full-length TMPRSS2.** 293T cells were transiently co-transfected with plasmids encoding HA of the H3 subtype and different amounts of full-length TMPRSS2 plasmid (construct 1: *TMPRSS2\_FLAG*) or truncated TMPRSS2 containing plasmid (construct 11: *Igκ\_ALDLRA\_TMPRSS2\_FLAG*) of different quantities. Cells transfected with an empty vector or solely plasmids encoding for H3-HA (-) served as negative controls. Cleavage of HA<sub>0</sub> was detected in crude cell lysates 48 h post transfection by western blotting using a primary  $\alpha$ -H3N2 and an HRP-conjugated secondary antibody. The red arrow indicates an antibody-specific background signal and the green arrow the HA<sub>1</sub> cleavage product at 55 kDa indicating TMPRSS2 activity.

The crude cell lysates were analyzed for H3-HA signals using the more sensitive rabbit derived primary  $\alpha$ -H3N2 antibody (also used in Figure 5-8 B). Results suggest an HA expression-dependency on the quantity of co-transfected TMPRSS2 construct. The higher the introduced TMPRSS2 plasmid DNA amount, the less intense appeared the HA<sub>0</sub> and HA<sub>1</sub> signals. H3-HA signals from co-transfection with 5 ng *TMPRSS2\_FLAG* plasmid-DNA (positive control) are stronger than from the 10 ng *TMPRSS2\_FLAG* plasmid DNA sample. However, the weaker

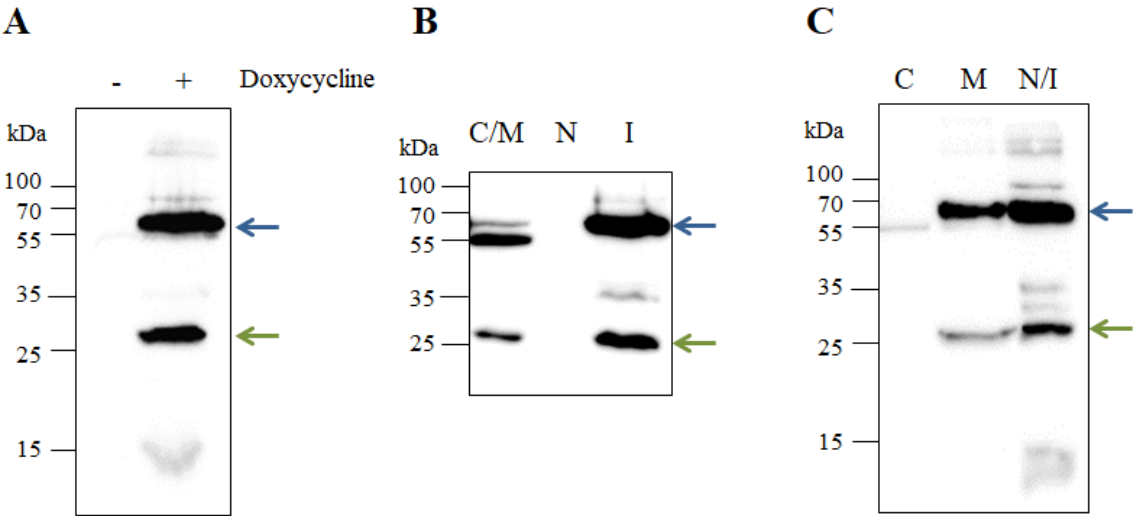
55 kDa HA<sub>1</sub> signal of the latter 10 ng sample in co-expression with the H3-HA allows distinguishing two different signals of very similar molecular weights. The slightly lower signal (red arrow) is referred to as background signal since it appears in all samples, including the negative control sample that was solely transfected with H3-HA plasmid DNA (lane 2 (-) in Figure 5-9). In consistency with the findings in Figure 5-8 B, a slightly more intense and broader 55 kDa signal is visible when HA-H3 was co-transfected with construct 11 in comparison to the negative control. A weak HA<sub>1</sub> signal (green arrow) in addition to the background signal (red arrow) is visible. Notwithstanding the signals are very flimsily, the additional signal at 55 kDa implies cleavage of HA<sub>0</sub> and hence indicate proteolytic activity of the truncated TMPRSS2 variant as it has been suggested from previous experiments where autocatalytic cleavage activity was observed with construct 11 (Figure 5-5 A, Figure 5-6 A and Figure 5-6 B). Even though autocatalytic cleavage activity of truncated TMPRSS2 (from construct 11) and of the full-length TMPRSS2 (construct 1) was comparable, only very weak H3-HA cleavage could be determined by the truncated TMPRSS2. It is conceivable that TMPRSS2 activity of truncated variants might simply not be adequately detectable in the given experimental setup. TMPRSS2 and HA are both membrane-anchored proteins. The removal of the TMPRSS2 anchoring domain could result in the unapproachability of the membrane-bound substrate hemagglutinin due to TMPRSS2 dislocation. When TMPRSS2 cannot spatially reach HA, HA<sub>0</sub> cannot be proteolytically cleaved. TMPRSS2 has been suggested to exclusively activate HA within the TGN or possibly in endosomes, which might indicate a local environmental dependency of proteolytic activity. Also, truncated TMPRSS2 might mature properly to be autocatalytically processed but could subsequently be identified as foreign by the cell recycling system resulting in the activation of its recycling. However, all TMPRSS2 variants were detectable in western blot experiments and did not show any signs of degradation. Altogether, HA<sub>0</sub> cleavage by co-expressing H3-HA and truncated TMPRSS2 variants is not a very suitable system to evaluate proteolytic TMPRSS2 activity. So far, no other activity tests are available for TMPRSS2 in cell culture experiments that could be used to further investigate and determine the specific activity of truncated TMPRSS2.

### 5.2.5. Subcellular localization of TMPRSS2

For a deeper insight into the subcellular localization of TMPRSS2, cell compartments of stably *TMPRSS2* expressing MDCK-H cells (MDCK-TMPRSS2 cells) were separated and analyzed. In addition, different TMPRSS2 variants were investigated in indirect immunofluorescence experiments for subcellular localization within Huh-7 cells after transient expression in order to detect possible dislocation of truncated TMPRSS2 variants.

Cell compartmentation

*TMPRSS2* expression in MDCK-H cells was induced by the addition of doxycycline (Figure 5-10 A). Proteins were isolated from different cell compartments, including the cytosol (C-fraction), membrane-bound organelles (M-fraction) and the cell nucleus (N-fraction) and separated via SDS-PAGE before being analyzed in western blot experiments with the  $\alpha$ -FLAG primary antibody (Figure 5-10 B and C). Precipitated proteins during isolation as well as the cell debris are summarized as insoluble fraction (I). Two different methods were applied for compartmentation. In Figure 5-10 B, the C- and M-fraction were isolated together by the *ProteoJET*<sup>TM</sup> Kit components and separated from the N-fraction and the I-fraction. By using the second method <sup>[275]</sup>, proteins from C- und M-fractions could be isolated separately (Figure 5-10 C). In this case, N- and I-fractions were isolated together for simplicity since no *TMPRSS2* was suspected and found in the N-fraction before (Figure 5-10 B).

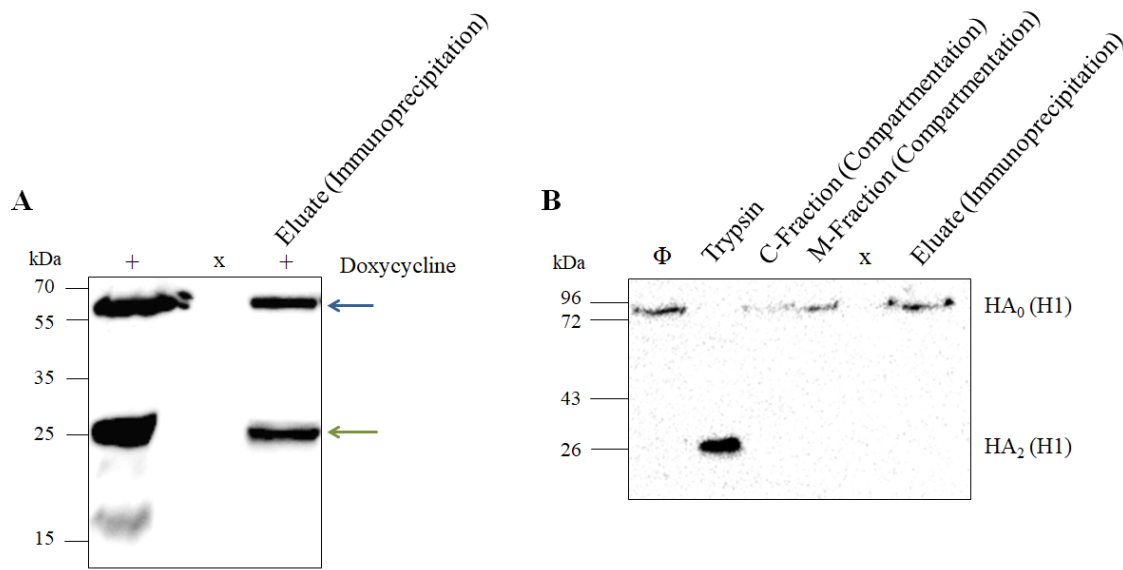


**Figure 5-10: Doxycycline induced *TMPRSS2* expression (A) and subcellular localization of full-length *TMPRSS2* by cell compartmentation (B+C).** MDCK-*TMPRSS2* cells were grown in T75 cell culture flasks in the absence (-) or presence (+) of doxycycline for 48 h. Cell compartments were isolated in **B** using the Fermentas *ProteoJET*<sup>TM</sup> Kit by following manufacturer's instructions or in **C** based on a protocol from <sup>[275]</sup>. *TMPRSS2* was detected by western blot analysis using  $\alpha$ -FLAG primary antibody and HRP-conjugated secondary antibody. Proteins from different cell compartments are indicated with C for cytosolic, M for membrane-bound organelles, N for nucleic and I for insoluble protein fractions. Since no *TMPRSS2* was found in the nucleic protein fraction in **B**, only C- and M- protein fractions were separated in the following compartmentation experiments as indicated in **C**. Blue arrow: glycosylated *TMPRSS2* zymogen, green arrow: *TMPRSS2*-SPD.

Active full-length *TMPRSS2* was only detectable upon doxycycline induction in MDCK-*TMPRSS2* crude cell lysates (see 27 kDa SPD signal in Figure 5-10 A). After cell compartmentation using the *ProteoJET*<sup>TM</sup> Kit, *TMPRSS2* was found in the C/M fraction and the I-fraction. In the latter insoluble fraction, *TMPRSS2* related signals are most likely a result of precipitated proteins due to the detergent-based isolation method. The separation of C- and M-fractions (Figure 5-10 C) shows a *TMPRSS2* signal at approximately 55 kDa in the C-

fraction matching the calculated size of the non-glycosylated zymogen form of TMPRSS2 (see also Table 5-1, construct 1). Matured active TMPRSS2 possessing the 27 kDa SPD was only detected in the M-fraction and again in the portion containing insoluble or precipitated proteins plus the nucleic proteins (N/I). The TMPRSS2 precursor in the M-fraction is, with approximately 65 - 70 kDa, of a slightly bigger than calculated size and matches the predicted size of glycosylated TMPRSS2 <sup>[76]</sup>. Indeed, this result is not surprising, since membrane-bound organelles (M-fraction) include the endoplasmic reticulum where PTMs take place. Therefore, compartmentation results confirm the presence of possibly active TMPRSS2 exclusively in membrane-bound organelles but not as soluble protein in the cytoplasm nor in the cell nucleus. Crude MDCK-H and MDCK-TMPRSS2 cells lysates as well as the respective isolated C- and M-fractions were tested for proteolytic activity by incubation with the Mes-DCha-Gly-Arg-pNA substrate and visually analyzed. All samples showed a color turnover after incubation overnight at RT with the strongest yellow staining in the crude cell lysate of MDCK-H cells and MDCK-TMPRSS2 cells, but no visible difference between the C- and M-fraction. Since a wide range of serine proteases cleave substrates after monobasic amino acid residues, the yellow staining gives no evidence of TMPRSS2 specific proteolytic activity.

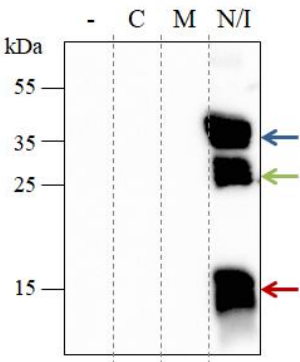
With a major objective in the head to create an appropriate system to isolate active TMPRSS2 for functional testing, TMPRSS2 was also isolated via immunoprecipitation from MDCK-TMPRSS2 cell lysates using ANTI-FLAG® M2 Affinity Gel (Sigma-Aldrich) that allows isolation of FLAG-tagged proteins. Although TMPRSS2 was captured successfully (see Figure 5-11 A) no proteolytic cleavage activity could be detected of these samples using the aforementioned Mes-DCha-Gly-Arg-pNA substrate. In addition, the C- and M-fractions, as well as the eluate from the immunoprecipitation experiments were incubated with 293T cells that were previously transfected with the H1-HA plasmid in order to investigate proteolytic TMPRSS2 specific HA cleavage activity. After 3 h incubation at 37 °C, H3-HA cleavage was only observed in case of the positive control containing externally added trypsin (see Figure 5-11 B) but not in samples where the C-, M- or eluate fractions were externally added. No TMPRSS2 activity could be detected outside the cell.



**Figure 5-11: Isolated full-length TMPRSS2 does not possess catalytic cleavage activity towards H1-HA. A:** TMPRSS2 was isolated from doxycycline induced MDCK-TMPRSS2 crude cell lysate by immunoprecipitation using the ANTI-FLAG M2 antibody (Sigma-Aldrich). Bound protein was eluted by incubation with glycine-HCl at pH 3.0 and directly neutralized using 10x TBS. Blue arrow indicates TMPRSS2 zymogen and the green arrow indicates the 27 kDa SPD. **B:** 293T cells were transiently transfected with an H1-HA encoding plasmid. After 48 h, cells were harvested by centrifugation and incubated with trypsin (1 h at 37 °C) or portions from cell compartmentation (C- and M-fractions of MDCK-TMPRSS2 cells) or the eluate from immunoprecipitation for 3 h at 37 °C. Cleavage of hemagglutinin is indicated by the appearance of a 25 kDa HA<sub>2</sub> signal. Color Prestained Protein Standard, Broad Range (11–245 kDa; NEB) was used as a marker in B.

To see if the removal of the anchoring system results in TMPRSS2 dislocation, MDCK-H cells were transiently transfected with the construct *Igκ<sub>Δ</sub>LDLRA\_TMPRSS2\_FLAG* (construct 11) and cell compartmentation experiments were performed (Figure 5-12). Truncated autocatalytically processed TMPRSS2 was exclusively found in the N/I-fraction that includes the nucleic and insoluble protein portions. Neither in the C- nor in the M-fraction was any TMPRSS2 signal detectable and TMPRSS2 repeated to precipitate completely during the separation of the different cell fractions. This might suggest an incompatibility with the buffer components that contain detergents that are standardly used to form stabilizing complexes with membrane proteins during isolation. Without the membrane anchoring domain, the detergents might have a rather destructive character. Also, aberrant protein stability might be the result of truncation by losing important interactions or by forming unfavored new ones. Also, a negative effect of the transfection method itself could be plausible. MDCK-H cells were transiently transfected with construct 11 using the strong transfection reagent Lipofectamine 2000. Full-length TMPRSS2 was isolated from stably transfected cells where expression was induced upon doxycycline addition. Although precipitation due to general protein overload is not excludable, only a partly precipitation would have been expected in this case. Ultimately, the localization of truncated TMPRSS2 within the cells by compartmentation was not possible in the given

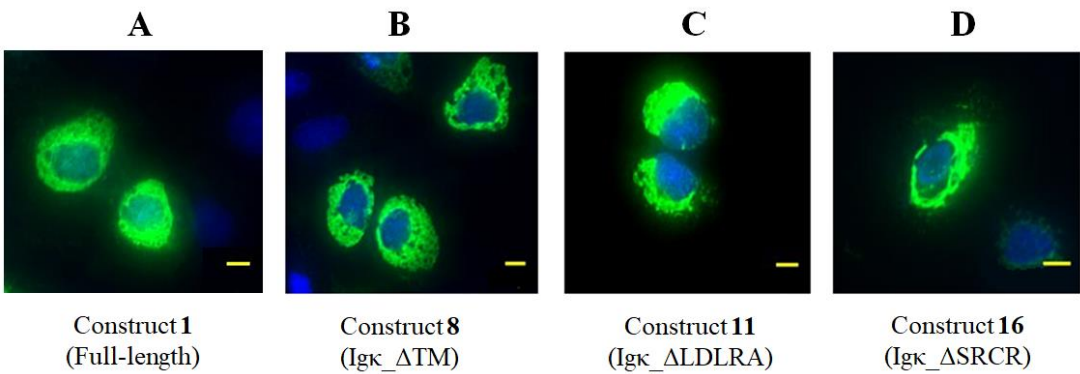
experimental setup. The reason for experimental failure remains speculative. Nevertheless, yet again, autoproteolytic cleavage activity of truncated TMPRSS2 using construct 11 has been confirmed (green arrow in Figure 5-12), which is generally associated with TMPRSS2 activity.



**Figure 5-12: Subcellular localization of truncated TMPRSS2 by cell compartmentation.** MDCK-H cells were transiently transfected with truncated TMPRSS2 (*Igκ<sub>κ</sub>\_ΔLDLRA\_TMPRSS2\_FLAG*). Cell compartments were separated 48 h after transfection based on a protocol from [275]. TMPRSS2 was detected by western blotting using α-FLAG primary and HRP-conjugated secondary antibodies. Proteins from different cell compartments are indicated with C for cytosolic, M for membrane-bound organelles, N for nucleus and I for insoluble protein fractions. Blue arrow: zymogen form, green arrow: SPD, red arrow: cleavage product/background signal.

*Indirect immunofluorescence*

Huh-7 cells were transiently transfected with different *TMPRSS2* constructs and proteins were visualized using the α-TMPRSS2 primary antibody and an Alexa Fluor® 488-conjugated secondary antibody (Figure 5-13 A-D). TMPRSS2 is detected as a green signal under the fluorescence microscope. Cell nuclei were visualized in blue following incubation with the dsDNA stain DAPI.



**Figure 5-13: Subcellular localization of truncated TMPRSS2 variants by indirect immunofluorescence.** Huh-7 cells were transiently transfected with the respective plasmids encoding for TMPRSS2 variants. TMPRSS2 was visualized 24 h post transfection using an α-TMPRSS2 primary and Alexa Fluor® 488-conjugated secondary antibody. Cell nuclei were counterstained with DAPI during incubation of the secondary antibody. **A:** *Igκ<sub>κ</sub>\_ΔTM\_TMPRSS2\_FLAG*, **B:** *Igκ<sub>κ</sub>\_ΔLDLRA\_TMPRSS2\_FLAG* and **C:** *Igκ<sub>κ</sub>\_ΔSRCR\_TMPRSS2\_FLAG*. Cell nuclei are blue, TMPRSS2 emerges in green. Yellow bars: 10 μm (magnitude 20x).

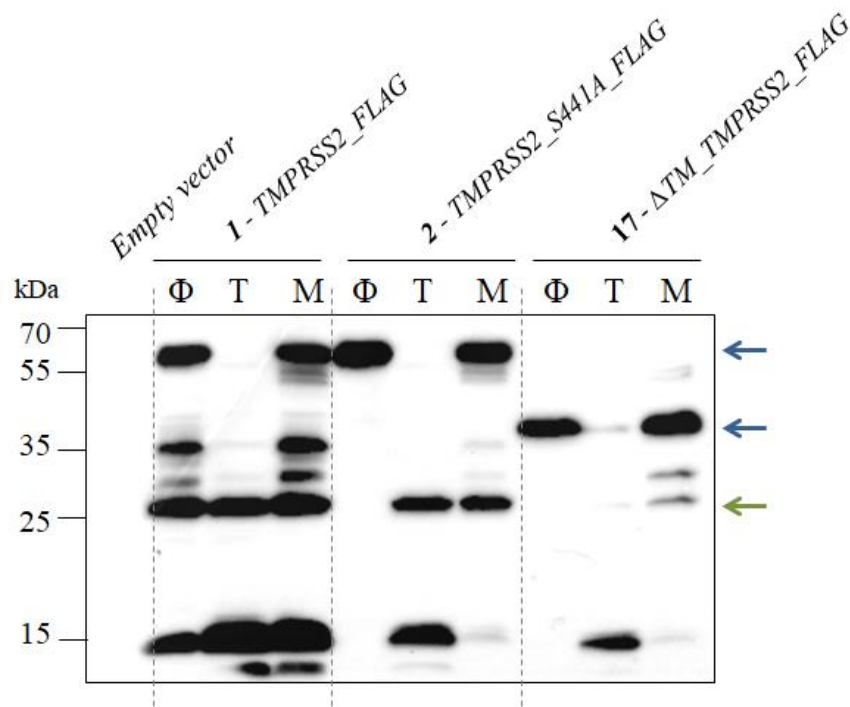


No aberrant location of truncated TMPRSS2, compared to the full-length TMPRSS2 (Figure 5-13 A) was detectable in Huh-7 cells. Full-length TMPRSS2 (construct 1) and the truncated TMPRSS2 variants from construct 8 (*Igκ\_ATM\_TMPRSS2\_FLAG*, Figure 5-13 B), construct 11 (*Igκ\_ALDLRA\_TMPRSS2\_FLAG*, Figure 5-13 C), as well as construct 16 (*Igκ\_ASRCR\_TMPRSS2\_FLAG*, Figure 5-13 D) are all located in cell compartments surrounding the cell nucleus, such as the ER and TGN. The removal of the anchoring TM domain does not seem to result in protein accumulation in the cytoplasm in which case the cell shape would have been indicated by the green color. However, at least a localization change within the single cell compartment is most likely since the protease variants can not be physically attached to the plasma membranes by the anchoring domain anymore. The applied technique and the optical magnification (20x) of the used microscope are not suitable to determine the precise location within a single cell compartment by microscoping. For a deeper insight, more experiments are necessary with different equipment. This could include the co-staining of TMPRSS2 and organelle protein-specific primary antibodies including, for example,  $\alpha$ -calnexin or  $\alpha$ -calreticulin for the ER,  $\alpha$ -GM130 for the Golgi-apparatus and  $\alpha$ -Rab4 for endosomes.

#### 5.2.6. Matriptase catalyzed activation of TMPRSS2

Androgen-induced TMPRSS2 has been shown to activate matriptase *in vitro* [237]. In Figure 5-14, 293T cells were harvested 48 h post transfection with different *TMPPRSS2* constructs (constructs 1, 2 and 17 of Table 5-1) and treated with recombinant matriptase or trypsin afterwards. Whereas the untreated full-length TMPRSS2 sample (construct 1) shows autocatalytic cleavage activity, only zymogen forms of the proteolytical inactive constructs 2 and 17 are present in the untreated samples. Upon treatment with either trypsin or matriptase (wt-matriptase from chapter 3) the SPD related 27 kDa signal additionally appears in samples of the constructs 2 and 17 (green arrow in Figure 5-14). Samples incubated with externally added trypsin thereby show (almost) complete cleavage of the inactive zymogen forms (no or only weak signal present after treatment, blue arrow) including also a strong cleavage product of approximately 15 kDa that is standardly present when using the  $\alpha$ -FLAG primary antibody (see Figure 5-6 B, Figure 5-10, Figure 5-11 and Figure 5-12). This 15 kDa signal is only vaguely visible for samples treated with matriptase, which is known to not cleave indiscriminately after nearly every arginine or lysine such as trypsin. These data suggest, that not only TMPRSS2 can activate matriptase, but also that TMPRSS2 is a possible matriptase substrate.





**Figure 5-14: Post expression proteolytic TMPRSS2 zymogen cleavage by externally added matriptase and trypsin *in vitro*.** 293T cells were transiently transfected with TMPRSS2 variants (constructs 1,2 and 17 in Table 5-1). Cells were harvested 48 h post transfection and treated with trypsin (T) for 1 h or wt-matriptase (M) for 3 h at 37 °C. Φ: cell lysates without additional protease treatment. The blue arrows indicate TMPRSS2 zymogen forms, green arrow indicates the 27 kDa SPD.

5.3. Summary of TMPRSS2 characterization

In total, 17 different TMPRSS2 variants with a FLAG-tag were investigated during this thesis. This includes 15 new mutated and/or truncated variants that were tested for their autocatalytic cleavage activity, proteolytic activity towards influenza HA, and partly for their subcellular localization. In the following, the results will be summarized and the characterization of TMPRSS2 will be further discussed. For an overview see Table 5-2.

**Table 5-2: Characterization of TMPRSS2 variants with FLAG-tags.**

Nr.	TMPRSS2 construct name	Amino acid residues	Autocatalytic cleavage activity	Proteolytic cleavage activity towards HA <sub>0</sub>
1	<i>TMPRSS2_FLAG (positive control)</i>	1 - 492	YES	YES
2	<i>TMPRSS2_S441A_FLAG (negative control)</i>	1 - 492	NO	NO
3	<i>TMPRSS2_C244S_FLAG</i>	1 - 492	NO	NO
4	<i>TMPRSS2_C365S_FLAG</i>	1 - 492	NO	NO
5	<i>TMPRSS2_C365S/C244S_FLAG</i>	1 - 492	NO	NO

Nr.	TMPRSS2 construct name	Amino acid residues	Autocatalytic cleavage activity	Proteolytic cleavage activity towards HA <sub>0</sub>
6	<i>TMPRSS2_V149I_FLAG</i>	1 - 492	YES	YES
7	<i>TMPRSS2_V479I_FLAG</i>	1 - 492	YES	YES
8	<i>Igκ_ΔTM_TMPRSS2_FLAG</i>	106 - 492	YES	UNCLEAR
9	<i>Igκ_ΔTM_TMPRSS2_S441A_FLAG</i>	106 - 492	NO	NO
10	<i>Igκ_ΔTM_TMPRSS2_C365S_FLAG</i>	106 - 492	NO	NO
11	<i>Igκ_ΔLDLRA_TMPRSS2_FLAG</i>	149 - 492	YES	YES
12	<i>Igκ_ΔLDLRA_TMPRSS2_C365S_FLAG</i>	149 - 492	NO	NO
13	<i>Igκ_aa173-492_TMPRSS2_FLAG</i>	173-492	NO	NO
14	<i>Igκ_aa186-492_TMPRSS2_FLAG</i>	186-492	NO	NO
15	<i>Igκ_aa231-492_TMPRSS2_FLAG</i>	231-492	NO	NO
16	<i>Igκ_ΔSRCR_TMPRSS2_FLAG</i>	243 - 492	NO	NO
17	<i>ΔTM_TMPRSS2_FLAG</i>	106 - 492	NO	NO

5.3.1. Autocatalytic cleavage activity of different TMPRSS2 variants

In a first round, TMPRSS2 variants containing a C-terminal 6xHis-tag were characterized in cell culture experiments. A TMPRSS2 variant without transmembrane domain but with the Igκ-SP instead (construct (c) in Figure 5-3) had been identified to mature properly, indicated by autocatalytic cleavage activity. A 27 kDa signal of the SPD is visible in the respective samples. In contrast, the single-site mutation of C365S (construct (e)) caused the loss of autocatalytic TMPRSS2 cleavage activity.

In a second set up, TMPRSS2 variants with C-terminal FLAG-tag were investigated (Table 5-1). Initial results could be confirmed: C365S mutated full-length TMPRSS2 (construct 4) is not able to mature properly. The cysteine residue C365 (as well as C244) have been identified as crucial for TMPRSS2 maturation, possibly due to a stabilizing disulfide bond that is normally formed between those two cysteine residues.

By analyzing truncated TMPRSS2 variants, the cytosolic part, the transmembrane domain and the LDLRA domain have been identified as not being essential for autocatalytic processing. Truncated variants of constructs 8 and 11 show the activity associated 27 kDa-SPD signal in the respective cell lysates. In contrast, further removal of 22 amino acid residues downstream the LDLRA domain resulted in the loss of TMPRSS2 autocatalytic activity. All four cysteine residues of the SRCR domain, that form two disulfide bonds, seem to be important in the maturation process. The single-site mutation of cysteine residue C365 also resulted in the loss of autocatalytic cleavage activity in the case of all truncated TMPRSS2 variants (constructs 10

and 12) compared to the analogous truncated variants without the mutation (constructs 8 and 11, respectively). However random mutations in the *TMPRSS2* gene at position V149 and V479 (constructs 6 and 7 with each valine residue being substituted with an isoleucine) gave evidence that the applied mutating technique itself did not cause the loss of autocatalytic cleavage activity since these *TMPRSS2* variants showed both, autocatalytic cleavage activity as well as proteolytic cleavage activity towards HA. Furthermore, a positive control (full-length *TMPRSS2*, construct 1 in Table 5-1) was included in every experiment to support the credibility of the results and to exclude general experimental failure.

### **5.3.2. Cleavage activity towards H1- and H3-HA**

In the case of full-length *TMPRSS2*, autocatalytic cleavage activity was identified as an indicator for proteolytic cleavage activity towards influenza HA of the subtypes H1 and H3 in co-expression experiments. Whereas the mutated *TMPRSS2* variants *TMPRSS2\_V149I* and *TMPRSS2\_V479I* showed cleavage activity in parallel to the positive control, the *TMPRSS2\_C365S* variant was not able to cleave HA<sub>0</sub> of subtype H1 (in parallel to the inactive negative control: *TMPRSS2\_S441A*). Altering the cysteine residue C365, hence hampering the respective disulfide bond formation, led to the loss of *TMPRSS2* activity. In contrast, the method was not perfectly suitable for truncated *TMPRSS2* variants. Only marginal HA<sub>0</sub> cleavage was observed in the case of the truncated *TMPRSS2* (from construct 11) even though strong autocatalytic cleavage activity occurred. If the truncated *TMPRSS2* variants are only partly active after maturation or are simply not able to spatially interact adequately with the also membrane-bound hemagglutinin due to an altered intracellular location remains to be determined. A different activity test applicable to detect possible *TMPRSS2* activity of non-membrane-bound *TMPRSS2* within the cell would be helpful.

### **5.3.3. Localization of *TMPRSS2***

Compartmentation experiments of MDCK-*TMPRSS2* cells revealed that autocatalytically cleaved *TMPRSS2* is exclusively present in membrane-bound organelles. Only the non-glycosylated inactive *TMPRSS2* zymogen could be detected in very little amounts in the cytosol and no *TMPRSS2* is present in the cell nucleus. These experiments support previous suggestions that *TMPRSS2* is active within the TGN. Compartmentation was not applicable for cells that were transiently transfected with truncated *TMPRSS2*. Truncated *TMPRSS2* precipitated repeatedly during isolation. In addition, results from indirect immunofluorescence experiments did not show any obvious differences between the localization of truncated *TMPRSS2* variants in comparison to the membrane-anchored full-length *TMPRSS2*. Truncated *TMPRSS2* localization seems also limited to cell organelles surrounding the cell nucleus.

#### 5.3.4. Isolation of recombinant active TMPRSS2 for functional studies

Three different approaches were investigated during this thesis in order to develop an appropriate system for the isolation of active TMPRSS2.

First, truncated TMPRSS2 without the anchoring TM domain was used with the aim to isolate soluble active protein from the cell cytosol or in secreted form from cell culture medium. Even though truncated TMPRSS2 showed autoactivation activity, proteolytic activity could not be ultimately confirmed. The cell lysis with the detergents digitonin and igepal led to protein precipitation. Also, no truncated TMPRSS2 was detectable as a secreted form in conditioned cell culture medium nor in the cytoplasm. Truncated TMPRSS2 did not show any obvious dislocation compared to the membrane-anchored protease form. The isolation of truncated but active protein was not possible. In addition, the isolation and purification of His-tagged truncated TMPRSS2 from crude 293T cells using Ni-NTA was also not applicable.

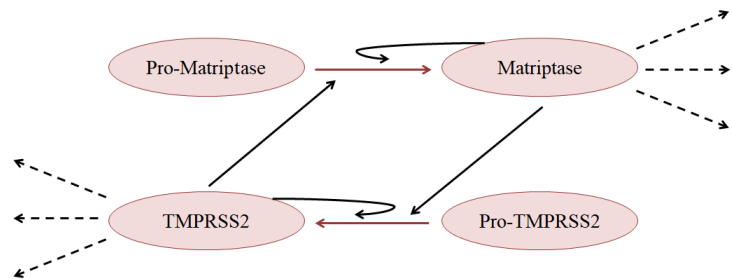
Second, soluble TMPRSS2 should be obtained from membrane-anchored TMPRSS2 with a single site mutation at cysteine residue C365. The hampered formation of the linking disulfide bond between SPD and plasma membrane in the case of C365S mutated TMPRSS2 did not result in active TMPRSS2 and therefore this approach was not of further interest.

Last, full-length TMPRSS2 was isolated via immunoprecipitation before and after compartmentation. While showing activity inside the cell (HA cleavage in co-expression experiments), no active TMPRSS2 could be isolated, when tested with chromogenic peptide substrates.

Under the given experimental conditions, none of the previous elaborated approaches to isolate active TMPRSS2 for enzyme kinetic and structural studies was suitable. Clearly, TMPRSS2 is more demanding than other serine proteases like matriptase or EK, which can easily be produced in large quantities using *E. coli* expression systems. To succeed in the future with active TMPRSS2 isolation, it is important to further characterize the protease and discover the needs for its stability outside the cell. This might include optimization of the environment, meaning buffer components, or the presence of stabilizing co-factors, such as a substrate or inhibitor.

#### 5.3.5. New insight into TMPRSS2 functionality

Although it was not possible to produce large amounts of active TMPRSS2, still new insight on TMPRSS2 functionality could be obtained during this thesis. This includes the identification of TMPRSS2 as a possible substrate of matriptase *in vitro* (as demonstrated in Figure 5-15) supporting TMPRSS2 involvement in possibly more than one cascade of proteolytic reactions.



**Figure 5-15: Matriptase can activate pro-TMPRSS2 *in vitro*.**

Furthermore, the importance of the TMPRSS2 stem region and single amino acids has been identified for TMPRSS2 activity. To highlight, in particular the SRCR domain and the disulfide bond between cysteine residues C244 and C365 seem essential for TMPRSS2 activity. Truncated as well as membrane-anchored TMPRSS2 was only found in cell organelles surrounding the cell nucleus. Obviously, truncated TMPRSS2 is not attached to the plasma-membrane anymore through the anchoring system, but it does not seem to be dislocated to other cell compartments in Huh-7 cells or even secreted. The failed isolation experiments of also full-length TMPRSS2 suggest, that TMPRSS2 might require a specific environment or specific stabilizing co-factors for activity outside the cell.



## 6. Conclusion and outlook

Influenza is one of the commonest infectious diseases affecting millions of people every year including 290,000 – 650,000 heavy casualties. Influenza viruses undergo constant genetic changes and every 10 – 50 years new influenza virus strains emerge that potentially cause a severe pandemic. In this modern interconnected world, experts believe the next influenza pandemic will be a “devastating global health event with far-reaching consequences”<sup>[1]</sup>. Novel effective anti-influenza drugs are in need. One strategy of influenza research is to focus on host-specific proteases that are essential for virus activation and spread. Trypsin-like serine proteases are crucial for influenza activation by mediating the cleavage of the viral surface glycoprotein HA and hence promoting the fusion potential of the virus. Therefore, their inhibition provides a promising therapeutic approach. The present work focused on the characterization of two relevant HA cleaving type-II transmembrane serine proteases matriptase and TMPRSS2.

### Matriptase characterization and inhibitor design

Chapter 3 and chapter 4 of this thesis engaged with the recombinant production of matriptase (chapter 3) in order to obtain pure functional enzyme of high quality for a SAR study with novel monobasic (hence potentially bioavailable) matriptase inhibitors of the 3-amidinophenylalanine type (chapter 4). Adequate amounts of high-quality matriptase enzyme were isolated using a new expression system and in total 5 matriptase crystals were available at the end of this thesis for structural analysis. The matriptase inhibitor design in this thesis focused on matriptase-affine compounds with a fair selectivity profile against the blood coagulation enzymes thrombin and fXa. In total, 18 new monobasic and potentially bioavailable, as well as four new dibasic compounds of the 3-amidinophenylalanine types were tested. Based on the last published crystal structure of this inhibitor type in complex with matriptase from 2006 (PDB code 2GV6) docking was used as structure-based virtual screening method for lead optimization of the compounds N-terminus. Selected compounds were suggested to interact with the carbonyl side chain of Gln175 of matriptase to achieve a higher affinity of matriptase compared to fXa.

The 4-*tert*-butylureido-piperidide could be identified as suitable C-terminus in combination with 3-fluoro-4-hydroxymethyl biphenylsulphonyl N-terminally in order to obtain excellent selectivity over thrombin. The binding mode of this compound (compound **55**) was crystallographically determined in complex with matriptase as well as trypsin. Trypsin proved as a suitable alternative to matriptase for detailed binding mode analysis of the compounds N-terminus. However, different preferences were detected for the C-terminus.

Dibasic compounds showed higher matriptase affinity and selectivity in comparison with the monobasic analogues. However, the tested monobasic compounds were still decent matriptase inhibitors that are additionally suitable for cell culture and animal studies in their benzamidine prodrug forms, which are well established from related inhibitors of thrombin. In addition, selected monobasic as well as dibasic compounds demonstrated strong suppression of the replication of certain H9N2 influenza viruses in a matriptase-expressing MDCK II cell model. These matriptase inhibitors could be potential lead structures for the development of new drugs against H9 strains for influenza.

### **TMPRSS2 characterization**

TMPRSS2 is widely discussed for its role in influenza activation. With a TMPRSS2 dependancy of HA-activation of certain subtypes, the characterization of this protease is an important prerequisite for being available as a target for influenza drug design. However, only little is known about the physiological function of TMPRSS2 and no experimental structure data are available at the moment to enable a structure-based drug development. Therefore, chapter 5 of this thesis focused on the characterization of TMPRSS2 in order to develop a strategy for the isolation of proteolytically active TMPRSS2 from cell culture. Even though, no functional TMPRSS2 could be recovered at the end of this work some new structural characteristics of TMPRSS2 were identified as crucial for functionality insight the cell. In general, TMPRSS2 without the cytosolic part, the transmembrane domain and the LDLRA domain is able to undergo autocatalytically activation if an artificial signal peptide was added N-terminal to enable entry into the endoplasmic reticulum. The presence of the cysteine-rich SRCR domain and the presence of the disulfide chain that connects the SPD and the stem region after activation cleavage have been identified as crucial for activity. N-terminal truncation of TMPRSS2 did not result in obvious dislocation within the cell: as the full-length positive control truncated TMPRSS2 was exclusively found in cell compartments surrounding the nucleus in immunofluorescence experiments. However, a reduced proteolytic cleavage activity towards H3-HA in co-expression experiments has been observed and might be a result of dislocation, since truncated TMPRSS2 is not bound to the biomembrane anymore. In addition, TMPRSS2 has been identified as a potential substrate of matriptase *in vitro*, which suggests possible participation in several zymogen cascades.



## 7. Experimental part

### 7.1. Materials and consumables

Unless otherwise stated, standard chemicals and solvents were purchased from the companies VWR, Fisher Scientific, Fluka, Sigma-Aldrich, Merck, or ROTH. Aqueous solutions of acids, bases or salts were prepared in deionized H<sub>2</sub>O, which was made by using a NOWA pure select system (KSN Water Technology, Nistertal). Sterile water was obtained by autoclaving deionized H<sub>2</sub>O for 20 min at 121 °C or was freshly sterile filtered using a 0.22 µm syringe filter.

#### 7.1.1. Consumables

Component	Supplier
6x DNA-Loading dye	Thermo Fisher Scientific, Waltham
Agar-agar	Roth, Karlsruhe
Agarose, SeaKem® LE	Lonza, Basel (CH)
Ammonium persulfate	ROTH, Karlsruhe
Ampicillin	ROTH, Karlsruhe
Benzamidine hydrochloride	Sigma-Aldrich, Taufkirchen
Bromphenol blue	Merck, Darmstadt
Cell culture flask (25 cm <sup>2</sup> , 75 cm <sup>2</sup> , 175 cm <sup>2</sup> )	Greiner Bio-One GmbH, Frickenhausen
Cell culture plates (Ø 10 cm)	Greiner Bio-One GmbH, Frickenhausen
Cell scraper	Greiner Bio-One GmbH, Frickenhausen
CellLytic M lysis buffer	Merck, Darmstadt
Chloramphenicol	ROTH, Karlsruhe
Coomassie Brilliant Blue R250	Merck, Darmstadt
Cover slides	Greiner Bio-One GmbH, Frickenhausen
Cuvettes	Sarstedt, Nuembrecht
DAPI	Institut of Virologie, Marburg University
Dialysis membrane	ROTH, Karlsruhe
Dimethyl sulfoxide (DMSO)	Sigma-Aldrich, Taufkirchen
DMEM	Gibco, Karlsruhe
Doxycycline	Clontech, Saint-Germain-en-Laye
DTT	ROTH, Karlsruhe
Erlenmeyer flasks, baffled (0,2 - 5 liter)	Schott, Mainz
FCS	Gibco, Karlsruhe
Fluoroshield™	Merck, Darmstadt
Geneticin (50 mg/ml)	Gibco, Karlsruhe
Glycerol	ROTH, Karlsruhe
GSH	ROTH, Karlsruhe
GSSG	ROTH, Karlsruhe
Guanidine-HCl	ROTH, Karlsruhe
HisPur™Ni-NTA Resin	Thermo Fisher Scientific, Waltham
IGEPAL® CA-360	Sigma-Aldrich, Taufkirchen
Imidazole	Sigma-Aldrich, Taufkirchen
IPTG	ROTH, Karlsruhe
Kanamycin	ROTH, Karlsruhe
L-Arginine	Sigma- Aldrich, Taufkirchen
L-Cystine	Merck, Darmstadt
L-Glutamine 200MM	Gibco, Karlsruhe
Lipofectamine™ 2000	Invitrogen, Karlsruhe

MicroSeedBeads™	Molecular Dimension, Maumee (Ohio)
Multi well plates (6, 12, 24, 96 wells)	Greiner Bio-One GmbH, Frickenhausen
Nunc™ Microwell™ 96-well microplates	Sigma-Aldrich, Taufkirchen
OptiMEM	Gibco, Karlsruhe
PEG 8000	Fluka, München
Penicillin-Streptomycin (each 5.000 IU/ml)	Gibco, Karlsruhe
Peptone	ROTH, Karlsruhe
Petri dishes	Merck, Darmstadt
Puromycin	InvivoGen, Toulouse (France)
PVDF-membrane	Millipore, Schwalbach
Roti®-Blue 5x	ROTH, Karlsruhe
Rotiphorese® 10x SDS-PAGE	ROTH, Karlsruhe
Rotiphorese® Gel 30	ROTH, Karlsruhe
Silicone grease (medium viscosity)	GE Bayer Silicones GmbH, Leverkusen
Skimmed milk powder	Saliter, Obergünzburg
Sterile filter	ROTH, Karlsruhe
SYRB® Safe DNA gel stain	Thermo Fisher Scientific, Waltham
TEMED	ROTH Karlsruhe
Triton™ X-100	Sigma-Aldrich, Taufkirchen
Tween® 20	Sigma-Aldrich, Taufkirchen
Urea	ROTH, Karlsruhe
Vivaflow® 200 crossflow cassette (MWCO: 10,000 Da)	Novagen, Darmstadt
Vivaspin Centrifugal Concentrator	Sartorius, Göttingen
Whatman™-filter paper	Sigma-Aldrich, Taufkirchen
Yeast extract	ROTH, Karlsruhe
β-Mercaptoethanol (ME)	Merck, Darmstadt

### 7.1.2. Enzymes, proteins molecular standards

Protein	Supplier
DNA Ligase (5 U/μl)	Thermo Fisher Scientific, Waltham
Benzonase	New England Biolabs (NEB), Ipswich
Bovine serum albumin (BSA)	ROTH, Karlsruhe
BSA 30 % solution	Sigma-Aldrich, Taufkirchen
EDTA-Trypsin	Gibco BRL, Karlsruhe
Endonuclease EcoRI (Fast digest)	Thermo Fisher Scientific, Waltham
Endonuclease NheI (Fast digest)	Thermo Fisher Scientific, Waltham
Endonuclease NotI (Fast digest)	Thermo Fisher Scientific, Waltham
Endonuclease XhoI (Fast digest)	Thermo Fisher Scientific, Waltham
Enteropeptidase (EK) (5xHis-tagged)	Kindly provided by Viktor Magdolen (TU Munich)
FastAP Thermosensitive Alkaline Phosphatase	Thermo Fisher Scientific, Waltham
TAGZymeDaPase Enzyme	QIAGEN, Hilden
Trypsin from bovine pancreas Type I (#T8003)	Sigma-Aldrich, Taufkirchen
Trypsin from porcine pancreas, lyophilized (8 U/mg)	Merck, Darmstadt
Thrombin from Bovine Plasma (#T4648)	Sigma-Aldrich, Taufkirchen
Factor Xa (human) 2530PL, 100 IU, 200.35 IU/mg	Enzyme Research, South Bend

#### Molecular standards (from Thermo Scientific Fisher)

- (A) PageRuler™ Prestained Protein Ladder
- (B) PageRuler™ Plus Prestained Protein Ladder
- (C) FastRuler™ Low range DNA Ladder, ready to use
- (D) FastRuler™ Middle range DNA Ladder, ready to use
- (E) FastRuler™ High range DNA Ladder, ready to use

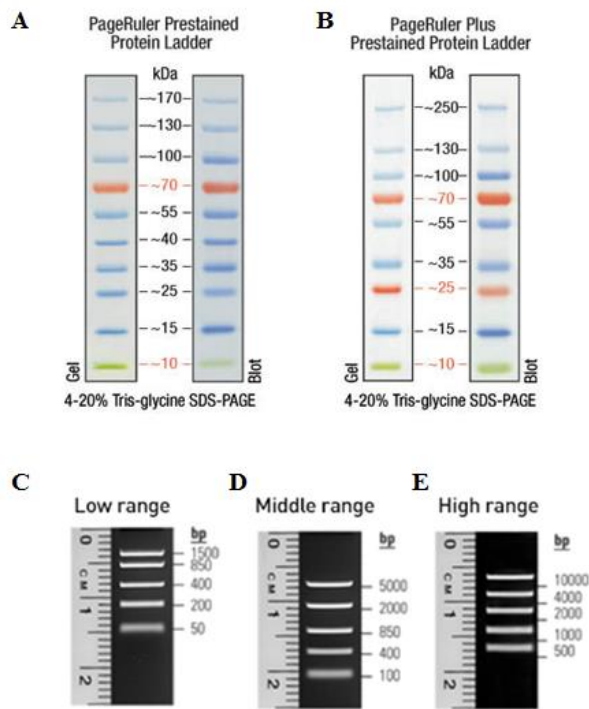


Figure 7-1: Molecular standards for SDS-PAGE (A-B) or agarose gels (C-E).

7.1.3. Primary and secondary antibodies

Primary antibodies		Working dilution
Rabbit $\alpha$ -H1 (pH1N1) (polyclonal)	Institute of Virology (Marburg University)	1:3000
Goat $\alpha$ -H3N2 influenza virus (polyclonal)	Antikoerper-online.de (Aachen)	1:6000
Rabbit $\alpha$ -H3N2 influenza virus (polyclonal)	Genetex. Irvine	1:2000
Rabbit $\alpha$ -FLAG® M2 (polyklonal)	Sigma-Aldrich, Taufkirchen	1:8000
Rabbit $\alpha$ -FLAG® (polyclonal)	Sigma-Aldrich, Taufkirchen	1:8000
Rat $\alpha$ -TMPRSS2 (polyclonal)	Institute of Virology (Marburg University)	1:4000 / 1:50
Mouse $\alpha$ -TMPRSS2 (polyclonal)	Institute of Virology (Marburg University)	1:3000
Mouse $\alpha$ -matriptase-SPD (monoclonal)	R&D Systems, Minneapolis	1:1000
Mouse $\alpha$ -6x-His Tag (monoclonal)	Thermo Fisher Scientific, Waltham	1:1000
Secondary antibodies		Working dilution
Goat $\alpha$ -rabbit IgG HRP conjugated	DAKO, Glostrup	1:8000
Rabbit $\alpha$ -goat IgG HRP conjugated	Sigma-Aldrich, Taufkirchen	1:8000
Donkey $\alpha$ -rat, Alexa Fluor® 488 conjugated	Life Technologies, Ober-Olm	1:150
Rabbit $\alpha$ -mouse, HRP conjugated	DAKO, Glostrup	1:8000

7.1.4. Bacterial strains

<i>E. coli</i> strain	Genotype	Supplier
Strains for cloning and plasmid preparation		
<i>JM109</i>	endA1 glnV44 thi-1 relA1 gyrA96 recA1 mcrB+ $\Delta$ (lac-proAB) e14- [F' traD36 proAB+ lacIq lacZ $\Delta$ M15] hsdR17(rK-mK+)	NEB
<i>TOP10</i>	F- ,mcrA, $\Delta$ (mrr-hsdRMS-mcrBC), $\phi$ 80lacZ $\Delta$ M15, $\Delta$ lacX74, deoR, recA1, araD139, $\Delta$ (ara-leu)7697, galU, galK, rpsL(StrR), endA1, nupG	Invitrogen
<i>XL2-Blue</i>	endA1 gyrA96(nalR) thi-1 recA1 relA1 lac glnV44 F'[:Tn10 proAB+ lacIq $\Delta$ (lacZ)M15 Amy CmR] hsdR17(rK- mK+)	Stratagene

<i>E. coli</i> strain	Genotype	Supplier
<i>XL10-Gold</i>	endA1 glnV44 recA1 thi-1 gyrA96 relA1 lac Hte Δ(mcrA)183 Δ(mcrCB-hsdSMR-mrr)173 tetR F'[proAB lacIqZΔM15 Tn10(TetR Amy CmR)]	Stratogene
Strains for protein expression		
<i>BL21 Gold (DE3)</i>	B F- ompThsdS(rB- mB-) dcm+ Tetr gal λ(DE3) endAHte	Stratogene
<i>BL21 CodonPlus (DE3)-RIL</i>	F- ompThsdS(rB- mB-) dcm+ Tetr gal λ(DE3) endAHte [argUileYleuWCamR]	Stratogene
<i>M15[pREP4]</i>	NalSStrSRifSThi- Lac- Ara+ Gal+ Mtl- F- RecA+ Uvr+ Lon+	QIAGEN

7.1.5. Cells lines

HEK 293T	Human embryonal kidney cells 293T, ATCC® number: CRL-11268
MDCK-H	Madin-Darby canine kidney cells ATCC®-CCL-34
MDCK-H TMPRSS2	Madin-Darby canine kidney cells induction of TMPRSS2 expression (Institute of Virology, Marburg University)
Huh-7	Human Hepato Cellular Carcinoma Cells ATTC®CCL-185

7.1.6. Plasmids

<b>Empty vectors</b>	
pCAGGS	Expression plasmid for eukaryotic cells, kindly provided by DR. MIYAZAKI <sup>[276]</sup> to the Institute of Virology (Marburg University)
pSecTag2/Hygro B	Expression plasmid for eukaryotic cells (Thermo Scientific), kindly provided by PROF. VIKTOR MAGDOLEN (TU Munich)
<b>TMPRSS2-plasmids</b>	
pCAGGS_TMPRSS2_FLAG	Plasmid for expression of TMPRSS2 with C-terminal FLAG epitope (GenBank: U75329) (Institute of Virology, Marburg University)
pCAGGS_S441A_TMPRSS2_FLAG	Plasmid for expression of inactive TMPRSS2 with S441A mutation and C-terminal FLAG epitope (Institute of Virology, Marburg University)
pTRE2pur-TMPRSS2-FLAG	Plasmid for expression of TMPRSS2 with C-terminal FLAG epitope (Institute of Virology, Marburg University)
pCAGGS_Δex5_TMPRSS2_FLAG	Plasmid for expression of LDLRA domain lacking TMPRSS2 (+V149I) mutation <sup>[236]</sup>
<b>Hemagglutinin-plasmids</b>	
pCAGGS-HA1	Plasmid for expression of the H1-HA of A/Hamburg/5/09 (H1N1) (GenBank: GQ166213.2) (Institute of Virology, Marburg University)
pCAGGS-HA3	Plasmid for expression of the H3-HA of A/HongKong/1/68 (H3N2) (GenBank: AF348176) (Institute of Virology, Marburg University)
<b>Matriptase-plasmids</b>	
pUC57_V1M_V2Y_C122S_matriptase	Synthesized by GeneCust (Ellange, Luxembourg)
pET24b_matriptase (wt)	Plasmid for expression of wt-matriptase SPD in <i>E. coli</i> <sup>[159]</sup>
pQE-30_C122S_matriptase	Plasmid for expression of C122S_matriptase SPD in <i>E. coli</i> , kindly provided by PROF. VIKTOR MAGDOLEN (TU Munich)

7.1.7. Oligonucleotides

Name	Sequence 5' – 3'
<b>Primer for cloning of truncated TMPRSS2 variants</b>	
deltaTM-EcoRI-f	CAGAATTCACTATGTGGAAGTTCATGGGCAGC
TMP2-ab A243-f	CAGAATTCACTATGGCCTGCGGGGTCAACTTG
NotI_deltaTM_f	CAGCGGCCGCTGGAAGTTCATGGGCAGCAAG
XhoI_IgK_dTM_f	CACTCGAGACTATGGAGACAGACACACTCCTG
EcoRI-TMPRSS2_f	CAGAATTCACCATGGCTTTGAACTCAGGG
NotI_dTM_dLDLRA_f	CAGCGGCCGCGTTCGCCTCTACGGACCAAAC
NotI_dSRCR_TMPRSS2_f	CAGCGGCCGCGCCTGCGGGGTCAACTTG
NotI_ab AS173_TMPRSS2_f	CAGCGGCCGCCAAGACGACTGGAACGAG

Name	Sequence 5' – 3'
NotI_ab AS186_TMPRSS2_f	CAGCGGCCGCAGGGACATGGGCTATAAG
NotI_abAS231_TMPRSS2_f	CAGCGGCCGCTCTTCAAAAGCAGTGG
TMP2-His-NotI-r	CGATGCGGCCGCTACCTAGTGGTGATGGTGATGGTGTCGCCGTCTGCCCTCAT
NotI-Flag TMPRSS2-r	CGATGCGGCCGCTACTTGTTCATCGTCATCCTTGTAGTCTCCGCCGTCTGCCCTCAT
TMPR_HIS_XhoI_r	CGATCTCGAGTACCTAGTGGTGATGGTGATGGTGTCGCCCGTCTGCCCTCAT
TMP2_His_NheI_r	GAGCTAGCTACCTAGTGGTGATGGTGATGGTGTCGCCCGTCTGCCCTCAT
deltaTM-start-r	GAAGTTTGGTCCGTAGAGGCGAACAC
NheI_FLAG_dTM_TMPRSS2_r	GAGCTAGCTACCTACTTGTTCATCGTCATCCTTGTAGTCTCCGCCGTCTGCCCTCAT
<b>Primer for site-directed mutagenesis of TMPRSS2</b>	
C365S_TMPRSS2_f	CCTAGTGAAACCAGTGTCTCTGCCCAACCCAGGC
C244S_TMPRSS2_f	CGCTGTATAGCCTCCGGGGTCAACTTG
V149I_TMPRSS2_f	GAGAATCGGTGTATTCGCCTCTACGG
V479I_TMPRSS2_f	GGAATGTGATGATTTTACGGACTGG
EKSubs-TMPRSS2_f	CAACTTGAACCTCAGACGACGATGACAAGATTGTGGGCGGCG
C365S_TMPRSS2_r	GCCTGGGTTGGGCAGAGACACTGGTTTCACTAGG
C244_TMPRSS2_r	CAAGTTGACCCCGGAGGCTATACAGCG
V149I_TMPRSS2_r	CCGTAGAGGCGAATACACCGATTCTC
V479I_TMPRSS2_r	CCAGTCCGTGAAAATCATCACATTCC
EKSubs-TMPRSS2_r	CGCCGCCCAACAATCTTGTTCATCGTCGTCTGAGTTCAAGTTG
<b>Primer for cloning of C122S matriptase</b>	
C122S_matriptase_nested1_f	ACAAGGTTGTGGGGGGGCACGGAT
C122S_matriptase_nested2_f	GATGACGATGACAAGGTTGTGGG
C122S_matriptase_nested3_f	TTGTTGGATCCGATGACGATGAC
C122S_matriptase_r	TTGTTAAGCTTATACCCCAGTGTT
<b>Primer for sequencing</b>	
pCAGGS_std_f	GC CTC TGC TAA CCA TGT TCA TGC C
pCAGGS_std_r	CAT ATG TCC TTC CGA GTG AGA GAC

All primer were customarily synthesized by Eurofins Genomic (Munich).

### 7.1.8. Cell culture medium

Growth medium (DMEM full medium)	1x Dulbeccos Modified Eagle Medium (DMEM) (500 ml)
+	10 % (v/v) fetal calf serum (FCS)
+	1 % (v/v) penicillin/streptomycin
+	1 % (v/v) L-glutamine
Transfection medium	OptiMEM
Selection medium	1x DMEM full medium
+	0.3 µg/ml (w/v) geneticin
+	0.2 µg (w/v) puromycin

### 7.1.9. Bacteria culture medium

All media were sterilized at 121 °C for 20 min. Antibiotics were only added after cooling down to 60 °C or immediately before inoculation with bacterial cells.

LB (Lysogeny Broth)-Medium	1 % bacto-Trypton
	0.5% yeast extrakt
	1% NaCl

LB/Amp-medium	LB medium with ampicillin (100µg/ml)
LB/Kan-medium	LB medium with kanamycin (30 µg/ml)
LB/Amp-agar plates	LB medium with 15 g/l bacto agar and ampicillin (100 µg/ml)

### 7.1.10. Commercially available reaction kits

CellTiter-Glo® 2.0 Cell Viability Assay	Promega, Mannheim
E.Z.N.A. Transfilter Plasmid Maxi Kit	Omega Bio Tek, Norcross
GeneJET Plasmid Maxiprep Kit	Thermo Scientific
Phusion High-Fidelity PCR Kit	ThermoFisher Scientific, Waltham
ProteoJET™ Cytoplasmic and Nuclear Protein Extraction Kit	ThermoFisher Scientific, Waltham
PureYield™ Plasmid Miniprep System	Promega, Mannheim
QuikChange Lightning Site-Directed Mutagenesis Kit	ThermoFisher Scientific, Waltham
Super Signal Dura Maximum Sensitivity Substrate	ThermoFisher Scientific, Waltham
Super Signal Femto Maximum Sensitivity Substrate	ThermoFisher Scientific, Waltham
Wizard® SV Gel and PCR Clean-Up System	Promega, Mannheim

### 7.1.11. Devices and special equipment

ÄKTA prime	GE Healthcare
Autoclave type FVA2	Fedegari
Centrifuge Avanti J-10 / J-25	Beckman Coulter
Centrifuge Biofuge fresco	Heraeus
ChemiDoc™ Imaging System	BioRad
Fluoroskan Ascent microplate reader	Thermo Scientific
HisTrap FF (+ crude) Ni-NTA column	GE Healthcare
Incubator shaker Innova 4.3	New Brunswick Scientific
Incubator shaker Innova 4230	New Brunswick Scientific
Microscope Axiovert 200M (fluorescence microscope)	Zeiss
Microscope SZ-PT (light microscope)	Olympus
Nanodrop 2000c	Thermo Scientific
thermo cycler (for PCR)	MJ Research
PD-10 desalting column (gravity flow)	GE Healthcare
Peristaltic pump P-1	GE Healthcare
pH-meter Five Easy™ FE20	Mettler Toledo
Pipettes (various)	Eppendorf
SDS-Page Mini-PROTEAN® Tetra Electrophoresis System	BioRad
Superdex™ 75 10/300 column for size-exclusion	GE Healthcare
Thermomixer Comfort 2ml	Eppendorf
Ultrasonic bath Elmasonic S 10/(H)	Elma
Ultrasonic Sonifier 250	Branson
Vortex mixer VF1	Janke & Kunkel
Balance type 572/45	Kern

### 7.1.12. Software and websites

Ascent Software	Thermo Fisher Scientific
Axio vision	Software for microscoping from ZEISS
Excel	Microsoft®
ExPASy applications	SIB Bioinformatics Resource Portal <sup>[277]</sup>
Fiji	image processing package ( <a href="http://fiji.sc/">http://fiji.sc/</a> )
ImageLab™ software	Bio Rad
MEROPS	Peptidase Database ( <a href="https://www.ebi.ac.uk/merops/">https://www.ebi.ac.uk/merops/</a> )
Origin 8.1	OriginLab® <a href="http://www.originlab.com">www.originlab.com</a>
PDB	Protein data bank <sup>[278]</sup>
PINGUI	<sup>[279]</sup> ( <a href="http://www.kolblab.org/scubidoo/pingui/">www.kolblab.org/scubidoo/pingui/</a> )
PyMol	Molecular Graphics System, Schrödinger
UniProt	<a href="https://www.uniprot.org/">https://www.uniprot.org/</a>

## 7.2. Molecular biological methods

Standard molecular biology methods were applied according to [280] and are not described in every detail.

### 7.2.1. Standard polymerase chain reaction (PCR)

Standard PCR reactions for cloning purposes within this thesis were performed by using the Phusion High-Fidelity PCR Kit following manufacturer's instructions.

Reaction mixture (50 µl)		Thermocycler protocol		Cycles
5xPhusion HF buffer (+MgCl <sub>2</sub> )	10 µl		T	t
dNTP mix, 10 mM	1 µl	Initial denaturation	98 °C	30 s
Forward primer (10 pmol/µl)	1.5 µl	Denaturation	98 °C	30 s
Reverse primer (10 pmol/µl)	1.5 µl	Primer annealing	65 °C	1 min
Template DNA	50-100 ng	Elongation	72 °C	2 min
Phusion DNA polymerase, 2 U/µL	1 µl	Final elongation	72 °C	10 min
H <sub>2</sub> O	to 50 µl	Storage	4 °C	∞

### 7.2.2. Site-directed mutagenesis using PCR

Single amino acid changes within the *TMPRSS2* gene were carried out using the QuikChange Lightning Site-Directed Mutagenesis Kit following the manufacturer instructions. Site directed mutagenesis was done after a 5-step protocol, starting with a PCR-reaction. After the digestion with DpnI nuclease, the plasmid was transformed into *E. coli*, isolated and finally analyzed by sequencing. Forward and reverse primer were designed with the triplet encoding for the to be exchanged amino acid being located in the middle and flanked by at least 10 template DNA complementary nucleotides on each site. The amount of 125 ng primer was used in every PCR reaction and calculated with the formula for ssDNA with 330 ng/nmol as the average molecular weight (MW) of a nucleotide:

$$\text{Equation (1)} \quad \text{x pmoles of oligo} = \frac{125 \text{ ng}}{330 \frac{\text{ng}}{\text{nmol}} \times \text{number of nucleotides}} \times 1000$$

A pTRE2 vector carrying the full-length *TMPRSS2* gene served as template DNA. PCR reactions were done in 50 µl volume following:

Reaction mixture		Thermocycler protocol		Cycles
10x QuikChange Lightning buffer	5 µl		T	t
dNTP mix	1 µl	Initial denaturation	95°C	30 s
QuikSolution reagent	1.5 µl	Denaturation	95°C	30 s
forward primer	125 ng	Primer annealing	60°C	1 min
reverse primer	125 ng	Elongation	68°C	6 min
pTET TMPRSS2	50-100 ng	Final elongation	68°C	2 min
QuikChange Lightning enzyme	1 µl	Storage	4°C	∞
H <sub>2</sub> O	to 50 µl			

After the PCR, 1 U DpnI enzyme was added to each reaction, gently mixed and incubated for 1 h at 37 °C. Since the template plasmid DNA was isolated from *dam*<sup>+</sup> *E. coli* strains, such as *XL2Blue*, *XL10Gold* or *JM109* strains, it is methylated. DpnI digests exclusively the methylated/hemimethylated parental template DNA and hence mainly mutated plasmid DNA was transformed into *E. coli* in the following step. Test tubes with 5 ml selective LB medium were each inoculated with a single colony and liquid cultures were grown overnight at 37 °C. Plasmid DNA was isolated using the PureYield™ Plasmid Miniprep System (Promega). The concentration and purity of isolated DNA was checked photometrically and the success of mutation was determined by sequencing (Eurofins Genomic).

7.2.3. Agarose gel electrophoresis

Electrophoresis through agarose is a standard technique to separate DNA molecules regarding their size. In this thesis, 1 % (w/v) agarose gels were prepared in 1 x TAE buffer and used to separate, identify and purify nucleic acid fragments. Samples in the FastDigest Green Buffer could be loaded directly. PCR-products and DNA samples digested with other enzymes than FastDigest enzymes were mixed with 1:5 volume of 6x Loading Dye prior loading. All agarose gels contained 1 µl of SYBR®Safe DNA gel stain to visualize the DNA under UV-light and were run in 1x TAE buffer. The ready-to-use Fast Ruler DNA Ladders (low, middle or high range, Figure 7-1) were used as molecular standards.

TAE buffer (1x)

40 mM Tris acetate, pH 8.0  
20 mM Acetic acid  
1 mM EDTA

7.2.4. DNA extraction and PCR product purification

The Wizard® SV Gel and PCR Clean-Up System (Promega) was used according to manufacturer’s instructions in order to purify PCR products or to extract DNA fragments from agarose gels. For the latter application, desired DNA band were cut out of the agarose gel and dissolved in 100-150 µl in the kit’s membrane binding buffer.



### 7.2.5. Restriction digestion of DNA

DNA digestion was carried out by restriction endonucleases following the manufacturer's instructions (7.1.2.) The linearization of vector DNA was done in a 40 µl reaction volume containing 1-2 µg plasmid DNA. The amplified insert DNA was digested in a 25 µl reaction volume containing 20.5 µl of the respective PCR product. In order to control the success of a ligation reaction, 100 - 300 ng of isolated plasmid DNA (from a single *E. coli* clone) was digested in a 10 µl volume and analyzed by agarose gel electrophoresis.

For digestions with FastDigest enzymes the 10x FastDigest Green Buffer was used for all combinations of enzymes. In case of others than FastDigest enzymes the buffers were chosen according to suggestions of the DoubleDigest Calculator of Thermo Scientific. FastDigest enzymes were incubated 30 min to 1 h and other restriction enzymes for 1 to 3 h at 37 °C, respectively. As a typical example, the following protocol shows the digestions of insert DNA and pCAGGS-MCS with EcoRI and NheI.

	<b>Insert</b> [µl]		<b>vector</b> [µl]
FastDigest Buffer (10x)	2.5	FastDigest Green Buffer (10x)	4
FastDigest EcoRI	1	FastDigest EcoRI	1
FastDigest NheI	1	FastDigest NheI	1
PCR-product	20.5	pCAGGS-MCS (680 ng/µl)	2.5
H <sub>2</sub> O	-	H <sub>2</sub> O	31.5

Incubation for 1 h at 37 °C

### 7.2.6. Treatment of DNA fragments with alkaline phosphatase

Linearized vector DNA was treated with alkaline phosphatase (FastAP, Thermo Scientific) to remove the 5' phosphate groups in order to lower the self-ligation rate. 1 µl FastAP was added to the restriction enzyme digestion mix after 30 min. Ligation of DNA fragments was performed using T4 ligase. For the ligation reaction, the digested insert DNA and the digested plus dephosphorylated vector DNA were mixed in a ratio of 5:1 in a reaction volume of 15 µl respectively. Samples were incubated overnight at 16 °C and further stored at 4 °C until transformation. Alternatively, samples were incubated for 2 h at room temperature with further incubation for at least 2 days at 4 °C.

T4 ligase buffer (10x)	1.5 µl
Insert:vector [5:1]	12.5 µl
T4 ligase	1 µl

### 7.2.7. Cloning of C122S-matriptase SPD

The pQE-30 plasmid carrying the C122S-matriptase construct was cloned in the facilities of PROF. VIKTOR MAGDOLEN (TU Munich) using the pUC57\_V1M\_V2Y\_C122S\_matriptase plasmid (custom synthesized by GeneCust) as template. In a three-step nested PCR reaction the 6xHis-tag and an EK-cleavage site were introduced upstream of the wild-type amino acids V16 and V17 using the respective primer as shown in 7.1.7.

### 7.2.8. Cloning of mutated full-length TMPRSS2 variants

For subcloning of the mutated full-length *TMPRSS2* genes into the pCAGGS-MCS mammalian expression vector a standard PCR reaction was performed using a sensed oligonucleotide introducing an EcoRI restriction site and the anti-sensed oligonucleotide to obtain a C-terminal FLAG tag and a NheI restriction site. PCR products and plasmid DNA were digested with EcoRI/NheI and purified with the Wizard<sup>®</sup> SV Gel and PCR Clean-Up System. The vector DNA was dephosphorylated, separated with agarose gel electrophoresis and extracted prior to purification. Following ligation reaction, the plasmid was transformed into *E. coli* and plated on selective LB/Amp agar plates. Several (5-8) overnight cultures per fragment were prepared by inoculating LB/Amp medium with single colonies. Plasmid DNA was extracted using the PureYield<sup>™</sup> Plasmid Miniprep System, digested with EcoRI/NheI and run in an agarose gel. Samples showing bands of the expected size of linearized vector DNA as well as insert DNA were sequenced using the standard pCAGGS sequencing oligonucleotides by Eurofins Genomic. Sequencing positive clones were re-transformed into *E. coli* and large amounts of plasmid DNA was prepared with the E.Z.N.A. Transfilter Plasmid Maxi Kit for transient transfection experiments.

### 7.2.9. Cloning of truncated TMPRSS2 variants

In order to produce truncated, but still active and correctly folded TMPRSS2, an artificial signal peptide was added N-terminally to ensure the entry into the secretory pathway of truncated TMPRSS2 variants. In this thesis, the secretion signal from the V-J2-C region of the mouse Ig kappa-chain (Igκ-SP, Igκ) was used. For this purpose, the respective *TMPRSS2* section was first introduced into the pSecTag2 HygroB plasmid harboring the Igκ-signal peptide upstream the multiple cloning site (MCS). In a second step, the *TMPRSS2* gene including the Igκ-SP N-terminally was amplified by a second PCR and subcloned into the pCAGGS-MCS vector as the expression vector of choice.

*Subcloning of truncated TMPRSS2 into pSecTag2 Hygro B*

In a first step, all truncated forms of TMPRRS were amplified by PCR using appropriate sensed primer to introduce NotI N-terminally and anti-sensed primer introducing a C-terminal 6xHis-tag and a XhoI restriction site. Fragments comprising of amino acids 106-492 ( $\Delta$ TM +  $\Delta$ TM\_S441A +  $\Delta$ TM\_C365S), 149-492 ( $\Delta$ LDLRA +  $\Delta$ LDLRA\_C365S), 173-492, 186-492, 232-492 and 243-492 ( $\Delta$ SRCR +  $\Delta$ SRCR\_C365S) were amplified by standard PCR using the pCAGGS-TMPRSS2\_FLAG plasmid as template and NotI\_deltaTM\_f, NotI\_dLDLRA\_f, NotI\_173-492\_f, NotI\_186-492\_f, NotI\_232-492\_f, NotI\_dSRCR\_f and TMPRSS2\_His\_XhoI\_r as oligonucleotides. Each fragment was digested with NotI and XhoI and purified with the Wizard® SV Gel and PCR Clean-Up System. The pSecTag2 HygroB plasmid DNA was linearized by NotI/XhoI digestion, dephosphorylated and purified with the aforementioned kit after agarose gel electrophoresis and subsequent gel extraction. After the ligation reaction, plasmid DNA was transformed into *E. coli* and plated on LB/Amp-agar plates. Several (5-8) overnight cultures per fragment were prepared by inoculating 5 ml LB/Amp-medium each with single colonies. Plasmid DNA was extracted using the PureYield™ Plasmid Miniprep System, digested with NotI and XhoI and analyzed by agarose gel electrophoresis. Samples showing bands of the expected size of linearized vector DNA and insert DNA were sequenced using the standard pCAGGs sequencing oligonucleotides (by Eurofins Genomic). Sequencing positive clones served as template DNA for the following subcloning into the pCAGGS-MCS expression vector.

*Subcloning of truncated Igk-TMPRSS2 into pCAGGS-MCS*

For subcloning the Igk containing truncated TMPRSS2 variant into the pCAGGS-MCS expression vector a standard PCR reaction was performed using a sensed oligonucleotide complementary to the Igk-SP sequence of the pSecTag2 HygroB plasmid. Sensed oligonucleotides introduce an XhoI restriction site and anti-sensed oligonucleotides were designed to either keep the 6xHis-tag C-terminally or substitute it with a FLAG-tag. In both cases, a NheI restriction site was introduced at the very end. PCR products and vector DNA were digested with XhoI plus NheI and purified with the Wizard® SV Gel and PCR Clean-Up System. The vector DNA was dephosphorylated, separated with gel electrophoresis and extracted prior to purification. After ligation, the plasmid DNA was transformed into *E. coli* and plated on LB/Amp-agar plates. Several (5-8) overnight cultures per fragment were prepared by inoculating 5 ml LB/Amp-medium with a single colony each. Plasmid DNA was extracted using the PureYield™ Plasmid Miniprep System, digested with XhoI and NheI and analyzed by agarose gel electrophoresis. Samples showing bands of the expected size of linearized vector

DNA as well as insert DNA were sequenced using the standard pCAGGs sequencing oligonucleotides (Eurofins Genomic). Sequencing positive clones were re-transformed into *E. coli* and large amounts of plasmid DNA was prepared with the E.Z.N.A. Transfilter Plasmid Maxi Kit for transient transfection experiments.

### **7.3. Microbiological methods**

#### **7.3.1. Growth and storage of bacterial strains**

Bacteria were grown at 37 °C on LB agar plates or in liquid LB medium that was supplemented for positive selection with appropriate antibiotics in the respective concentrations. All media were sterilized for 20 min at 121 °C. For permanent storage, cell strains carrying the plasmid of interest were grown overnight in selective LB medium and were flash frozen in liquid nitrogen after being mixed with glycerol at a final concentration of 25 %.

#### **7.3.2. Heat-shock based transformation of chemically competent *E. coli* cells**

For transformation, 20-40 µl of competent *E. coli* cells were thawed on ice and mixed with either 50-100 ng experimental DNA or 2-4 µl of ligation mixture. Tubes were incubated on ice for 10-30 min, heat-shocked for 30 s at 42 °C and immediately transferred on ice for at least 2 min. 760 – 780 µl LB medium at RT was added and the cells were incubated at 37 °C for 1 h with shaking at 350 rpm. Cells were pelleted by centrifugation for 1 min at 10.000 rpm. 700 µl LB medium was discarded and in the remaining liquid resuspended cells were plated on selective LB agar. Plates were incubated bottom up overnight at 37 °C and used for maximal three weeks stored at 4 °C.

#### **7.3.3. Recombinant production of matriptase in *E. coli***

Overnight cultures of 200 – 300 ml sterilized selective LB medium were inoculated with single colonies, containing the respective expression plasmid, and incubated at 37 °C overnight with continuous shaking. Alternatively, glycerol stocks were used for inoculation. The overnight cultures were diluted 1:40-1:50 into 6-12 liters fresh selective LB medium and were further incubated at 37 °C with vigorous shaking at 120-150 rpm. The protein expression was induced by the addition of 1 mM IPTG (from a freshly prepared 1 M IPTG stock) when OD<sub>600</sub> reached 0.6 – 0.9. Bacterial cells were harvested 4 h after induction by centrifugation at 4 °C and 8000 rpm for 15 min. If not processed immediately after harvesting, pellets were flash frozen in liquid nitrogen and stored at -80 °C until use. For expression of wild type matriptase-SPD a glycerol stock of *BL21(DE3) CodonPlus (DE3)-RIL* cells containing the pET24-matriptase

plasmid was used. The pQE-30-matriptase-C122S plasmid was expressed by *M15[pREP4]* cells after fresh transformation.

## 7.4. Cell culture methods

### 7.4.1. Thawing, freezing and passaging of cells

A frozen 1 ml cell suspension aliquot was shortly thawed in a water bath at 37 °C and mixed with 3 ml DMEM cell culture medium. After centrifugation for 5 min at 900 x g and RT the cell pellet was resuspended in 1 ml DMEM full medium and transferred to a T75 culture flask containing DMEM full medium. Cells were incubated at 37 °C and 5 % CO<sub>2</sub> until a confluency of 95-100 % was reached. For passaging cells, medium was aspirated and cells were washed with DMEM full medium once and detached by incubating with 2 ml trypsin/EDTA at 37 °C. After 5-15 min, the reaction was blocked by adding 8 ml of DMEM full medium and the resuspended cells were transferred into fresh DMEM full medium. For 1:10 dilution, 1 ml of cell suspension was transferred to a T75 flask. For freezing cells, detached cells were first centrifuged for 5 min at 900 x g. The cell pellet was resuspended in 4 ml DMEM full medium supplemented with 25 % DMSO as cryo protectant. 1 ml aliquots were frozen in cryo vials, by storing the vials in an isopropanol freezing container for 24 h at -80 °C before being finally transferred into a liquid nitrogen dewar for long-term storage.

### 7.4.2. Transient transfection of eukaryotic cells with Lipofectamine™ 2000

Lipofectamine™2000 is used for transfection of different cell lines with plasmid DNA. As a cationic lipid it forms a complex with negatively charged DNA molecules, which then enters the cellular membrane through endocytosis. Cells were seeded into appropriate well plates or cell culture flasks and transfected when the desired cell confluency was reached.

Cell line	Confluency for transfection [%]
HEK 293T	90 - 95
HUH-7	70 - 80
MDCK-H	70 - 80

For all transfections, the serum-free OptiMem was used as medium. Quantity of Lipofectamine2000 and plasmid DNA in up-scaling approaches were adjusted regarding cell number and surface area of the used culture vessel.

Culture vessel	Surface area per well <sup>1</sup>	Shared reagent		DNA	Lipofectamine 2000™
		Vol. of plating	Vol. of dilution		
12-well	4 cm <sup>2</sup>	1 ml	2 x 100 µl	1.6 µg	2 µl
6-well	10 cm <sup>2</sup>	2 ml	2 x 250 µl	4.0 µg	4 µl
10 cm	60 cm <sup>2</sup>	15 ml	2 x 1.5 ml	24 µg	60 µl
T75	75 cm <sup>2</sup>	15 ml	2 x 1.5 ml	36 µg	75 µl

<sup>1</sup> Surface areas may vary depending on the manufacturer.

Modified after: <https://www.thermofisher.com>.

The transfection procedure was performed as recommended by the manufacturer. The following protocol for the transfection in a 12-well format is to be seen as an example.

1. 1.6 µg plasmid DNA of each to be transfected sample was mixed with 100 µl OptiMEM in 1.5 ml reaction tubes
2. Lipofectamine Master-Mix (MM) was prepared (100 µl OptiMEM mixed with 2 µl Lipofectamine2000 reagent per number of transfections, e.g. 5 samples: 500 µl OptiMEM + 10 µl Lipofectamine2000)
3. 5 min incubation at room temperature of each tube individually
4. 100 µl Lipofectamine MM was added to each 100 µl OptiMEM/DNA sample
5. 20 min incubation at room temperature (Lipofectamine/DNA complex formation)
6. Growth medium from 293T cells was changed to 800 µl OptiMEM meanwhile
7. 200 µl OptiMEM containing the Lipofectamine/DNA complex was added (dropwise) to the corresponding well
8. Cells were incubated at 37 °C for 24-48 h with a medium change after 4-6 h post transfection to 1 ml DMEM full medium.

## 7.5. Biochemical methods

### 7.5.1. Purification and refolding of wt-matriptase from *BL2(DE3)CodonPlus RIL cells*

#### *Cell lysis and purification*

Cell pellets from 6 liters cultured LB medium were used per purification procedure. Frozen cells were thawed at RT and each pellet of 2 liters cell culture was resuspended in 40 ml IB-buffer containing 3 mM DTT. Cells were broken by sonication on ice for two times 4 min at 70 % output. To digest the nucleic acids, 1 µg/ml benzonase was added and the cell suspension was incubated for 30 min on ice. Inclusion bodies were harvested by centrifugation for 30 min at 20.000 rpm, washed twice with IB-buffer, and finally resuspended in 5 ml/g pellet of 6 M Guanidine lysis buffer A (pH 8.0). To complete denaturation, the suspension was incubated overnight at RT on the end-over-end rotator. Insoluble components were separated by centrifugation for 1 h at 20.000 rpm. The supernatant was loaded with a peristaltic pump at a flow rate of 1 ml/min onto a 5 ml Ni-NTA column, which had been equilibrated previously with the guanidine lysis buffer. The column was connected to the Äkta system and washed with Urea wash buffer A (pH 6.3) until a constant baseline signal was reached. Proteins were

eluted with Urea elution buffer A at a pH of 4.5 and fractions covering the elution peak were pooled and concentrated to at least 2.5 mg/ml. The pH was adjusted to 8.0 with 5 M NaOH and the protein solution was incubated overnight with 7.5 mM GSH and 0.75 mM GSSG at RT. To remove GSH and GSSG, protein solution was dialyzed against Urea dialysis Buffer (pH 5.0) at RT for 8 h with changing the buffer thrice.

<b>IB-buffer</b> (pH 7.5)	<b>Guanidine lysis buffer A</b> (pH 8.0)	<b>Urea wash buffer A</b> (pH 6.3)	<b>Urea elution buffer A</b> (pH 4.5)	<b>Urea dialysis buffer</b> (pH 5.0)
50 mM Tris-HCl	6 M Guanidine-HCl	8 M Urea	8 M Urea	6 M Urea
154 mM NaCl	0.01 M Tris-HCl	0.01 M Tris-HCl	0.01 M Tris-HCl	0.01 M Tris-HCl
3 mM DTT	0.1 M NaH <sub>2</sub> PO <sub>4</sub>	0.1 M NaH <sub>2</sub> PO <sub>4</sub>	0.1 M NaH <sub>2</sub> PO <sub>4</sub>	0.1 M NaH <sub>2</sub> PO <sub>4</sub>
	10 mM $\beta$ -ME	10 mM $\beta$ -ME	10 mM $\beta$ -ME	

#### *Refolding of denaturated proteins by rapid dilution*

The purified and denaturated proteins were renatured using the rapid dilution technique. Therefore, the suspension was dropwise diluted 50 – 100 times into ice-cold RD-Refolding buffer under moderate stirring. After incubation for 3 days at RT, the protein solution was filtered and concentrated to a final volume of approximately 100 ml with the Vivaflow®200 crossflow cassette (MWCO: 10.000 Da). Prior to the 6xHis-tag removal, and hence matriptase activation, buffer was exchanged to 1x TAGZyme buffer using PD10 columns under gravity flow. The protein concentration was adjusted to > 0.3 mg/ml.

<b>RD-Refolding buffer</b> (pH 7.5)	<b>TAGZyme buffer</b> (pH 7.0)
50 mM Tris-HCl	20 mM NaH <sub>2</sub> PO <sub>4</sub>
500 mM L-Arginin	150 mM NaCl
100 mM NaCl	
20 mM CaCl <sub>2</sub>	
1 mM EDTA	
0.5 mM Cystein	

#### *Activation of MK(H)<sub>6</sub>-matriptase by the dipeptidyl aminopeptidase-1 (DAPase-I)*

For the 6xHis-tag removal, 2.5 mU TAGZyme™ DAPase™ (QIAGEN, Hilden) was used per 50  $\mu$ g of protein. First, DAPase was diluted to 1 U/ml in TAGZyme buffer, mixed 1:2 with 2 mM cysteamine-HCl (part of the kit) and activated for 1 h at 37 °C. Respective amount of activated DAPase was mixed with the matriptase solution and incubated for 2.5 h at 30 °C. The recombinant C-terminally His-tagged DAPase removes dipeptides sequentially from the N-terminus of proteins. Some amino acids in P1' or P2' position stop or strongly inhibit cleavage. In case of matriptase only the N-terminal His-tag will be removed, since amino acid residues with hydrophobic side chains (such as valine residues 615 and 616) inhibit DAPase activity.

Precipitates were removed by centrifugation and filtration. The clear protein solution was checked visually for protease activity by incubation with the chromogenic Mes-DCha-Gly-Arg-pNA and was further stored at 4 °C for 16 h. His-tagged DAPase and non-activated matriptase were removed via reverse IMAC. Protein solution was therefore mixed 1:1 with Ni-NTA lysis buffer and pumped over a Ni-NTA column. While DAPase and non-activated matriptase bind to the Ni-NTA resin of the column, the flow through contains the active protease. The volume of active matriptase was reduced to < 5ml by concentration and subsequently further purified by gel filtration using a Superdex®75 with 1x PBS pH 7.5 as running solution. Fractions of active matriptase were pooled and stored at a concentration of 0.3 mg/ml at -80 °C. Bound proteins on the Ni-NTA column were washed and eluted with Ni-NTA elution buffer to allow column reuse.

Ni-NTA lysis buffer (pH 8.0)	Ni-NTA wash buffer (pH 6.3)	Ni-NTA elution buffer (pH 4.5)
50 mM NaH <sub>2</sub> PO <sub>4</sub>	50 mM NaH <sub>2</sub> PO <sub>4</sub>	50 mM NaH <sub>2</sub> PO <sub>4</sub>
300 mM NaCl	150 mM NaCl	150 mM NaCl
10 mM Imidazole		

### 7.5.2. Purification and refolding of C122S-matriptase from *M15 cells*

#### *Cell lysis and purification*

Cell pellets from 8-12 liters cultured medium were used per purification procedure. Frozen cells were defrosted at RT and each pellet of 2-3 liters cell culture was resuspended in 40 ml of 6 M Guanidine lysis buffer B containing 1 mM β-ME. To complete cell lysis and inclusion body denaturation, the suspension was incubated overnight at RT on the end-over-end rotator. Insoluble components were separated by centrifugation for 1 h at 20.000 rpm. To reduce viscosity, the supernatant was filtered through a 0.8 µm syringe filter or was passed several times through a gauche needle before being loaded onto Ni-NTA column with a peristaltic pump at 1-2 ml/min. The column was washed with minimum 5 column volumes of 8 M Urea lysis buffer before being connected to the Äkta system. The column was washed with Urea wash buffer B until a constant baseline signal was reached and proteins were eluted with Urea elution buffer B at pH 4.5. The fractions covering the elution peak were pooled and diluted with Urea wash buffer B to a protein concentration of 0.1-0.5 mg/ml or a maximum volume of 100 ml. To ensure reduction of all cysteine residues prior refolding, 20 mM DTT was added. The protein solution was incubated overnight at RT on the end-over-end rotator or for at least 3 h before starting the refolding process.



<b>Guanidine lysis buffer B</b> (pH 8.0)	<b>Urea lysis buffer</b> (pH 8.0)	<b>Urea wash buffer B</b> (pH 6.3)	<b>Urea elution buffer B</b> (pH 4.5)
6 M Guanidine-HCl	8 M Urea	8 M Urea	8 M Urea
0.01 M Tris-HCl	0.01 M Tris-HCl	0.01 M Tris-HCl	0.01 M Tris-HCl
0.1 M NaH <sub>2</sub> PO <sub>4</sub>	0.1 M NaH <sub>2</sub> PO <sub>4</sub>	0.1 M NaH <sub>2</sub> PO <sub>4</sub>	0.1 M NaH <sub>2</sub> PO <sub>4</sub>
1 mM $\beta$ -ME	1 mM $\beta$ -ME	1 mM $\beta$ -ME	1 mM $\beta$ -ME

### *Refolding of denaturized proteins by step wise dialysis*

Denaturized and purified proteins were refolded via stepwise dialysis. The protein solution was transferred into a dialysis tube with a molecular weight cut off (MWCO) of 10 – 12 kDa and dialyzed in different refolding buffers with decreasing urea concentrations at 4 °C. For the initial dialysis step, the protein solution was incubated in 3 liters Dialysis buffer I (pH 7.5), containing 4 M urea, for at least 6 h (or overnight). The dialysis tube was next incubated in 3 liters of 2 M urea containing Dialysis buffer II. This buffer was supplemented with GSH and GSSG in a ratio of 10:1 (2.5 mM / 0.25 mM) as oxido-shuffling system to support the correct forming of disulfide bridges during refolding. After 2 days, the dialysis tube was transferred to 3 liters of 1 M urea containing dialysis Dialysis buffer III, supplemented with 1.25 mM GSH and 0.125 mM GSSG, and was further incubated for 2 days at 4 °C before being transferred to 4 liters of the urea free Dialysis buffer IV. Protein solution was dialyzed overnight with a buffer exchange after 4 h. The refolded protein was concentrated to a volume of 50 – 70 ml or a maximum protein concentration of 1 mg/ml.

<b>Dialysis buffer I</b> (pH 7.5)	<b>Dialysis buffer II</b> (pH 7.5)	<b>Dialysis buffer III</b> (pH 7.5)	<b>Dialysis buffer IV</b> (pH 7.8)
4 M Urea	2 M Urea	1 M Urea	100 mM NaCl
0.25 M L-Arginin	0.5 M L-Arginin	0.25 M L-Arginin	50 mM Tris-Cl
100 mM NaCl	100 mM NaCl	100 mM NaCl	0.005 % Tween 20
50 mM Tris-HCl	50 mM Tris-HCl	50 mM Tris-Cl	
20 mM CaCl <sub>2</sub>	20 mM CaCl <sub>2</sub>	20 mM CaCl <sub>2</sub>	
1 mM EDTA	1 mM EDTA	1 mM EDTA	
0.005 % Tween 20	0.005 % Tween 20	0.005 % Tween 20	
	2.5 mM GSH	1.5 mM GSH	
	0.25 mM GSSG	0.15 mM GSSG	

### *Activation of MRGS(H)<sub>6</sub>GSDDDDK-C122S-matriptase by enteropeptidase (EK)*

For activation of C122S-matriptase, the protein solution was mixed with 5xHis-tagged EK (provided by Viktor Magdolen (TU Munich, Germany) and produced by Wolfgang Skala (following Skala *et al.*, 2013)) in a matriptase/enteropeptidase ratio of 500-700 : 1. The solution was incubated overnight at RT on the end-over-end rotator. Precipitate was removed by

filtration and the clear protein solution was checked visually for activity by incubation with the chromogenic substrate Mes-DCha-Gly-Arg-pNA. The activated matriptase was separated from the non-activated protein as well as 5xHis-tagged EK by reverse IMAC. The protein solution was loaded onto a 5 ml Ni-NTA column and the flow through containing the activated protein was collected and concentrated to a volume smaller than 5 ml. C122S-matriptase was further purified by gel filtration using a Superdex<sup>®</sup>75 which was run with Storage buffer (C122S-matriptase) at RT. Active matriptase of 0.28 mg/ml in C122S-matriptase storage buffer was flash frozen in liquid nitrogen and stored at -80 °C. For crystallization experiments, the freshly produced matriptase was concentrated to 5 - 10 mg/ml or flash frozen at ~ 1 mg/ml for later experiments.

**Storage buffer (C122S-matriptase)**  
(pH 7.8)

100 mM NaCl  
50 mM Tris-Cl  
5 % Glycerol (v/v)

### 7.5.3. Isolation of plasmid DNA from *E. coli*

Plasmid DNA was isolated from *E. coli* cells grown overnight in the LB/Amp medium. Routine plasmid isolation was carried out by alkaline lysis method using the PureYield<sup>™</sup> Plasmid Miniprep System (Promega) or the E.Z.N.A. Transfilter Plasmid Maxi Kit (Omega). DNA was isolated and purified via anion exchange resin according to manufacturer's instructions respectively.

#### *Miniprep (PureYield<sup>™</sup> Plasmid Miniprep System)*

Small amounts of DNA that were used as template DNA for cloning purposes as well as for sequencing experiments were isolated from 2-4 ml overnight cultures and was eluted into 30 – 40 µl H<sub>2</sub>O.

#### *Maxiprep (E.Z.N.A. Transfilter Plasmid Maxi Kit)*

For transient expression experiments, larger DNA amounts were required. Therefore, 200-600 ml overnight cultures were used to isolate DNA, which was finally eluted with a volume of 500 – 1000 µl H<sub>2</sub>O each.

#### 7.5.4. Ultraviolet-Visible (UV-Vis) Spectroscopy using Nanodrop™2000c

A nanodrop™2000 spectrophotometer (Thermo Scientific) was used following the manufacturer's manual to measure quantity and purity of nucleic acids and the concentrations of proteins as well as bacterial culture density following the Beer-Lambert law respectively.

##### *Determination of quantity and purity of total DNA*

Nucleic acids absorb UV light due to the heterocyclic rings of the nucleotides with a maximum absorption at 260 nm. For quantity/quality determination of the DNA preparation, 2 µl were loaded onto the nanodrop's pedestal (after zeroizing with H<sub>2</sub>O) and measured at wavelengths of 260 and 280 nm. At a concentration of 50 µg/ml and a 1 cm path length, dsDNA has an OD<sub>260</sub> of 1. Based on this assumption, DNA is quantified and recorded by the nanodrop software as ng/µl automatically. For purity assessment the A<sub>260</sub>/A<sub>280</sub> ratio is commonly used, with A<sub>260</sub>/A<sub>280</sub> = 1.8 considering the DNA sample pure from contaminations, such as proteins. DNA samples with A<sub>260</sub>/A<sub>280</sub> of 1.8 - 2.0 were considered sufficiently pure for further experiments.

##### *Determination of protein concentration*

Protein concentration is measured using the Protein A<sub>280</sub> method, which is applicable to purified proteins that contain the amino acids Trp and Tyr or Cys-Cys disulfide bonds and therefore exhibit absorbance at 280 nm. After blank measurement with the protein's storage buffer, 2 µl protein sample were loaded to the nanodrop's pedestal and measured at wavelengths 230 nm, 260 nm and 280 nm. The chosen sample type was "Other protein (ε+ MW) ", with the extinction coefficient/1000 (ε) and molecular weight (MW) in kilodaltons being calculated with the ExPASy ProtParam tool based on the protein amino acid sequence.

	before activation MW in [kDa]	after activation MW in [kDa]	extinction coefficient / 1000 [(mg/ml) <sup>-1</sup> cm <sup>-1</sup> ]
wt-matriptase	28.41	27.50	53.32
C122S-matriptase	27.50	27.50	53.32

##### *Determination of bacterial concentration in culture (OD<sub>600</sub> measurement)*

The number of microbial cells were determined by measuring bacterial culture density at 600 nm using the cuvette option. Sterilized LB medium served as blank.

### 7.5.5. Preparation of 293T cell lysates for western blotting

Cell lysis prior to western blot analysis was performed by sonication directly in 1x SDS buffer or chemically by using the CelLytic M lysis buffer (Merck). 293T cells were therefore harvested by resuspension in the growth medium 48 h after transfection with subsequent centrifugation for 5 min at 8.000 rpm.

#### *Cell lysis by sonication in 1x SDS sample buffer*

Transiently transfected cells were washed in PBS<sub>def</sub> once, pelleted and finally resuspended in an adequate volume of PBS<sub>def</sub> (30 - 60 µl), depending on the pellet size. Cells were mixed with an equal volume of 2x SDS sample buffer and sonified for 1 min.

#### *Cell lysis with CelLyticM buffer*

Transiently transfected 293T cells were washed in PBS<sub>def</sub>, pelleted and resuspended in an adequate volume of CelLyticM lysis buffer (50 – 80 µl). After 30 min incubation on ice with light shaking, cell debris was pelleted by centrifugation for 5 min at 13.000 rpm at 4 °C. 50 µl of supernatant was transferred to a fresh reaction tube and mixed with 10 µl of 6x SDS sample buffer supplemented with DTT.

#### **1x PBS<sub>def</sub> (pH 7.5)**

8 g NaCl  
0.2 g KCl  
1.15 g Na<sub>2</sub>HPO<sub>4</sub>  
0.2 g KH<sub>2</sub>PO<sub>4</sub>  
Add 1 liter H<sub>2</sub>O

#### **2x SDS Buffer (+β-ME)**

1.2 ml 1 M Tris-HCl (pH 6.8)  
4 ml 10 % SDS  
2 ml Glycerol  
0.02 % (w/v) Bromophenol blue  
0.2 ml β-ME  
2.6 ml H<sub>2</sub>O  
Aliquots frozen at -20 °C

#### **6x SDS Buffer (+DTT)**

1.2 ml 1 M Tris-HCl (pH 6.8)  
1.2 g SDS  
4.7 ml Glycerol  
6 mg Bromophenol blue  
0.93 g DTT  
2.1 ml H<sub>2</sub>O  
Aliquots frozen at -20 °C

### 7.5.6. Sodiumdodecylsulfat-polyacrylamid gelelectrophoresis (SDS-PAGE)

In this thesis, SDS-PAGE was used to separate proteins before western blotting or to confirm presence and purity of samples during protein purification procedures. Either way, lysates or protein solutions were mixed with 2x - 6x SDS buffer (to a final concentration: 1x) and were separated in 12 % - 15 % polyacrylamide gels. Prior loading, samples were incubated at 95 °C for 5 – 20 min (depending on sample viscosity) and centrifuged for 5 min at 13.000 rpm.

#### *SDS-PAGE for western blotting*

For western blotting purposes polyacrylamide gels made of a 4.4 % stacking gel and 12 % separation gel were prepared as followed:

stacking gel (4.4%)		seperation gel (12%)	
0.5 M Tris-HCl (pH 6.8)	1.3 ml	1.5 M Tris-HCl (pH 8.8)	2.5 ml
Rotiphorese®Gel30	750 µl	Rotiphorese®Gel30	4 ml
dH <sub>2</sub> O	2.9 ml	dH <sub>2</sub> O	3.3 ml
10 % SDS	100 µl	10 % SDS	100 µl
10% APS	60 µl	10% APS	100 µl
TEMED	10 µl	TEMED	10 µl

10 µl – 20 µl of protein samples in 1x SDS buffer were loaded onto the gels and proteins were separated beginning with 10 min at 100 V and completed by running 40 – 50 min at 150 V. Western blotting was performed as described below.

#### *SDS-PAGE for validation of protein expression or purification*

In order to check presence or purity of proteins, 15 % separation gels were used. Three portions of protein samples were diluted in one portion 4x SDS sample buffer (1x final concentration), incubated for 5 min at 95 °C, separated at 120 – 160 V for 60 – 90 min and finally visualized by either staining at RT in Coomassie staining solution for 1 h or in a 1x RotiBlue Solution overnight. Gels were destained in the corresponding destaining solution to remove background color of the gel and photographed for analysis (smartphone camera).

4x SDS sample buffer	Coomassie staining solution	Coomassie destaining solution
2 ml 1M Tris-HCl (pH 6.8)	40 % Methanol	20 % Methanol
0.8 g SDS	10 % Glacial acetic acid	10 % Glacial acetic acid
4 ml Glycerol	0.1 % (1g/L) Coomassie R250	
0.4 ml β-ME		
4 mg Bromophenol blue	<b>1x RotiBlue solution (1L)</b>	<b>RotiBlue destaining solution</b>
H <sub>2</sub> O to 10 ml	20 % 5x RotiBlue Solution	25 % Methanol

#### **7.5.7. Immunodetection by western blotting**

The sensitive detection of specific proteins from crude cell lysates, supernatants or purified protein samples was performed by western blotting and immunostaining using the “semi-dry” blotting technique <sup>[281]</sup>. Proteins were first separated by SDS-PAGE and subsequently transferred to a polyvinylidene difluoride membrane (PVDF) using a semi-dry blotter (Bio-Rad, Munich). Prior to blotting, the membrane was activated for 30 s in methanol and subsequently incubated in Transfer buffer in parallel with the SDS-gel and six Whatman™-filter paper per membrane. The proteins were transferred at 40 mA per membrane for 70 min. After the transfer, the membrane was immediately incubated for at least 1 h in WB blocking buffer to saturate unspecific binding sites. Primary antibody was added and incubation took place overnight at 4 °C with smooth movement. After being washed three times for 10 min in PBS/Tween, the membrane was incubated with the respective HRP conjugated secondary

antibody in WB blocking buffer for 1 h at RT and finally washed thrice with PBS/Tween as described above. Antibody dilutions were used as described in section 7.1.3. Detection was performed using the enhanced chemiluminescence (ECL) method with SuperSignal™ WestFemto or SuperSignal™ West Dura as ECL-substrates according to manufacturer's instructions. Blots were incubated for at least 5 minutes with the substrate in darkness before signals were detected in the ChemiDoc™ Imaging System using the ImageLab™ software.

Transfer buffer	WB blocking buffer	PBS/Tween
5.8 g Tris base	1x PBS	1x PBS
2.9 g Glycine	0.1 % Tween-80	0.1 % Tween-80
200 ml Ethanol	7 % Skimmed milk	
fill to 1 liter with H <sub>2</sub> O		

**7.5.8. Immunofluorescence of transiently transfected Huh-7 cells**

*Fixation, permeabilization and blocking of Huh-7 cells*

For fixation of transiently transfected Huh-7 cells (grown on cover slides in 24-well plates), the medium was removed carefully 24 h after transfection and the cells were washed with ice cold 1x PBS<sub>def</sub> twice. Cells were coated with ice cold acetone/methanol mixture (ratio 1:1) and incubated on ice for 20 min. The acetone/methanol solution was removed, the cells washed two times with 1x PBS<sub>def</sub> and incubated for 1-2 h with IF blocking buffer.

IF blocking buffer
1x PBS
2 % BSA

*Incubation with primary and secondary antibody*

Cells were incubated with the primary rat  $\alpha$ -TMPRSS2 antibody in a 1:50 dilution in IF-blocking buffer in a dark and wet environment. After 1 h, cells were washed with 1x PBS<sub>def</sub> thrice and subsequently incubated 1 h with a mixture of an  $\alpha$ -rat Alexa-488 conjugated secondary antibody in a 1:150 dilution and 1:50 diluted DAPI both in IF blocking buffer in the dark wet environment. Cells were washed with 1x PBS<sub>def</sub> twice and once with H<sub>2</sub>O before being covered with Fluroshield™ mounting medium on a microscope slide. Cells were incubated at RT in the dark overnight, analyzed and photographed with the camera enclosed to the ApoTome fluorescence microscope (Zeiss) and further stored at 4 °C.

**7.5.9. Compartmentation of stably and transiently transfected MDCK-H cells**

Compartmentation was performed using the buffer supplements digitonin and igepal <sup>[275]</sup> to fractionate proteins from different cell compartments or by using the ProteoJET™ Cytoplasmic and Nuclear Protein Extraction Kit (Thermo Scientific) following manufacturer's instructions.

MDCK-TMPRSS2 cells or transiently transfected MDCK-H cells with truncated constructs were harvested carefully 48 h after induction or transfection with a cell scraper and washed twice with chilled 1x PBS<sub>def</sub>. All steps were performed on ice or 4 °C with pre-chilled buffers.

#### *Compartmentation using digitonin and igepal*

Fractionizing cell compartments by sequential lysis of cell membranes was performed after the method of BAGHIROVA and team <sup>[275]</sup> (as shown in Figure 7-2). Cytosolic proteins were isolated by using a digitonin supplemented buffer. The steroidal saponin binds with cholesterol and other  $\beta$ -hydroxysterols and hence permeabilizes the cholesterol rich plasma membrane by pore formation, but not the cholesterol poor membranes of cellular organelles <sup>[282]</sup>. In a second step, proteins of membrane bound organelles were isolated by using low concentrations of the non-ionic and non-denaturing detergent igepal to permeabilize the membrane of the ER, Golgi and mitochondrion while keeping the nucleus membrane intact. Proteins from 1/2 T75 culture flask or from a 10 cm dish were isolated by using 400  $\mu$ l of Buffer A and B respectively. Each isolated fraction was analyzed by western blotting or further used for immunoprecipitation or treatment of H1-HA transfected 293T cells.

Buffer A	Buffer B
150 mM NaCl	150 mM NaCl
50 mM HEPES (pH 7.4)	50 mM HEPES (pH 7.4)
25 $\mu$ g/ml Digitonin	1 % (v/v) Igepal
1 M Hexlene glycol	1 M Hexlene glycol

#### *Compartmentation using the ProteoJET™ Cytoplasmic and Nuclear Protein Extraction Kit*

Cells of 1/2 T75 culture flash were used to separate cytoplasmic and nuclear proteins by following the manufacturer's instruction. All buffers were supplemented as recommended with 0.1 M DTT, but not treated with protease inhibitors. The cytoplasmic fraction contained cytosolic, as well as proteins from membrane-bound organelles.

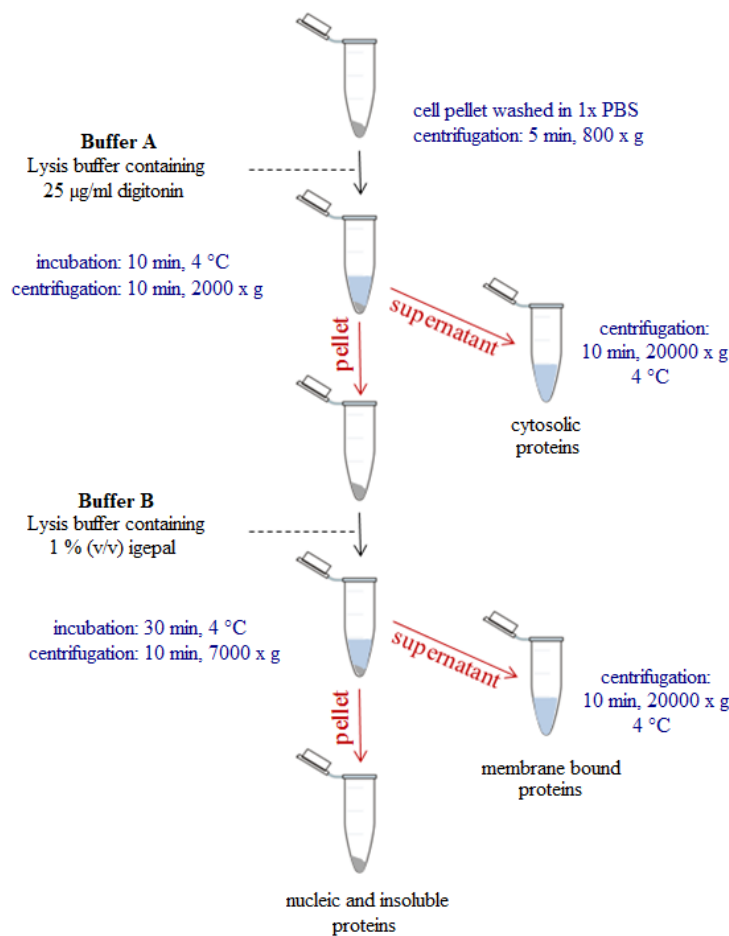


Figure 7-2: Cell compartmentation using digitonin and igepal containing lysis buffers.

7.5.10. Immunoprecipitation using the ANTI-FLAG M2 Affinity Gel

FLAG-tagged TMPRSS2 was isolated from stably transfected MDCK-TMPRSS2 cells 48 h post induction with doxycyclin according to manufacturer's instructions (Merck). Cell lysate was incubated with equilibrated ANTI-FLAG® M2 Affinity Gel overnight at 4 °C on the end-over-end rotator. The next day, the affinity gel was washed with TBS thrice, and bound proteins were eluted by incubation with 0.1 M glycine (pH 3.0). After 5 min incubation at RT, affinity gel was separated by centrifugation and the supernatant was directly mixed 1:10 with 10x TBS to neutralize the low pH glycine. Eluted proteins were immediately visually checked for activity by incubating with the chromogenic substrate Mes-DCha-Gly-Arg-pNA. Alternatively, bound proteins were eluted directly by incubation in 2x SDS buffer for 10 min at 95 °C and analyzed via SDS-PAGE and western blotting.

10x TBS	1x TBS
500 mM Tris-HCl (pH 7.5)	50 mM Tris-HCl (pH 7.5)
1.5 M NaCl	150 mM NaCl



### 7.5.11. Enzyme kinetic measurements

All enzyme kinetic measurements were performed using synthetic fluorogenic 7-amino-4-methylcoumarin (AMC) substrates at  $\lambda_{\text{ex}} = 355$  nm and  $\lambda_{\text{em}} = 460$  nm in a Fluoroskan Ascent reader (Thermo Scientific). All substrates were synthesized in house and prepared as a 10 mM stock solution in 100 % ultrapure water or in case of lower soluble compounds in 50 % DMSO. Stock solutions of synthesized inhibitors were prepared in a concentration of 10 mM in 100 % DMSO, which were subsequently diluted with Enzyme kinetic buffer to appropriate final concentrations. The lowest inhibitor concentration in the assay was chosen to be at least 10-fold higher than the used enzyme concentration. The reaction volume of 140  $\mu\text{l}$  consisted of 100  $\mu\text{l}$  Enzyme kinetic buffer (with or without respective inhibitor concentration), 20  $\mu\text{l}$  substrate and 20  $\mu\text{l}$  enzyme solution. After adding the enzyme, the substrate turnover was detected at RT in 15 s intervals for 10-20 min by measuring the fluorescence increase. The resulting steady-state rates were calculated using the Ascent software, inhibition constants were calculated with Excel or Origin 8.1. All enzymes were diluted in the enzyme dilution buffer EVP (*“Enzymverdünnungspuffer”*) to reach a substrate turnover of at least 100 RFU after 10 min of measurement.

EVP	Enzyme kinetic buffer
154 mM NaCl	50 mM Tris-HCl (pH 8.0)
0.1 % BSA	154 mM NaCl
	0.01 % Triton X-100

**Table 7-1: Used substrates and enzymes for enzyme kinetic measurements**

Substrate sequence P4-P1'						Enzyme (origin)	Assay concentration			
MI	P4	P3	P2	P1	P1'		[E] <sup>#</sup> [pM]	[S <sub>max</sub> ] <sup>*</sup> [μM]	S1 <sup>#</sup> [μM]	S2 <sup>#</sup> [μM]
109	Tos	Gly	Pro	Arg	AMC	Thrombin (bovine)	61	20	10	5
505	H	dArg	Pro	Arg	AMC	Matriptase, wt (human)	202	100	-	-
						Matriptase, C122S (human)	189	100	-	-
506	H	dArg	Gly	Arg	AMC	Matriptase, wt (human)	202	100	-	-
						Matriptase, C122S (human)	189	100	-	-
507	Mes	dArg	Pro	Arg	AMC	Matriptase, wt (human)	202	50	14.29	7.15
						Matriptase, C122S (human)	189	100	14.29	7.15
						Factor Xa (human)	110	50	50	25
508	Mes	dArg	Gly	Arg	AMC	Matriptase, wt (human)	202	50	-	-
						Matriptase, C122S (human)	189	20	-	-
						Trypsin (porcine)	77	40	20	10
1314	Mes	dArg	Val	Arg	AMC	Matriptase, wt (human)	202	100	-	-
						Matriptase, C122S (human)	189	100	-	-
1319	H	dhPhe	Pro	Arg	AMC	Matriptase, wt (human)	202	100	-	-
						Matriptase, C122S (human)	189	100	-	-

<sup>\*</sup>highest substrate concentration for determination of  $K_m$  and  $V_{\text{max}}$

<sup>#</sup> final substrate concentration for  $K_i$  determination or enzyme concentration for kinetic measurements

*Determination of  $K_m$  und  $V_{max}$* 

In order to compare the kinetics of both matriptase variants and to detect the appropriate substrate for inhibition assays the standard values  $K_M$ ,  $V_{max}$ , and  $V_{max}/K_M$  were determined using the Michaelis-Menten equation (1), with  $V_{max}$  as the maximum reaction rate,  $[S]$  for substrate concentration and  $K_M$  as Michaelis-Menten constant.

Equation (2) 
$$v = \frac{V_{max} [S]}{K_m + [S]}$$

The  $K_M$  values corresponds to the substrate concentration, which provides the half-maximal rate ( $V_{max}/2$ ) under the used measurement conditions. Progress curves were measured with six different substrate concentrations prepared by a serial 1:1 dilution in ultrapure  $H_2O$  starting with the highest concentration ( $S_{max}$  in Table 7-1 ).

*Determination of the inhibitory constant  $K_i$* 

$K_i$  values of all substrates in this thesis were calculated by fitting the determined steady-state rates ( $v$ ) at 2 different substrate ( $S$ ) and 5 different inhibitor ( $I$ ) concentrations using Equation 3. In parallel, also Dixon plots were used for  $K_i$  value calculations and visualization of the data points <sup>[283]</sup>. The given  $K_i$  values represent the mean of standardly three independent measurements (independent dilutions) of at least 2 different weighted portions.

Equation (3) 
$$v = \frac{V_{max} * [S]}{K_m \left(1 + \frac{[I]}{K_i}\right) + [S]}$$

The highest inhibitor concentration was chosen to have an inhibitory effect of approximately 50 % ( $I_1=100$  %) and was subsequently diluted to  $I_2 = 75$  %,  $I_3 = 50$  % and  $I_4 = 25$  % with Enzyme kinetic buffer. The fifth value represented the steady-state rate in absence of the inhibitor ( $I_0 = 0$  %).

**7.6. Biostructural methods****7.6.1. Crystallization by vapor diffusion**

For in house crystallization, hanging drop vapor diffusion technique was performed at 18 °C. If not otherwise stated, the reservoirs of 24 well-plates were filled with 500 - 800  $\mu$ l respective crystallization buffer. Protein and buffer were mixed 1:1 on clean cover slides in a maximum volume of 4  $\mu$ l. Cover slides were arranged on top of the reservoirs with the drop facing the

reservoir solution. To ensure proper closure, wells were sealed with medium viscous silicone grease. Crystal formation and growth was monitored with a light microscope. For experimental hanging drop crystallization set up, matriptase was concentrated to 5-10 mg/ml and incubated with 10-20 mM benzamidine 2 HCl or 1-2 mM of inhibitor **17** (Table 4-1) for 30 min on ice before being mixed 1:1 with the respective crystallization buffer. Different crystallization buffers were prepared and used for screening of varying pH, precipitants, salts and additives and are not all described in detail.

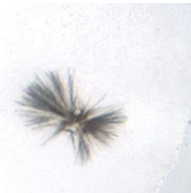

### **7.6.2. Microseeding experiments with C122S-matriptase inhibitor complex**

For microseeding experiments a seeding stock was prepared using MicroSeedBeads™ (Molecular Dimensions). A tangly crystal, grown in 0.1 M Tris HCl (pH 8.5), 35 % (w/v) PEG 5000 and 0.2 M MgCl<sub>2</sub> was pulverized by vortexing in 50 µl of original reservoir solution. After a serial ten times dilution in fresh buffer, seed stocks were flash frozen in liquid nitrogen and stored at -80°C unless used directly. Before use, seeding stock were vortexed thoroughly and centrifuged for 1 min at 10.000 rpm. C122S-matriptase was concentrated to 3-4 mg/ml and mixed with respective crystallization buffer and seeding stock on a cover slide in the ratio 3:2:1 (v/v/v) (e.g., 1.2 µl protein, 0.8 µl buffer and 0.4 µl seeding stock).

### **7.6.3. High-throughput (HT) nanodrop crystallization of C122S-matriptase**

To identify suitable crystallization conditions, freshly prepared C122S-matriptase was sent to the crystallization laboratory MarXtal (Marburg, Germany). Using a HT nanodrop crystallization robot, commercially available screens, including the Morpheus Screen (Molecular Dimensions; exact conditions unknown), NeXtal Tubes JCSG Core Suite III and Core Suite IV screens (QIAGEN, Hilden). With each screen 96 different conditions were tested in a sitting drop set up. Matriptase was concentrated to 7 mg/ml and incubated for 1 h with either 10 mM benzamine 2 HCl or 1 mM of inhibitor **17** (Table 4-1) on ice before being filtered in order to remove precipitants and other particles. To monitor crystal growth, pictures were taken periodically using the Formulatrix Rock Imager™ starting directly after the drop set up. Further pictures were taken once per day for the first 5 days and then after one week, two weeks and finally after three weeks and upon request (after 3 months and 1 year).

#### 7.6.4. Microseed Matrix Screening (MMS) with C122S-matriptase

Seeding set up	Drop I	Drop II
Seeding material		
Commercial screen well	JCSG Core Suite IV No. 74	JCSG Core Suite III well E4
Crystallization buffer	0.1 M MES (pH 6.0) 1.0 M C <sub>4</sub> H <sub>4</sub> KNaO <sub>6</sub>	0.1 M MES (pH 6.5) 1.6 M MgCl <sub>2</sub>

Based on a protocol from D'ARCY <sup>[193]</sup> crystalline C122S-matriptase material from a previous experiment was homogenized thorough vortexing with MicroSeedBeads™ (Molecular Dimensions) in 100 µl of the respective crystallization buffer. The seeding stock was diluted 1:10, vortexed, centrifuged and the supernatant was used in the crystal set up with the respective crystallization buffer and C122S-matriptase at 6.4 mg/ml in a ratio of 1:2:3 (100 nl seeding stock; 200 nl crystallization buffer; 300 nl C122S-matriptase). Pictures were taken periodically using the Formulatrix Rock Imager™ as described above. The seed stock and the 1:10 dilution were flash frozen in liquid nitrogen and stored at -80°C for further use.

#### 7.6.5. Soaking and freezing of C-122S matriptase crystal

Based on earlier protocols <sup>[159]</sup> soaking of the C122S-matriptase crystal was performed for 3 h at 18 °C in the crystallization buffer containing 1 mM of the respective compound. Crystals were then briefly dipped in the glycerol containing cryo buffer and subsequently frozen in liquid nitrogen.

C122S-matriptase soaking buffer		
Component	Stock solutions	Final concentration
Sodium acetate	1 M	0.1 M
Sodium formate	7 M	2 M
Compound <b>55</b>	100 mM in DMSO	1 mM

C122S-matriptase cryo buffer		
Component	Stock solutions	Final concentration
Sodium acetate	1 M	0.1 M
Sodium formate	7 M	2 M
Glycerol	100 %	16 %

### 7.6.6. Co-crystallization and freezing of bovine trypsin with compound **55** and **56**

Co-crystallization of bovine  $\beta$ -trypsin with compound **55** (Table 4-8) and **56** (Table 4-9) was performed as described previously <sup>[128]</sup>. For the preparation of the protein-inhibitor solution, trypsin (#T8003 Sigma-Aldrich, Germany) was dissolved in 10 mM  $\text{CaCl}_2$  and 1 mM HCl to a concentration of 20 - 40 mg/ml and incubated on ice for 30 min. Inhibitors were prepared as 10 mM solutions in 30 % DMSO and diluted 1:10 in a solution consisting of 50 % protein solution and 40 %  $\text{H}_2\text{O}$  (= Protein-inhibitor solution). In total, three different Protein-inhibitor solutions (a-b) were prepared with varying protein concentrations. Each Protein-inhibitor solution was tested in combination with two different crystallization buffers containing different PEG 8000 concentrations (20 or 25 %).

Protein-inhibitor solution			
	Stock solutions	Volume [ $\mu\text{l}$ ]	Final concentration
<b>Protein:</b>	a) 20 mg/ml b) 30 mg/ml c) 40 mg/ml	50 $\mu\text{l}$	a) 10 mg/ml b) 15 mg/ml c) 20 mg/ml
<b>Inhibitor:</b>	10 mM	10 $\mu\text{l}$	1 mM
<b>H<sub>2</sub>O:</b>	-	40 $\mu\text{l}$	-
		$\Sigma = 100 \mu\text{l}$	

Trypsin - Crystallization buffer			
Component	Stock solutions	Final buffer composition	
		Buffer 1 (B1)	Buffer 2 (B2)
Imidazole (pH 8.0)	1 M (pH: 8.0)	0.1 M	0.1 M
$(\text{NH}_4)_2\text{SO}_4$	3.8 M	0.1 M	0.1 M
PEG 8000	50 % (w/v)	20 % (v/v)	25 % (v/v)
$\text{NaN}_3$	5 % (w/v)	0.1 % (v/v)	0.1 % (v/v)

One 24-well plate per inhibitor was prepared by filling 500  $\mu\text{l}$  B1 or B2 in the respective well. 2  $\mu\text{l}$  of protein-inhibitor solution was mixed with 2  $\mu\text{l}$  buffer on cover slides. Each protein concentration was screened for the same buffer conditions 4 times. Crystals appeared in all wells overnight and reached full size after one week. Freezing of the crystals was performed by CrystalsFirst GmbH (Marburg) using a cryo buffer of unknown composition.

### 7.6.7. Data collection, processing and refinement

All data sets were collected at the synchrotron BESSY II (Berlin-Adlershof, HZB, Germany) <sup>[284]</sup>. The data sets were processed and scaled using XDS. Molecular replacement was done using Phaser from the CCP4 suite <sup>[285, 286]</sup> using PDB entry 2GV6 for matriptase <sup>[159]</sup> and 2ZFS for trypsin <sup>[287]</sup> as starting models. The refinement of the C122S-matriptase and trypsin crystals

was performed by CrystalsFirst GmbH (Marburg). The final models, giving the best possible explanation of the electron density, were deposited in the PDB with code 6T9T for the matriptase/compound **55**-complex, 6T9V for trypsin in complex with compound **55** and 6T9U for trypsin in complex with inhibitor **56**.

## 7.7. Synthesis of compounds

The synthesis of the 3-aminidinophenylalanine-derived inhibitors was done by OLIVER PILGRAM following previously described methods <sup>[159, 218]</sup>.

## 7.8. Computational methods

The computational part was carried out in collaboration with the the computational chemist DR. FLORENT CHEVILLARD.

### 7.8.1. Creation of core compound

The crystal structure of matriptase in complex with compound **20** (Figure 4-2; PDB code: 2GV6) was used as starting point for docking studies in scope of this thesis. In order to target specific single amino acids in the distal matriptase S3/4 binding pocket, the N-terminal naphthylsulfonyl of compound **20** was substituted with a 3-iodo-benzene to which an additional aryl moiety was attached by SUZUKI coupling (as done for similar compounds earlier including inhibitor **17** <sup>[217, 218, 224]</sup>). As this “virtual” compound is too big and flexible for docking studies solely 1-iodo-3-(methylsulfonyl)benzene (Figure 7-3) was used as a surrogate compound as the benzene ring and sulfone group were sufficient as core segment to have a shape reference for visual inspection of the docking poses.

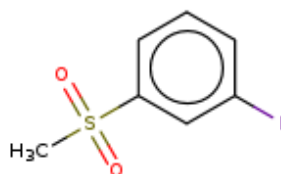


Figure 7-3: 1-iodo-3-(methylsulfonyl)benzene

### 7.8.2. Dataset creation

As 3-iodophenylsulfonyl cyanophenylalanine contains an aryl halide moiety suitable for modifications by Suzuki coupling, compatible extensions are any boronic acid. Three building block catalogs were downloaded from Chembridge (<https://www.chembridge.com>), Sigma Aldrich (<https://www.sigmaaldrich.com>) and Enamine ([www.enamine.net](http://www.enamine.net)) as SMILES files.

Boronic acids were extracted from those datasets using the "Filter your library" module from the PINGUI toolbox ([www.kolblab.org/scubidoo/pingui/](http://www.kolblab.org/scubidoo/pingui/)). After removing duplicates by using the isomeric canonical SMILES employing a python script based on RDKit the initial possible building block amount was reduced from 3054 to 2630. Starting from the 1-iodo-3-(methylsulfonyl)benzene as surrogate, virtual products were then generated using the "Create your virtual library" module from PINGUI by selecting "Suzuki" as reaction and uploading the aforementioned boronic acids datasets. In total, 2151 virtual products were downloaded as SMILES and prepared for docking.

### **7.8.3. Data preparation, docking and selection of molecules**

Docking calculations were performed based on the matriptase crystal structure in complex with compound **20** (Figure 4-2) Virtual products were protonated using QUACPAC ([openeye.com](http://openeye.com)) and 3D conformers were generated using OMEGA ([openeye.com](http://openeye.com)) with a maximum of 100 conformers for each virtual product. All virtual product conformers were placed within the binding site of matriptase using HYBRID ([openeye.com](http://openeye.com)). The top 500 virtual product docking poses (based on the docking score) were carefully visually inspected. For each pose, the main quality criterion was the overlap of the sulfone benzene group with the reference compound **20**.





## Summary in German

Matriptase und TMPRSS2 sind humane Transmembran Serinproteasen, die in der Lage sind, verschiedene Subtypen des Influenza-A Oberflächenproteins Hemagglutinin proteolytisch zu spalten und somit die Vermehrung von Influenzaviren in der Wirtszelle ermöglichen. Die Charakterisierung dieser Wirtsproteasen ist eine wichtige Voraussetzung zur Entwicklung wirksamer und selektiver Inhibitoren als potentielle Wirkstoffe zur Therapie von Influenzainfektionen. In der vorliegenden Arbeit wurde ein Expressionssystem zur Herstellung rekombinanter C122S mutierter Matriptase etabliert und das erhaltene Protein im Folgenden für die enzymkinetische Charakterisierung verschiedener Inhibitoren sowie zur Strukturanalyse eingesetzt. Des Weiteren wurden verschiedene verkürzte und/oder mutierte TMPRSS2 Formen hergestellt und auf ihre proteolytische Aktivität mit dem Ziel untersucht, eine isolierbare TMPRSS2 Variante für enzymkinetische Messungen bereit zu stellen.

### Herstellung rekombinanter Matriptase

In dem Wirtsorganismus *E. coli* wurden zwei verschiedene Matriptasevarianten produziert und miteinander enzymkinetisch verglichen. Die Serinproteasedomänen der Wildtyp-Matriptase und der mutierten C122S-Matriptase wurden jeweils mit 6 verschiedenen AMC-Substraten vermessen und zeigten dasselbe enzymkinetische Profil. Die etwas schlechteren Ausbeuten bei der Produktion der C122S-Matriptasevariante wurde durch die deutlich kostengünstigere Herstellung kompensiert. Im Folgenden wurden größere Mengen der C122S-Matriptase produziert und für Kristallisationsexperimente eingesetzt. Durch den Einsatz der Hochdurchsatz Kristallisationsrobotik konnte ein geeigneter Kristall zur Strukturanalyse gefunden werden. Mittels „*Micoseed Matrix Screening*“ konnten in einem zweiten Ansatz 4 weitere gut geformte Matriptase Proteinkristalle erhalten werden, indem zwei für die Röntgenstrukturanalyse qualitativ nicht geeignete und zerkleinerte Proteinkristalle dem neuen Kristallisationsansatz beigelegt wurden.

### Neue 3-Amidinophenylalanin-Derivate als selektive Matriptaseinhibitoren

Im Fokus der Entwicklung neuer monobasischer 3-Amidinophenylalanin Derivate im Rahmen dieser Arbeit standen potente Matriptase Verbindungen die eine hohe Selektivität gegenüber den Serinproteasen der Blutgerinnungskaskade Faktor Xa und Thrombin aufweisen. Insgesamt wurden 18 monobasische und somit potentiell bioverfügbare, als auch 4 dibasische Verbindungen getestet. Auf Basis einer Kristallstruktur der Matriptase in Komplex mit einem Inhibitor des 3-Amidinophenylalanin-Typs (PDB code: 2GV6, Verbindung **20**) wurde *Docking*

genutzt um mögliche Optimierungen des N-Terminus zu identifizieren. Es wurden Verbindungen mit einer 3-Hydroxymethylgruppe ausgesucht, da eine Interaktion mit dem Gln175 der Matriptase vorhergesagt wurde, die zu einer erhöhten Matriptase Affinität, aber nicht fXa Affinität, beitragen soll. Die monobasische Verbindung **55** mit einem 4-*tert*-Butylureido Piperidin am C- und 3-Fluor-4-Hydroxymethyl Biphenylsulphonyl am N-Terminus zeigte eine hervorragende Selektivität gegenüber Thrombin (1533-fach). Der kristallographische Bindemodus dieser Verbindung in Komplex mit Matriptase zeigte eine Interaktion der 4-Hydroxymethylgruppe mit dem Gln175, was zu einer moderaten Selektivität gegenüber fXa führte (12-fach). Der Bindemodus von Verbindung **55** (als auch Verbindung **56**) wurde zudem in Komplex mit Trypsin bestimmt. Das gut verfügbare Trypsin erwies sich als geeignete Alternative zu Matriptase, wenn eine Interaktion des Inhibitors mit der S3/4 Bindetasche im Fokus steht. Der Inhibitor C-Terminus zeigte dagegen eine abweichende Orientierung in beiden Proteasen. Die Trypsin/Verbindung **56**-Kristallstruktur zeigte außerdem, dass ein N-terminales 3-Hydroxymethyl theoretisch mit dem Gln175 in Matriptase interagieren würde. Verbindungen mit einem 3-Hydroxymethyl zeigten allerdings eine verminderte Matriptase Affinität (als auch Selektivität) in enzymkinetischen Messungen.

### **Charakterisierung von TMPRSS2**

Aufgrund erfolgloser Isolierungsversuche aktiver TMPRSS2 aus *E.coli* Bakterien wurden verschiedene Strategien entwickelt um die anspruchsvolle Transmembran Serinprotease aus Säugerzellen zu isolieren. Dazu wurden zunächst verkürzte als auch mutierte TMPRSS2-Varianten eingesetzt um strukturelle Eigenschaften zu identifizieren, die essentiell für eine Proteaseaktivität sind. Die Hauptstrategie basierte auf dem Entfernen von mindestens der Transmembrandomäne um lösliches und somit leichter isolierbares Enzym zu erhalten. Obwohl im Rahmen dieser Thesis keine aktive TMPRSS2 isoliert wurde, konnten durch diese Versuche neue Erkenntnisse zur TMPRSS2 Funktionalität gewonnen werden. Es wurde gezeigt, dass der zytosolische Teil, die Transmembrandomäne als auch die LDLRA Domäne keine Voraussetzung zur autokatalytischen Aktivierung von TMPRSS2 sind. Die SRCR Domäne als auch die Disulfidbrücke, welche die Proteasendomäne mit der Stammregion nach der Aktivierungsspaltung verbindet, sind hingegen essentiell für TMPRSS2 Aktivität in der Zelle. Trotz Entfernung der in der Zellmembran verankernden Transmembrandomäne konnte in Immunofluoreszenz-Versuchen keine offensichtliche Dislokalisierung der verkürzten Varianten im Vergleich zum Wildtyp festgestellt werden. Des Weiteren wurde TMPRSS2 *in vitro* als ein Substrat der Matriptase identifiziert. Dies deutet auf die Involvierung von TMPRSS2 in mehreren Zymogenkaskaden hin.

## Bibliography

- [1] WHO. Global influenza strategy 2019-2030. **2019**
- [2] Clark, N.; Lynch, J. Influenza: Epidemiology, Clinical Features, Therapy, and Prevention. *Seminars in respiratory and critical care medicine* **2011**, 32, 373-92.
- [3] Targonski, P. V.; Poland, G. A. Influenza. In *International Encyclopedia of Public Health (Second Edition)*, Quah, S. R., Ed. Academic Press: Oxford, 2017; pp 238-246.
- [4] Johnson, N. P.; Mueller, J. Updating the accounts: global mortality of the 1918-1920 "Spanish" influenza pandemic. *Bulletin of the history of medicine* **2002**, 76, 105-15.
- [5] Simonsen, L.; Spreeuwenberg, P.; Lustig, R.; Taylor, R. J.; Fleming, D. M.; Kroneman, M.; Van Kerkhove, M. D.; Mounts, A. W.; Paget, W. J. Global mortality estimates for the 2009 Influenza Pandemic from the GLaMOR project: a modeling study. *PLoS medicine* **2013**, 10, e1001558.
- [6] Buda, S.; Prahm, K.; Dürrwald, R.; Biere, B.; Schilling, J.; Buchholz, U.; an der Heiden, M.; Haas, W. Bericht zur Epidemiologie der Influenza in Deutschland Saison 2017/18. *Robert Koch-Institut* **2018**.
- [7] WHO. Up to 650 000 people die of respiratory diseases linked to seasonal flu each year. **2017**
- [8] Tong, S.; Li, Y.; Rivaller, P.; Conrardy, C.; Castillo, D. A.; Chen, L. M.; Recuenco, S.; Ellison, J. A.; Davis, C. T.; York, I. A.; Turmelle, A. S.; Moran, D.; Rogers, S.; Shi, M.; Tao, Y.; Weil, M. R.; Tang, K.; Rowe, L. A.; Sammons, S.; Xu, X.; Frace, M.; Lindblade, K. A.; Cox, N. J.; Anderson, L. J.; Rupprecht, C. E.; Donis, R. O. A distinct lineage of influenza A virus from bats. *Proceedings of the National Academy of Sciences of the United States of America* **2012**, 109, 4269-74.
- [9] Fouchier, R. A.; Munster, V.; Wallensten, A.; Bestebroer, T. M.; Herfst, S.; Smith, D.; Rimmelzwaan, G. F.; Olsen, B.; Osterhaus, A. D. Characterization of a novel influenza A virus hemagglutinin subtype (H16) obtained from black-headed gulls. *Journal of virology* **2005**, 79, 2814-22.
- [10] Tong, S.; Zhu, X.; Li, Y.; Shi, M.; Zhang, J.; Bourgeois, M.; Yang, H.; Chen, X.; Recuenco, S.; Gomez, J.; Chen, L.-M.; Johnson, A.; Tao, Y.; Dreyfus, C.; Yu, W.; McBride, R.; Carney, P. J.; Gilbert, A. T.; Chang, J.; Guo, Z.; Davis, C. T.; Paulson, J. C.; Stevens, J.; Rupprecht, C. E.; Holmes, E. C.; Wilson, I. A.; Donis, R. O. New world bats harbor diverse influenza A viruses. *PLoS Pathog* **2013**, 9, e1003657-e1003657.
- [11] Noda, T. Native Morphology of Influenza Virions. *Frontiers in Microbiology* **2011**, 2, 269.
- [12] Horimoto, T.; Kawaoka, Y. Influenza: lessons from past pandemics, warnings from current incidents. *Nature reviews. Microbiology* **2005**, 3, 591-600.
- [13] Steinmetzer, T.; Hards, K.; Böttcher-Friebertshäuser, E.; Garten, W. Strategies for the Development of Influenza Drugs: Basis for New Efficient Combination Therapies. *Topics in Medicinal Chemistry* **2015**, 15, 143-182.
- [14] Shi, Y.; Wu, Y.; Zhang, W.; Qi, J.; Gao, G. F. Enabling the 'host jump': structural determinants of receptor-binding specificity in influenza A viruses. *Nature Reviews Microbiology* **2014**, 12, 822-831.
- [15] Luo, M. Influenza virus entry. *Advances in experimental medicine and biology* **2012**, 726, 201-21.

- [16] Matrosovich, M.; Stech, J.; Klenk, H. D. Influenza receptors, polymerase and host range. *Revue scientifique et technique (International Office of Epizootics)* **2009**, 28, 203-17.
- [17] Sriwilaijaroen, N.; Suzuki, Y. Molecular basis of the structure and function of H1 hemagglutinin of influenza virus. *Proc Jpn Acad Ser B Phys Biol Sci* **2012**, 88, 226-249.
- [18] Bui, M.; Whittaker, G.; Helenius, A. Effect of M1 protein and low pH on nuclear transport of influenza virus ribonucleoproteins. *Journal of virology* **1996**, 70, 8391-8401.
- [19] Wang, P.; Palese, P.; O'Neill, R. E. The NPI-1/NPI-3 (karyopherin alpha) binding site on the influenza A virus nucleoprotein NP is a nonconventional nuclear localization signal. *Journal of virology* **1997**, 71, 1850-6.
- [20] Weber, F.; Kochs, G.; Gruber, S.; Haller, O. A classical bipartite nuclear localization signal on Thogoto and influenza A virus nucleoproteins. *Virology* **1998**, 250, 9-18.
- [21] Garten, W.; Klenk, H. D. Cleavage Activation of the Influenza Virus Hemagglutinin and Its Role in Pathogenesis. *Avian Influenza* **2008**, 27, 156-167.
- [22] Kuiken, T. Is low pathogenic avian influenza virus virulent for wild waterbirds? *Proceedings of the Royal Society B: Biological Sciences* **2013**, 280, 20130990.
- [23] Sonnberg, S.; Webby, R. J.; Webster, R. G. Natural history of highly pathogenic avian influenza H5N1. *Virus Research* **2013**, 178, 63-77.
- [24] Spackman, E. A Brief Introduction to Avian Influenza Virus. In *Animal Influenza Virus*, Springer New York: 2014; pp 61-68.
- [25] Puzelli, S.; Rossini, G.; Facchini, M.; Vaccari, G.; Di Trani, L.; Di Martino, A.; Gaibani, P.; Vocale, C.; Cattoli, G.; Bennett, M.; McCauley, J. W.; Rezza, G.; Moro, M. L.; Rangoni, R.; Finarelli, A. C.; Landini, M. P.; Castrucci, M. R.; Donatelli, I. Human infection with highly pathogenic A(H7N7) avian influenza virus. *Emerg Infect Dis* **2014**, 20, 1745-1749.
- [26] Shi, J.; Xie, J.; He, Z.; Hu, Y.; He, Y.; Huang, Q.; Leng, B.; He, W.; Sheng, Y.; Li, F.; Song, Y.; Bai, C.; Gu, Y.; Jie, Z. A detailed epidemiological and clinical description of 6 human cases of avian-origin influenza A (H7N9) virus infection in Shanghai. *PLoS One* **2013**, 8, e77651-e77651.
- [27] Webster, R. G.; Bean, W. J.; Gorman, O. T.; Chambers, T. M.; Kawaoka, Y. Evolution and ecology of influenza A viruses. *Microbiological Reviews* **1992**, 56, 152.
- [28] Potter, C. W. A history of influenza. *Journal of Applied Microbiology* **2001**, 91, 572-579.
- [29] Cox, N. J.; Subbarao, K. Global Epidemiology of Influenza: Past and Present. *Annual Review of Medicine* **2000**, 51, 407-421.
- [30] Al Hajjar, S.; McIntosh, K. The first influenza pandemic of the 21st century. *Ann Saudi Med* **2010**, 30, 1-10.
- [31] Hannoun, C. The evolving history of influenza viruses and influenza vaccines. *Expert review of vaccines* **2013**, 12, 1085-94.
- [32] Shope, R. E. Immunization Experiments with Swine Influenza Virus. *J Exp Med* **1936**, 64, 47-61.

- [33] Soema, P. C.; Kompier, R.; Amorij, J.-P.; Kersten, G. F. A. Current and next generation influenza vaccines: Formulation and production strategies. *European Journal of Pharmaceutics and Biopharmaceutics* **2015**, 94, 251-263.
- [34] Kidd, I. Influenza viruses: Update on epidemiology, clinical features, treatment and vaccination. *Current opinion in pulmonary medicine* **2014**, 20, 242-246.
- [35] Baranovich, T.; Bahl, J.; Marathe, B. M.; Culhane, M.; Stigger-Rosser, E.; Darnell, D.; Kaplan, B. S.; Lowe, J. F.; Webby, R. J.; Govorkova, E. A. Influenza A viruses of swine circulating in the United States during 2009-2014 are susceptible to neuraminidase inhibitors but show lineage-dependent resistance to adamantanes. *Antiviral Research* **2015**, 117, 10-19.
- [36] Zhang, Y.; Liu, Q.; Wang, D.; Chen, S.; Wang, S. Simultaneous detection of oseltamivir- and amantadine-resistant influenza by oligonucleotide microarray visualization. *PLoS One* **2013**, 8, e57154.
- [37] Baek, Y. H.; Song, M. S.; Lee, E. Y.; Kim, Y. I.; Kim, E. H.; Park, S. J.; Park, K. J.; Kwon, H. I.; Pascua, P. N.; Lim, G. J.; Kim, S.; Yoon, S. W.; Kim, M. H.; Webby, R. J.; Choi, Y. K. Profiling and characterization of influenza virus N1 strains potentially resistant to multiple neuraminidase inhibitors. *Journal of virology* **2015**, 89, 287-99.
- [38] Moscona, A. Global Transmission of Oseltamivir-Resistant Influenza. *New England Journal of Medicine* **2009**, 360, 953-956.
- [39] Sheu, T. G.; Deyde, V. M.; Okomo-Adhiambo, M.; Garten, R. J.; Xu, X.; Bright, R. A.; Butler, E. N.; Wallis, T. R.; Klimov, A. I.; Gubareva, L. V. Surveillance for neuraminidase inhibitor resistance among human influenza A and B viruses circulating worldwide from 2004 to 2008. *Antimicrobial agents and chemotherapy* **2008**, 52, 3284-92.
- [40] Corti, D.; Voss, J.; Gamblin, S. J.; Codoni, G.; Macagno, A.; Jarrossay, D.; Vachieri, S. G.; Pinna, D.; Minola, A.; Vanzetta, F.; Silacci, C.; Fernandez-Rodriguez, B. M.; Agatic, G.; Bianchi, S.; Giacchetto-Sasselli, I.; Calder, L.; Sallusto, F.; Collins, P.; Haire, L. F.; Temperton, N.; Langedijk, J. P.; Skehel, J. J.; Lanzavecchia, A. A neutralizing antibody selected from plasma cells that binds to group 1 and group 2 influenza A hemagglutinins. *Science (New York, N.Y.)* **2011**, 333, 850-6.
- [41] Li, F.; Ma, C.; Wang, J. Inhibitors Targeting the Influenza Virus Hemagglutinin. *Current medicinal chemistry* **2015**, 22, 1361-1382.
- [42] Garten, W.; Bosch, F. X.; Linder, D.; Rott, R.; Klenk, H. D. Proteolytic activation of the influenza virus hemagglutinin: The structure of the cleavage site and the enzymes involved in cleavage. *Virology* **1981**, 115, 361-74.
- [43] Igarashi, M.; Ito, K.; Yoshida, R.; Tomabeche, D.; Kida, H.; Takada, A. Predicting the Antigenic Structure of the Pandemic (H1N1) 2009 Influenza Virus Hemagglutinin. *PLoS One* **2010**, 5, e8553.
- [44] Copeland, C. S.; Doms, R. W.; Bolzau, E. M.; Webster, R. G.; Helenius, A. Assembly of influenza hemagglutinin trimers and its role in intracellular transport. *The Journal of cell biology* **1986**, 103, 1179-91.
- [45] Isin, B.; Doruker, P.; Bahar, I. Functional Motions of Influenza Virus Hemagglutinin: A Structure-Based Analytical Approach. *Biophysical journal* **2002**, 82, 569-581.
- [46] Yamaguchi, M.; Danev, R.; Nishiyama, K.; Sugawara, K.; Nagayama, K. Zernike phase contrast electron microscopy of ice-embedded influenza A virus. *Journal of structural biology* **2008**, 162, 271-6.

- [47] Chen, J.; Lee, K. H.; Steinhauer, D. A.; Stevens, D. J.; Skehel, J. J.; Wiley, D. C. Structure of the hemagglutinin precursor cleavage site, a determinant of influenza pathogenicity and the origin of the labile conformation. *Cell* **1998**, 95, 409-17.
- [48] Wiley, D. C.; Skehel, J. J. The structure and function of the hemagglutinin membrane glycoprotein of influenza virus. *Annual review of biochemistry* **1987**, 56, 365-94.
- [49] Steinhauer, D. A. Role of Hemagglutinin Cleavage for the Pathogenicity of Influenza Virus. *Virology* **1999**, 258, 1-20.
- [50] Skehel, J. J.; Waterfield, M. D. Studies on the primary structure of the influenza virus hemagglutinin. *Proceedings of the National Academy of Sciences of the United States of America* **1975**, 72, 93-7.
- [51] Huang, R. T.; Wahn, K.; Klenk, H. D.; Rott, R. Fusion between cell membrane and liposomes containing the glycoproteins of influenza virus. *Virology* **1980**, 104, 294-302.
- [52] Maeda, T.; Ohnishi, S.-i. Activation of influenza virus by acidic media causes hemolysis and fusion of erythrocytes. *FEBS Letters* **1980**, 122, 283-287.
- [53] Cross, K.; Langley, W.; J Russell, R.; J Skehel, J.; Steinhauer, D. *Composition and Functions of the Influenza Fusion Peptide*. **2009**; Vol. 16, p 766-78.
- [54] Galloway, S.; Liang, B.; Steinhauer, D. Activation of the Hemagglutinin of Influenza Viruses. In *Activation of Viruses by Host Proteases*, Böttcher-Friebertshäuser E., G. W., Klenk H, Ed. Springer, Cham: **2018**; pp 3-26.
- [55] Garten, W.; Klenk, H. D. Understanding influenza virus pathogenicity. *Trends in microbiology* **1999**, 7, 99-100.
- [56] Kido, H.; Yokogoshi, Y.; Sakai, K.; Tashiro, M.; Kishino, Y.; Fukutomi, A.; Katunuma, N. Isolation and characterization of a novel trypsin-like protease found in rat bronchiolar epithelial Clara cells. A possible activator of the viral fusion glycoprotein. *The Journal of biological chemistry* **1992**, 267, 13573-9.
- [57] Lazarowitz, S. G.; Goldberg, A. R.; Choppin, P. W. Proteolytic cleavage by plasmin of the HA polypeptide of influenza virus: Host cell activation of serum plasminogen. *Virology* **1973**, 56, 172-180.
- [58] Böttcher, E.; Matrosovich, T.; Beyerle, M.; Klenk, H.-D.; Garten, W.; Matrosovich, M. Proteolytic activation of influenza viruses by serine proteases TMPRSS2 and HAT from human airway epithelium. *Journal of virology* **2006**, 80, 9896-9898.
- [59] Chaipan, C.; Kobasa, D.; Bertram, S.; Glowacka, I.; Steffen, I.; Tsegaye, T. S.; Takeda, M.; Bugge, T. H.; Kim, S.; Park, Y.; Marzi, A.; Pöhlmann, S. Proteolytic activation of the 1918 influenza virus hemagglutinin. *Journal of virology* **2009**, 83, 3200-11.
- [60] Hatesuer, B.; Bertram, S.; Mehnert, N.; Bahgat, M. M.; Nelson, P. S.; Pöhlmann, S.; Schughart, K. Tmprss2 is essential for influenza H1N1 virus pathogenesis in mice. *PLoS Pathog* **2013**, 9, e1003774-e1003774.
- [61] Sakai, K.; Ami, Y.; Tahara, M.; Kubota, T.; Anraku, M.; Abe, M.; Nakajima, N.; Sekizuka, T.; Shirato, K.; Suzuki, Y.; Ainai, A.; Nakatsu, Y.; Kanou, K.; Nakamura, K.; Suzuki, T.; Komase, K.; Nobusawa, E.; Maenaka, K.; Kuroda, M.; Hasegawa, H.; Kawaoka, Y.; Tashiro, M.; Takeda, M. The host protease TMPRSS2 plays a major role in in vivo replication of emerging H7N9 and seasonal influenza viruses. *Journal of virology* **2014**, 88, 5608-5616.

- [62] Tarnow, C.; Engels, G.; Arendt, A.; Schwalm, F.; Sediri, H.; Preuss, A.; Nelson, P. S.; Garten, W.; Klenk, H. D.; Gabriel, G.; Böttcher-Friebertshauser, E. TMPRSS2 is a host factor that is essential for pneumotropism and pathogenicity of H7N9 influenza A virus in mice. *Journal of virology* **2014**, 88, 4744-51.
- [63] Baron, J.; Tarnow, C.; Mayoli-Nüssle, D.; Schilling, E.; Meyer, D.; Hammami, M.; Schwalm, F.; Steinmetzer, T.; Guan, Y.; Garten, W.; Klenk, H.-D.; Böttcher-Friebertshäuser, E. Matriptase, HAT, and TMPRSS2 activate the hemagglutinin of H9N2 influenza A viruses. *Journal of virology* **2013**, 87, 1811-1820.
- [64] Hamilton, B. S.; Gludish, D. W. J.; Whittaker, G. R. Cleavage activation of the human-adapted influenza virus subtypes by matriptase reveals both subtype and strain specificities. *Journal of virology* **2012**, 86, 10579-10586.
- [65] Tsukada, H.; Blow, D. M. Structure of alpha-chymotrypsin refined at 1.68 Å resolution. *Journal of molecular biology* **1985**, 184, 703-11.
- [66] Huber, R.; Bode, W. Structural basis of the activation and action of trypsin. *Accounts of Chemical Research* **1978**, 11, 114-122.
- [67] Garten, W.; Braden, C.; Arendt, A.; Peitsch, C.; Baron, J.; Lu, Y.; Pawletko, K.; Hards, K.; Steinmetzer, T.; Böttcher-Friebertshäuser, E. Influenza virus activating host proteases: Identification, localization and inhibitors as potential therapeutics. *European Journal of Cell Biology* **2015**, 94, 375-383.
- [68] Antalis, T. M.; Bugge, T. H.; Wu, Q. Membrane-anchored serine proteases in health and disease. *Progress in molecular biology and translational science* **2011**, 99, 1-50.
- [69] Uhlén, M.; Fagerberg, L.; Hallström, B. M.; Lindskog, C.; Oksvold, P.; Mardinoglu, A.; Sivertsson, Å.; Kampf, C.; Sjöstedt, E.; Asplund, A.; Olsson, I.; Edlund, K.; Lundberg, E.; Navani, S.; Szigartyo, C. A.-K.; Odeberg, J.; Djureinovic, D.; Takanen, J. O.; Hober, S.; Alm, T.; Edqvist, P.-H.; Berling, H.; Tegel, H.; Mulder, J.; Rockberg, J.; Nilsson, P.; Schwenk, J. M.; Hamsten, M.; von Feilitzen, K.; Forsberg, M.; Persson, L.; Johansson, F.; Zwahlen, M.; von Heijne, G.; Nielsen, J.; Pontén, F. Tissue-based map of the human proteome. *Science (New York, N.Y.)* **2015**, 347, 1260419.
- [70] Cal, S.; Quesada, V.; Garabaya, C.; Lopez-Otin, C. Polyserase-I, a human polyprotease with the ability to generate independent serine protease domains from a single translation product. *Proceedings of the National Academy of Sciences of the United States of America* **2003**, 100, 9185-90.
- [71] List, K.; Hobson, J. P.; Molinolo, A.; Bugge, T. H. Co-localization of the channel activating protease prostasin/(CAP1/PRSS8) with its candidate activator, matriptase. *Journal of Cellular Physiology* **2007**, 213, 237-245.
- [72] Oberst, M. D.; Singh, B.; Ozdemirli, M.; Dickson, R. B.; Johnson, M. D.; Lin, C.-Y. Characterization of Matriptase Expression in Normal Human Tissues. *Journal of Histochemistry & Cytochemistry* **2003**, 51, 1017-1025.
- [73] Chen, S.; Sen, S.; Young, D.; Wang, W.; Moravec, C. S.; Wu, Q. Protease corin expression and activity in failing hearts. *American journal of physiology. Heart and circulatory physiology* **2010**, 299, H1687-92.
- [74] Tsuzuki, S.; Murai, N.; Miyake, Y.; Inouye, K.; Hirayasu, H.; Iwanaga, T.; Fushiki, T. Evidence for the occurrence of membrane-type serine protease 1/matriptase on the basolateral sides of enterocytes. *The Biochemical journal* **2005**, 388, 679-687.
- [75] Wang, X.-T.; Engel, P. C. An optimised system for refolding of human glucose 6-phosphate dehydrogenase. *BMC Biotechnology* **2009**, 9, 19.

- [76] Afar, D. E. H.; Vivanco, I.; S Hubert, R.; Kuo, J.; Chen, E.; Saffran, D.; Raitano, A.; Jakobovits, A. Catalytic Cleavage of the Androgen-regulated TMPRSS2 Protease Results in Its Secretion by Prostate and Prostate Cancer Epithelia. *Cancer research* **2001**, 61, 1686-1692.
- [77] Jacquinet, E.; Rao, N. V.; Rao, G. V.; Zhengming, W.; Albertine, K. H.; Hoidal, J. R. Cloning and characterization of the cDNA and gene for human epitheliasin. *European Journal of Biochemistry* **2001**, 268, 2687-2699.
- [78] Hooper, J. D.; Clements, J. A.; Quigley, J. P.; Antalis, T. M. Type II transmembrane serine proteases. Insights into an emerging class of cell surface proteolytic enzymes. *The Journal of biological chemistry* **2001**, 276, 857-60.
- [79] Lu, D.; Yuan, X.; Zheng, X.; Sadler, J. E. Bovine proenteropeptidase is activated by trypsin, and the specificity of enteropeptidase depends on the heavy chain. *The Journal of biological chemistry* **1997**, 272, 31293-300.
- [80] Lin, C. Y.; Wang, J. K.; Torri, J.; Dou, L.; Sang, Q. A.; Dickson, R. B. Characterization of a novel, membrane-bound, 80-kDa matrix-degrading protease from human breast cancer cells. Monoclonal antibody production, isolation, and localization. *The Journal of biological chemistry* **1997**, 272, 9147-52.
- [81] Takeuchi, T.; Harris, J. L.; Huang, W.; Yan, K. W.; Coughlin, S. R.; Craik, C. S. Cellular localization of membrane-type serine protease 1 and identification of protease-activated receptor-2 and single-chain urokinase-type plasminogen activator as substrates. *The Journal of biological chemistry* **2000**, 275, 26333-42.
- [82] Yamaoka, K.; Masuda, K.; Ogawa, H.; Takagi, K.; Umemoto, N.; Yasuoka, S. Cloning and characterization of the cDNA for human airway trypsin-like protease. *The Journal of biological chemistry* **1998**, 273, 11895-901.
- [83] Lee, M.-S.; Tseng, I.-C.; Wang, Y.; Kiyomiya, K.-i.; Johnson, M. D.; Dickson, R. B.; Lin, C.-Y. Autoactivation of matriptase in vitro: requirement for biomembrane and LDL receptor domain. *American Journal of Physiology-Cell Physiology* **2007**, 293, C95-C105.
- [84] Knappe, S.; Wu, F.; Rose Madlansacay, M.; Wu, Q. Identification of Domain Structures in the Propeptide of Corin Essential for the Processing of Proatrial Natriuretic Peptide. *The Journal of biological chemistry* **2004**, 279, 34464-34471.
- [85] Gladysheva, I.; M King, S.; Houn, A. N-Glycosylation modulates the cell-surface expression and catalytic activity of corin. *Biochemical and biophysical research communications* **2008**, 373, 130-5.
- [86] Liao, X.; Wang, W.; Chen, S.; Wu, Q. Role of Glycosylation in Corin Zymogen Activation. *Journal of Biological Chemistry* **2007**, 282, 27728-27735.
- [87] Somoza, J. R.; Ho, J. D.; Luong, C.; Ghate, M.; Sprengeler, P. A.; Mortara, K.; Shrader, W. D.; Sperandio, D.; Chan, H.; McGrath, M. E.; Katz, B. A. The structure of the extracellular region of human hepsin reveals a serine protease domain and a novel scavenger receptor cysteine-rich (SRCR) domain. *Structure (London, England : 1993)* **2003**, 11, 1123-31.
- [88] Kyrieleis, O. J. P.; Huber, R.; Ong, E.; Oehler, R.; Hunter, M.; Madison, E. L.; Jacob, U. Crystal structure of the catalytic domain of DESC1, a new member of the type II transmembrane serine proteinase family. *The FEBS Journal* **2007**, 274, 2148-2160.



- [89] Antalis, T. M.; Buzza, M. S.; Hodge, K. M.; Hooper, J. D.; Netzel-Arnett, S. The cutting edge: membrane-anchored serine protease activities in the pericellular microenvironment. *The Biochemical journal* **2010**, 428, 325-346.
- [90] Bugge, T. H.; Antalis, T. M.; Qingyu, W. Type II Transmembrane Serine Proteases. *Journal of Biological Chemistry* **2009**, 284, 23177-81.
- [91] Netzel-Arnett, S.; Hooper, J.; Szabo, R.; Madison, E.; Quigley, J. P.; Bugge, T. H.; Antalis, T. M. Membrane anchored serine proteases: a rapidly expanding group of cell surface proteolytic enzymes with potential roles in cancer. *Cancer and Metastasis Reviews* **2003**, 22, 237-58.
- [92] Yamashina, I. The action of enterokinase on trypsinogen. *Biochimica et biophysica acta* **1956**, 20, 433-4.
- [93] Holzinger, A.; Maier, E. M.; Buck, C.; Mayerhofer, P. U.; Kappler, M.; Haworth, J. C.; Moroz, S. P.; Hadorn, H. B.; Sadler, J. E.; Roscher, A. A. Mutations in the proenteropeptidase gene are the molecular cause of congenital enteropeptidase deficiency. *American journal of human genetics* **2002**, 70, 20-5.
- [94] Jin, X.; Hirosaki, T.; Lin, C.-Y.; Dickson, R. B.; Higashi, S.; Kitamura, H.; Miyazaki, K. Production of soluble matriptase by human cancer cell lines and cell surface activation of its zymogen by trypsin. *Journal of Cellular Biochemistry* **2005**, 95, 632-647.
- [95] Olsen, J. V.; Ong, S. E.; Mann, M. Trypsin cleaves exclusively C-terminal to arginine and lysine residues. *Molecular & cellular proteomics : MCP* **2004**, 3, 608-14.
- [96] Kazama, Y.; Hamamoto, T.; Foster, D. C.; Kisiel, W. Hepsin, a putative membrane-associated serine protease, activates human factor VII and initiates a pathway of blood coagulation on the cell surface leading to thrombin formation. *The Journal of biological chemistry* **1995**, 270, 66-72.
- [97] Jin, X.; Yagi, M.; Akiyama, N.; Hirosaki, T.; Higashi, S.; Lin, C.-Y.; Dickson, R. B.; Kitamura, H.; Miyazaki, K. Matriptase activates stromelysin (MMP-3) and promotes tumor growth and angiogenesis. *Cancer Science* **2006**, 97, 1327-1334.
- [98] List, K.; Bugge, T. H.; Szabo, R. Matriptase: potent proteolysis on the cell surface. *Molecular medicine* **2006**, 12, 1-7.
- [99] Macao, B.; Johansson, D. G. A.; Hansson, G. C.; Härd, T. Autoproteolysis coupled to protein folding in the SEA domain of the membrane-bound MUC1 mucin. *Nature Structural & Molecular Biology* **2006**, 13, 71-76.
- [100] Benaud, C.; Dickson, R. B.; Lin, C. Y. Regulation of the activity of matriptase on epithelial cell surfaces by a blood-derived factor. *Eur J Biochem* **2001**, 268, 1439-47.
- [101] Lai, C. H.; Lai, Y. J.; Chou, F. P.; Chang, H. H.; Tseng, C. C.; Johnson, M. D.; Wang, J. K.; Lin, C. Y. Matriptase Complexes and Prostatic Complexes with HAI-1 and HAI-2 in Human Milk: Significant Proteolysis in Lactation. *PLoS One* **2016**, 11, e0152904.
- [102] Tseng, I. C.; Chou, F.-P.; Su, S.-F.; Oberst, M.; Madayiputhiya, N.; Lee, M.-S.; Wang, J.-K.; Sloane, D. E.; Johnson, M.; Lin, C.-Y. Purification from human milk of matriptase complexes with secreted serpins: mechanism for inhibition of matriptase other than HAI-1. *Am J Physiol Cell Physiol* **2008**, 295, C423-C431.
- [103] List, K.; Szabo, R.; Molinolo, A.; Nielsen, B. S.; Bugge, T. H. Delineation of matriptase protein expression by enzymatic gene trapping suggests diverging roles in barrier function, hair formation, and squamous cell carcinogenesis. *Am J Pathol* **2006**, 168, 1513-1525.

- [104] Tseng, I. C.; Xu, H.; Chou, F.-P.; Li, G.; Vazzano, A. P.; Kao, J. P. Y.; Johnson, M. D.; Lin, C.-Y. Matriptase activation, an early cellular response to acidosis. *The Journal of biological chemistry* **2010**, 285, 3261-3270.
- [105] Madison, E.; Kobe, A.; Gething, M.-J.; Sambrook, J.; Goldsmith, E. Converting Tissue Plasminogen Activator to a Zymogen: A Regulatory Triad of Asp-His-Ser. *Science (New York, N.Y.)* **1993**, 262, 419-21.
- [106] Friis, S.; Tadeo, D.; Le-Gall, S. M.; Jürgensen, H. J.; Sales, K. U.; Camerer, E.; Bugge, T. H. Matriptase zymogen supports epithelial development, homeostasis and regeneration. *BMC Biology* **2017**, 15, 46.
- [107] Godiksen, S.; Soendergaard, C.; Friis, S.; Jensen, J. K.; Bornholdt, J.; Sales, K. U.; Huang, M.; Bugge, T. H.; Vogel, L. K. Detection of active matriptase using a biotinylated chloromethyl ketone peptide. *PLoS One* **2013**, 8, e77146-e77146.
- [108] Inouye, K.; Tomoishi, M.; Yasumoto, M.; Miyake, Y.; Kojima, K.; Tsuzuki, S.; Fushiki, T. Roles of CUB and LDL receptor class A domain repeats of a transmembrane serine protease matriptase in its zymogen activation. *Journal of biochemistry* **2013**, 153, 51-61.
- [109] Inouye, K.; Yasumoto, M.; Tsuzuki, S.; Mochida, S.; Fushiki, T. The optimal activity of a pseudozymogen form of recombinant matriptase under the mildly acidic pH and low ionic strength conditions. *Journal of biochemistry* **2010**, 147, 485-92.
- [110] Tamberg, T.; Hong, Z.; De Schepper, D.; Skovbjerg, S.; Dupont, D. M.; Vitved, L.; Schar, C. R.; Skjoedt, K.; Vogel, L. K.; Jensen, J. K. Blocking the proteolytic activity of zymogen matriptase with antibody-based inhibitors. *The Journal of biological chemistry* **2019**, 294, 314-326.
- [111] Shi, Y. E.; Torri, J.; Yieh, L.; Wellstein, A.; Lippman, M. E.; Dickson, R. B. Identification and characterization of a novel matrix-degrading protease from hormone-dependent human breast cancer cells. *Cancer research* **1993**, 53, 1409-15.
- [112] Bhatt, A. S.; Welm, A.; Farady, C. J.; Vasquez, M.; Wilson, K.; Craik, C. S. Coordinate expression and functional profiling identify an extracellular proteolytic signaling pathway. *Proceedings of the National Academy of Sciences of the United States of America* **2007**, 104, 5771-6.
- [113] Chou, F.-P.; Chen, Y.-W.; Zhao, X.; Xu-Monette, Z.; Young, K.; Gartenhaus, R.; Wang, J.-K.; Kataoka, H.; H Zuo, A.; J Barndt, R.; Johnson, M.; Lin, C.-Y. Imbalanced Matriptase Pericellular Proteolysis Contributes to the Pathogenesis of Malignant B-Cell Lymphomas. *Am J Pathol* **2013**, 183, 1306-17.
- [114] Lanchec, E.; Desilets, A.; Beliveau, F.; Flamier, A.; Mahmoud, S.; Bernier, G.; Gris, D.; Leduc, R.; Lavoie, C. The type II transmembrane serine protease matriptase cleaves the amyloid precursor protein and reduces its processing to beta-amyloid peptide. *The Journal of biological chemistry* **2017**, 292, 20669-20682.
- [115] List, K.; Haudenschild, C. C.; Szabo, R.; Chen, W.; Wahl, S. M.; Swaim, W.; Engelholm, L. H.; Behrendt, N.; Bugge, T. H. Matriptase/MT-SP1 is required for postnatal survival, epidermal barrier function, hair follicle development, and thymic homeostasis. *Oncogene* **2002**, 21, 3765-79.
- [116] Kamata, Y.; Taniguchi, A.; Yamamoto, M.; Nomura, J.; Ishihara, K.; Takahara, H.; Hibino, T.; Takeda, A. Neutral cysteine protease bleomycin hydrolase is essential for the breakdown of deiminated filaggrin into amino acids. *The Journal of biological chemistry* **2009**, 284, 12829-12836.

- [117] List, K.; Kosa, P.; Szabo, R.; Bey, A. L.; Wang, C. B.; Molinolo, A.; Bugge, T. H. Epithelial integrity is maintained by a matriptase-dependent proteolytic pathway. *Am J Pathol* **2009**, 175, 1453-1463.
- [118] List, K.; Szabo, R.; Wertz, P. W.; Segre, J.; Haudenschild, C. C.; Kim, S. Y.; Bugge, T. H. Loss of proteolytically processed filaggrin caused by epidermal deletion of Matriptase/MT-SP1. *The Journal of cell biology* **2003**, 163, 901-10.
- [119] Milner, J.; Patel, A.; Rowan, A. D. Emerging Roles of Serine Proteinases in Tissue Turnover in Arthritis. *Arthritis and rheumatism* **2008**, 58, 3644-56.
- [120] Milner, J. M.; Elliott, S.-F.; Cawston, T. E. Activation of procollagenases is a key control point in cartilage collagen degradation: Interaction of serine and metalloproteinase pathways. *Arthritis & Rheumatism* **2001**, 44, 2084-2096.
- [121] Kang, J. Y.; Dolled-Filhart, M.; Ocal, I. T.; Singh, B.; Lin, C.-Y.; Dickson, R. B.; Rimm, D. L.; Camp, R. L. Tissue Microarray Analysis of Hepatocyte Growth Factor/Met Pathway Components Reveals a Role for Met, Matriptase, and Hepatocyte Growth Factor Activator Inhibitor 1 in the Progression of Node-negative Breast Cancer. *Cancer research* **2003**, 63, 1101-1105.
- [122] Lee, J.-W.; Yong Song, S.; Choi, J.-J.; Lee, S.-J.; Kim, B.-G.; Park, C.-S.; Lee, J.-H.; Lin, C.-Y.; Dickson, R. B.; Bae, D.-S. Increased expression of matriptase is associated with histopathologic grades of cervical neoplasia. *Human Pathology* **2005**, 36, 626-633.
- [123] Saleem, M.; Adhami, V. M.; Zhong, W.; Longley, B. J.; Lin, C. Y.; Dickson, R. B.; Reagan-Shaw, S.; Jarrard, D. F.; Mukhtar, H. A novel biomarker for staging human prostate adenocarcinoma: overexpression of matriptase with concomitant loss of its inhibitor, hepatocyte growth factor activator inhibitor-1. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* **2006**, 15, 217-27.
- [124] Santin, A. D.; Cane', S.; Bellone, S.; Bignotti, E.; Palmieri, M.; De Las Casas, L. E.; Anfossi, S.; Roman, J. J.; O'Brien, T.; Pecorelli, S. The novel serine protease tumor-associated differentially expressed gene-15 (matriptase/MT-SP1) is highly overexpressed in cervical carcinoma. *Cancer* **2003**, 98, 1898-1904.
- [125] Alef, T.; Torres, S.; Hausser, I.; Metze, D.; Türsen, Ü.; Lestringant, G. G.; Hennies, H. C. Ichthyosis, Follicular Atrophoderma, and Hypotrichosis Caused by Mutations in ST14 Is Associated with Impaired Profilaggrin Processing. *Journal of Investigative Dermatology* **2009**, 129, 862-869.
- [126] Basel-Vanagaite, L.; Attia, R.; Ishida-Yamamoto, A.; Rainshtein, L.; Ben Amitai, D.; Lurie, R.; Pasmanik-Chor, M.; Indelman, M.; Zvulunov, A.; Saban, S.; Magal, N.; Sprecher, E.; Shohat, M. Autosomal recessive ichthyosis with hypotrichosis caused by a mutation in ST14, encoding type II transmembrane serine protease matriptase. *American journal of human genetics* **2007**, 80, 467-477.
- [127] Förbs, D.; Thiel, S.; Stella, M.; Stuerzebecher, A.; Schweinitz, A.; Steinmetzer, T.; Stürzebecher, J.; Uhland, K. In vitro inhibition of matriptase prevents invasive growth of cell lines of prostate and colon carcinoma. *International journal of oncology* **2005**, 27, 1061-70.
- [128] Maiwald, A.; Hammami, M.; Wagner, S.; Heine, A.; Klebe, G.; Steinmetzer, T. Changing the selectivity profile - from substrate analog inhibitors of thrombin and factor Xa to potent matriptase inhibitors. *Journal of enzyme inhibition and medicinal chemistry* **2016**, 31, 89-97.

- [129] Mitchell, A. C.; Kannan, D.; Hunter, S. A.; Parra Sperberg, R. A.; Chang, C. H.; Cochran, J. R. Engineering a potent inhibitor of matriptase from the natural hepatocyte growth factor activator inhibitor type-1 (HAI-1) protein. *The Journal of biological chemistry* **2018**, 293, 4969-4980.
- [130] Uhland, K. Matriptase and its putative role in cancer. *Cellular and molecular life sciences : CMLS* **2007**, 63, 2968-78.
- [131] Lee, S. L.; Dickson, R. B.; Lin, C. Y. Activation of hepatocyte growth factor and urokinase/plasminogen activator by matriptase, an epithelial membrane serine protease. *The Journal of biological chemistry* **2000**, 275, 36720-5.
- [132] Kilpatrick, L. M.; Harris, R. L.; Owen, K. A.; Bass, R.; Ghorayeb, C.; Bar-Or, A.; Ellis, V. Initiation of plasminogen activation on the surface of monocytes expressing the type II transmembrane serine protease matriptase. *Blood* **2006**, 108, 2616-2623.
- [133] Camerer, E.; Barker, A.; Duong, D. N.; Ganesan, R.; Kataoka, H.; Cornelissen, I.; Darragh, M. R.; Hussain, A.; Zheng, Y. W.; Srinivasan, Y.; Brown, C.; Xu, S. M.; Regard, J. B.; Lin, C. Y.; Craik, C. S.; Kirchhofer, D.; Coughlin, S. R. Local protease signaling contributes to neural tube closure in the mouse embryo. *Developmental cell* **2010**, 18, 25-38.
- [134] Chen, M.; Chen, L.-M.; Lin, C.-Y.; Chai, K. X. The epidermal growth factor receptor (EGFR) is proteolytically modified by the Matriptase–Prostasin serine protease cascade in cultured epithelial cells. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **2008**, 1783, 896-903.
- [135] Riddick, A. C. P.; Shukla, C. J.; Pennington, C. J.; Bass, R.; Nuttall, R. K.; Hogan, A.; Sethia, K. K.; Ellis, V.; Collins, A. T.; Maitland, N. J.; Ball, R. Y.; Edwards, D. R. Identification of degradome components associated with prostate cancer progression by expression analysis of human prostatic tissues. *Br J Cancer* **2005**, 92, 2171-2180.
- [136] Jin, X.; Yagi, M.; Akiyama, N.; Hirosaki, T.; Higashi, S.; Lin, C.-Y.; B Dickson, R.; Kitamura, H.; Miyazaki, K. Matriptase activates stromelysin (MMP-3) and promotes tumor growth and angiogenesis. *Cancer science* **2007**, 97, 1327-34.
- [137] Ge, W.; Hu, H.; Ding, K.; Sun, L.; Zheng, S. Protein Interaction Analysis of ST14 Domains and Their Point and Deletion Mutants. *Journal of Biological Chemistry* **2006**, 281, 7406-12.
- [138] Tanimoto, H.; Shigemasa, K.; Tian, X.; Gu, L.; Beard, J. B.; Sawasaki, T.; O'Brien, T. J. Transmembrane serine protease TADG-15 (ST14/Matriptase/MT-SP1): expression and prognostic value in ovarian cancer. *Br J Cancer* **2005**, 92, 278-283.
- [139] Vogel, L. K.; Saebo, M.; Skjelbred, C. F.; Abell, K.; Pedersen, E. D.; Vogel, U.; Kure, E. H. The ratio of Matriptase/HAI-1 mRNA is higher in colorectal cancer adenomas and carcinomas than corresponding tissue from control individuals. *BMC cancer* **2006**, 6, 176.
- [140] Welman, A.; Sproul, D.; Mullen, P.; Muir, M.; Kinnaird, A. R.; Harrison, D. J.; Faratian, D.; Brunton, V. G.; Frame, M. C. Diversity of matriptase expression level and function in breast cancer. *PLoS One* **2012**, 7, e34182.
- [141] Zeng, L.; Cao, J.; Zhang, X. Expression of serine protease SNC19/matriptase and its inhibitor hepatocyte growth factor activator inhibitor type 1 in normal and malignant tissues of gastrointestinal tract. *World journal of gastroenterology* **2005**, 11, 6202-7.

- [142] List, K.; Szabo, R.; Molinolo, A.; Sriuranpong, V.; Redeye, V.; Murdock, T.; Burke, B.; Nielsen, B. S.; Gutkind, J. S.; Bugge, T. H. Deregulated matriptase causes ras-independent multistage carcinogenesis and promotes ras-mediated malignant transformation. *Genes & development* **2005**, 19, 1934-50.
- [143] Varela, F.; E. Hyland, T.; List, K. Physiological Functions and Role of Matriptase in Cancer: Strategies Directed at MET and RON Receptor Tyrosine Kinase Pathways. In *Extracellular Targeting of Cell Signaling in Cancer*, 2018; pp 91-124.
- [144] Milner, J. M.; Patel, A.; Davidson, R. K.; Swingle, T. E.; Desilets, A.; Young, D. A.; Kelso, E. B.; Donell, S. T.; Cawston, T. E.; Clark, I. M.; Ferrell, W. R.; Plevin, R.; Lockhart, J. C.; Leduc, R.; Rowan, A. D. Matriptase is a novel initiator of cartilage matrix degradation in osteoarthritis. *Arthritis & Rheumatism* **2010**, 62, 1955-1966.
- [145] Beaulieu, A.; Gravel, É.; Cloutier, A.; Marois, I.; Colombo, É.; Désilets, A.; Verreault, C.; Leduc, R.; Marsault, É.; Richter, M. V. Matriptase proteolytically activates influenza virus and promotes multicycle replication in the human airway epithelium. *Journal of virology* **2013**, 87, 4237-4251.
- [146] Wood, M. P.; Cole, A. L.; Eade, C. R.; Chen, L.-M.; Chai, K. X.; Cole, A. M. The HIV-1 gp41 ectodomain is cleaved by matriptase to produce a chemotactic peptide that acts through FPR2. *Immunology* **2014**, 142, 474-483.
- [147] Mosleh, N.; Dadras, H.; Mohammadi, A. Molecular quantitation of H9N2 avian influenza virus in various organs of broiler chickens using TaqMan real time PCR. *Journal of molecular and genetic medicine : an international journal of biomedical research* **2009**, 3, 152-157.
- [148] Guo, Y. J.; Krauss, S.; Senne, D. A.; Mo, I.; Lo, K. S.; Xiong, X. P.; Norwood, M.; Shortridge, K. F.; Webster, R. G.; Guan, Y. Characterization of the Pathogenicity of Members of the Newly Established H9N2 Influenza Virus Lineages in Asia. *Virology* **2000**, 267, 279-288.
- [149] Fang, J.-D.; Chou, H.-C.; Tung, H.-H.; Huang, P.-Y.; Lee, S.-L. Endogenous expression of matriptase in neural progenitor cells promotes cell migration and neuron differentiation. *The Journal of biological chemistry* **2011**, 286, 5667-5679.
- [150] Straus, M. R.; Whittaker, G. A peptide-based approach to evaluate the adaptability of influenza A virus to humans based on its hemagglutinin proteolytic cleavage site. *PLoS One* **2017**, 12, e0174827.
- [151] Galloway, S. E.; Reed, M. L.; Russell, C. J.; Steinhauer, D. A. Influenza HA subtypes demonstrate divergent phenotypes for cleavage activation and pH of fusion: implications for host range and adaptation. *PLoS Pathog* **2013**, 9, e1003151-e1003151.
- [152] Lambertz, R. L. O.; Gerhauser, I.; Nehlmeier, I.; Leist, S. R.; Kollmus, H.; Pöhlmann, S.; Schughart, K. Tmprss2 knock-out mice are resistant to H10 influenza A virus pathogenesis. *Journal of General Virology* **2019**, 100, 1073-1078.
- [153] Horimoto, T.; Nakayama, K.; Smeekens, S. P.; Kawaoka, Y. Proprotein-processing endoproteases PC6 and furin both activate hemagglutinin of virulent avian influenza viruses. *Journal of virology* **1994**, 68, 6074-8.
- [154] Guan, Y.; Poon, L. L. M.; Cheung, C. Y.; Ellis, T. M.; Lim, W.; Lipatov, A. S.; Chan, K. H.; Sturm-Ramirez, K. M.; Cheung, C. L.; Leung, Y. H. C.; Yuen, K. Y.; Webster, R. G.; Peiris, J. S. M. H5N1 influenza: a protean pandemic threat. *Proceedings of the National Academy of Sciences of the United States of America* **2004**, 101, 8156-8161.

- [155] Taubenberger, J. K.; Morens, D. M.; Fauci, A. S. The next influenza pandemic: can it be predicted? *JAMA* **2007**, 297, 2025-2027.
- [156] Xu, K. M.; Li, K. S.; Smith, G. J. D.; Li, J. W.; Tai, H.; Zhang, J. X.; Webster, R. G.; Peiris, J. S. M.; Chen, H.; Guan, Y. Evolution and Molecular Epidemiology of H9N2 Influenza A Viruses from Quail in Southern China, 2000 to 2005. *Journal of virology* **2007**, 81, 2635.
- [157] Matrosovich, M. N.; Krauss, S.; Webster, R. G. H9N2 Influenza A Viruses from Poultry in Asia Have Human Virus-like Receptor Specificity. *Virology* **2001**, 281, 156-162.
- [158] Scheiblaue, H.; Reinacher, M.; Tashiro, M.; Rott, R. Interactions between Bacteria and Influenza A Virus in the Development of Influenza Pneumonia. *The Journal of Infectious Diseases* **1992**, 166, 783-791.
- [159] Steinmetzer, T.; Schweinitz, A.; Stürzebecher, A.; Dönnecke, D.; Uhland, K.; Schuster, O.; Steinmetzer, P.; Müller, F.; Friedrich, R.; Than, M. E.; Bode, W.; Stürzebecher, J. Secondary Amides of Sulfonylated 3-Amidinophenylalanine. New Potent and Selective Inhibitors of Matriptase. *Journal of Medicinal Chemistry* **2006**, 49, 4116-4126.
- [160] Takeuchi, T.; Shuman, M. A.; Craik, C. S. Reverse biochemistry: use of macromolecular protease inhibitors to dissect complex biological processes and identify a membrane-type serine protease in epithelial cancer and normal tissue. *Proceedings of the National Academy of Sciences of the United States of America* **1999**, 96, 11054-11061.
- [161] Désilets, A.; Longpré, J.-M.; Beaulieu, M.-E.; Leduc, R. Inhibition of human matriptase by eglin c variants. *FEBS letters* **2006**, 580, 2227-32.
- [162] Friedrich, R.; Fuentes-Prior, P.; Ong, E.; Coombs, G.; Hunter, M.; Oehler, R.; Pierson, D.; Gonzalez, R.; Huber, R.; Bode, W.; Madison, E. L. Catalytic domain structures of MT-SP1/matriptase, a matrix-degrading transmembrane serine proteinase. *The Journal of biological chemistry* **2002**, 277, 2160-8.
- [163] Zhao, B.; Yuan, C.; Li, R.; Qu, D.; Huang, M.; Ngo, J. C. Crystal structures of matriptase in complex with its inhibitor hepatocyte growth factor activator inhibitor-1. *The Journal of biological chemistry* **2013**, 288, 11155-64.
- [164] Kadokura, H.; Katzen, F.; Beckwith, J. Protein disulfide bond formation in prokaryotes. *Annual review of biochemistry* **2003**, 72, 111-35.
- [165] Messens, J.; Collet, J. F. Pathways of disulfide bond formation in Escherichia coli. *The international journal of biochemistry & cell biology* **2006**, 38, 1050-62.
- [166] Häußler, D.; Schulz-Fincke, A.-C.; Beckmann, A.-M.; Keils, A.; Gilberg, E.; Mangold, M.; Bajorath, J.; Stirnberg, M.; Steinmetzer, T.; Gütschow, M. A Fluorescent-Labeled Phosphono Bisbenzguanidine As an Activity-Based Probe for Matriptase. *Chemistry – A European Journal* **2017**, 23, 5205-5209.
- [167] Studier, F. W.; Moffatt, B. A. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *Journal of molecular biology* **1986**, 189, 113-30.
- [168] Baneyx, F. Recombinant protein expression in Escherichia coli. *Current opinion in biotechnology* **1999**, 10, 411-21.
- [169] Graumann, K.; Premstaller, A. Manufacturing of recombinant therapeutic proteins in microbial systems. *Biotechnology Journal* **2006**, 1, 164-186.
- [170] Iost, I.; Guillerez, J.; Dreyfus, M. Bacteriophage T7 RNA polymerase travels far ahead of ribosomes in vivo. *J Bacteriol* **1992**, 174, 619-622.

- [171] Lopez, P.; Iost, I.; Dreyfus, M. Use of a Transfer-Rna as a Transcriptional Reporter - the T7 Late Promoter Is Extremely Efficient in Escherichia-Coli but Its Transcripts Are Poorly Expressed. *Nucleic Acids Research* **1994**, 22, 2434-2434.
- [172] Clark, E. D. B. Refolding of recombinant proteins. *Current opinion in biotechnology* **1998**, 9, 157-63.
- [173] Vallejo, L. F.; Rinas, U. Strategies for the recovery of active proteins through refolding of bacterial inclusion body proteins. *Microb Cell Fact* **2004**, 3, 11-11.
- [174] Yamaguchi, H.; Miyazaki, M. Refolding Techniques for Recovering Biologically Active Recombinant Proteins from Inclusion Bodies. *Biomolecules* **2014**, 4, 235-51.
- [175] Tsumoto, K.; Ejima, D.; Kumagai, I.; Arakawa, T. Practical considerations in refolding proteins from inclusion bodies. *Protein expression and purification* **2003**, 28, 1-8.
- [176] Tsumoto, K.; Umetsu, M.; Kumagai, I.; Ejima, D.; Philo, J. S.; Arakawa, T. Role of Arginine in Protein Refolding, Solubilization, and Purification. *Biotechnology Progress* **2004**, 20, 1301-1308.
- [177] Hudson, D. A.; Gannon, S. A.; Thorpe, C. Oxidative protein folding: from thiol-disulfide exchange reactions to the redox poise of the endoplasmic reticulum. *Free radical biology & medicine* **2015**, 80, 171-182.
- [178] Camacho, C. J.; Thirumalai, D. Modeling the role of disulfide bonds in protein folding: Entropic barriers and pathways. *Proteins: Structure, Function, and Bioinformatics* **1995**, 22, 27-40.
- [179] Clark, E. D. Protein refolding for industrial processes. *Current opinion in biotechnology* **2001**, 12, 202-7.
- [180] Kohyama, K.; Matsumoto, T.; Imoto, T. Refolding of an unstable lysozyme by gradient removal of a solubilizer and gradient addition of a stabilizer. *The Journal of Biochemistry* **2009**, 147, 427-431.
- [181] Timasheff, S. Protein-solvent preferential interactions, protein hydration, and the modulation of biochemical reactions by solvent components. *Proceedings of the National Academy of Sciences of the United States of America* **2002**, 99, 9721-6.
- [182] Dechavanne, V.; Barrillat, N.; Borlat, F.; Hermant, A.; Magnenat, L.; Pâquet, M.; Antonsson, B.; Chevalet, L. A high-throughput protein refolding screen in 96-well format combined with design of experiments to optimize the refolding conditions. *Protein expression and purification* **2011**, 75, 192-203.
- [183] Vincentelli, R.; Canaan, S.; Campanacci, V.; Valencia, C.; Maurin, D.; Frassinetti, F.; Scappucini-Calvo, L.; Bourne, Y.; Cambillau, C.; Bignon, C. High-throughput automated refolding screening of inclusion bodies. *Protein Science* **2004**, 13, 2782-2792.
- [184] Wang, Y.; van Oosterwijk, N.; Ali, A. M.; Adawy, A.; Anindya, A. L.; Dömling, A. S. S.; Groves, M. R. A Systematic Protein Refolding Screen Method using the DGR Approach Reveals that Time and Secondary TSA are Essential Variables. *Scientific Reports* **2017**, 7, 9355.
- [185] Skala, W.; Goettig, P.; Brandstetter, H. Do-it-yourself histidine-tagged bovine enterokinase: a handy member of the protein engineer's toolbox. *J Biotechnol* **2013**, 168, 421-425.

- [186] Bender, M. L.; Cantón, M. L. B.; Blakeley, R. L.; Brubacher, L. J.; Feder, J.; Gunter, C. R.; Kézdy, F. J.; Killheffer, J. V.; Marshall, T. H.; Miller, C. G.; Roeske, R. W.; Stoops, J. K. The Determination of the Concentration of Hydrolytic Enzyme Solutions:  $\alpha$ -Chymotrypsin, Trypsin, Papain, Elastase, Subtilisin, and Acetylcholinesterase1. *Journal of the American Chemical Society* **1966**, 88, 5890-5913.
- [187] Graham Knight, C. Active-site titration of peptidases. *Methods in enzymology* **1995**, 248, 85-101.
- [188] Jameson, G. W.; Roberts, D. V.; Adams, R. W.; Kyle, W. S.; Elmore, D. T. Determination of the operational molarity of solutions of bovine alpha-chymotrypsin, trypsin, thrombin and factor Xa by spectrofluorimetric titration. *The Biochemical journal* **1973**, 131, 107-117.
- [189] Brown, C. M.; Ray, M.; Eroy-Reveles, A. A.; Egea, P.; Tajon, C.; Craik, C. S. Peptide length and leaving-group sterics influence potency of peptide phosphonate protease inhibitors. *Chem Biol* **2011**, 18, 48-57.
- [190] Farady, C. J.; Egea, P. F.; Schneider, E. L.; Darragh, M. R.; Craik, C. S. Structure of an Fab-protease complex reveals a highly specific non-canonical mechanism of inhibition. *Journal of molecular biology* **2008**, 380, 351-60.
- [191] Goswami, R.; Mukherjee, S.; Wohlfahrt, G.; Ghadiyaram, C.; Nagaraj, J.; Chandra, B. R.; Sistla, R. K.; Satyam, L. K.; Samiulla, D. S.; Moilanen, A.; Subramanya, H. S.; Ramachandra, M. Discovery of Pyridyl Bis(oxy)dibenzimidamide Derivatives as Selective Matriptase Inhibitors. *ACS medicinal chemistry letters* **2013**, 4, 1152-7.
- [192] Yuan, C.; Chen, L.; Meehan, E. J.; Daly, N.; Craik, D. J.; Huang, M.; Ngo, J. C. Structure of catalytic domain of Matriptase in complex with Sunflower trypsin inhibitor-1. *BMC structural biology* **2011**, 11, 30.
- [193] D'Arcy, A.; Bergfors, T.; Cowan-Jacob, S. W.; Marsh, M. Microseed matrix screening for optimization in protein crystallization: what have we learned? *Acta Crystallogr F Struct Biol Commun* **2014**, 70, 1117-1126.
- [194] Bergfors, T. Seeds to crystals. *Journal of structural biology* **2003**, 142, 66-76.
- [195] Ireton, G. C.; Stoddard, B. L. Microseed matrix screening to improve crystals of yeast cytosine deaminase. *Acta crystallographica. Section D, Biological crystallography* **2004**, 60, 601-5.
- [196] D'Arcy, A.; Villard, F.; Marsh, M. An automated microseed matrix-screening method for protein crystallization. *Acta crystallographica. Section D, Biological crystallography* **2007**, 63, 550-4.
- [197] Goswami, R.; Wohlfahrt, G.; Mukherjee, S.; Ghadiyaram, C.; Nagaraj, J.; Satyam, L. K.; Subbarao, K.; Gopinath, S.; Krishnamurthy, N. R.; Subramanya, H. S.; Ramachandra, M. Discovery of O-(3-carbamimidoylphenyl)-l-serine amides as matriptase inhibitors using a fragment-linking approach. *Bioorganic & medicinal chemistry letters* **2015**, 25, 616-20.
- [198] Schneider, E. L.; Lee, M. S.; Baharuddin, A.; Goetz, D. H.; Farady, C. J.; Ward, M.; Wang, C. I.; Craik, C. S. A reverse binding motif that contributes to specific protease inhibition by antibodies. *Journal of molecular biology* **2012**, 415, 699-715.
- [199] Gustafsson, D.; Bylund, R.; Antonsson, T.; Nilsson, I.; Nystrom, J. E.; Eriksson, U.; Bredberg, U.; Teger-Nilsson, A. C. A new oral anticoagulant: the 50-year challenge. *Nature reviews. Drug discovery* **2004**, 3, 649-59.



- [200] Froriep, D.; Clement, B.; Bittner, F.; Mendel, R.; Reichmann, D.; Schmalix, W.; Havemeyer, A. Activation of the anti-cancer agent upamostat by the mARC enzyme system. *Xenobiotica* **2013**, 43, 780-784.
- [201] Colombo, E.; Désilets, A.; Duchêne, D.; Chagnon, F.; Najmanovich, R.; Leduc, R.; Marsault, E. Design and synthesis of potent, selective inhibitors of matriptase. *ACS medicinal chemistry letters* **2012**, 3, 530-534.
- [202] Wu, S.-R.; Teng, C.-H.; Tu, Y.-T.; Ko, C.-J.; Cheng, T.-S.; Lan, S.-W.; Lin, H.-Y.; Lin, H.-H.; Tu, H.-F.; Hsiao, P.-W.; Huang, H.-P.; Chen, C.-H.; Lee, M.-S. The Kunitz Domain I of Hepatocyte Growth Factor Activator Inhibitor-2 Inhibits Matriptase Activity and Invasive Ability of Human Prostate Cancer Cells. *Scientific Reports* **2017**, 7, 15101.
- [203] Long, Y.-Q.; L Lee, S.; Lin, C.-Y.; J Enyedy, I.; Wang, S.; Li, P.; B Dickson, R.; P Roller, P. Synthesis and evaluation of the sunflower derived trypsin inhibitor as a potent inhibitor of the type II transmembrane serine protease, matriptase. *Bioorganic & medicinal chemistry letters* **2001**, 11, 2515-9.
- [204] Li, P.; Jiang, S.; Lee, S.-L.; Lin, C.-Y.; Johnson, M.; B Dickson, R.; J Michejda, C.; P Roller, P. Design and Synthesis of Novel and Potent Inhibitors of the Type II Transmembrane Serine Protease, Matriptase, Based upon the Sunflower Trypsin Inhibitor-1. *Journal of medicinal chemistry* **2007**, 50, 5976-83.
- [205] Fittler, H.; Avrutina, O.; Empting, M.; Kolmar, H. Potent inhibitors of human matriptase-1 based on the scaffold of sunflower trypsin inhibitor. *Journal of Peptide Science* **2014**, 20, 415-420.
- [206] Fittler, H.; Avrutina, O.; Glotzbach, B.; Empting, M.; Kolmar, H. Combinatorial tuning of peptidic drug candidates: high-affinity matriptase inhibitors through incremental structure-guided optimization. *Organic & biomolecular chemistry* **2013**, 11, 1848-1857.
- [207] Avrutina, O.; Fittler, H.; Glotzbach, B.; Kolmar, H.; Empting, M. Between two worlds: a comparative study on in vitro and in silico inhibition of trypsin and matriptase by redox-stable SFTI-1 variants at near physiological pH. *Organic & biomolecular chemistry* **2012**, 10, 7753-62.
- [208] Sun, J.; Pons, J.; Craik, C. Potent and Selective Inhibition of Membrane-Type Serine Protease 1 by Human Single-Chain Antibodies. *Biochemistry* **2003**, 42, 892-900.
- [209] Enyedy, I. J.; Lee, S.-L.; Kuo, A. H.; Dickson, R. B.; Lin, C.-Y.; Wang, S. Structure-Based Approach for the Discovery of Bis-benzamidines as Novel Inhibitors of Matriptase. *Journal of Medicinal Chemistry* **2001**, 44, 1349-1355.
- [210] Goswami, R.; Mukherjee, S.; Ghadiyaram, C.; Wohlfahrt, G.; Sistla, R. K.; Nagaraj, J.; Satyam, L. K.; Subbarao, K.; Palakurthy, R. K.; Gopinath, S.; Krishnamurthy, N. R.; Ikonen, T.; Moilanen, A.; Subramanya, H. S.; Kallio, P.; Ramachandra, M. Structure-guided discovery of 1,3,5 tri-substituted benzenes as potent and selective matriptase inhibitors exhibiting in vivo antitumor efficacy. *Bioorganic & medicinal chemistry* **2014**, 22, 3187-203.
- [211] Furtmann, N.; Häußler, D.; Scheidt, T.; Stirnberg, M.; Steinmetzer, T.; Bajorath, J.; Gütschow, M. Limiting the Number of Potential Binding Modes by Introducing Symmetry into Ligands: Structure-Based Design of Inhibitors for Trypsin-Like Serine Proteases. *Chemistry - A European Journal* **2015**, 22, 610-625.
- [212] Sisay, M. T.; Steinmetzer, T.; Stirnberg, M.; Maurer, E.; Hammami, M.; Bajorath, J.; Gütschow, M. Identification of the First Low-Molecular-Weight Inhibitors of Matriptase-2. *Journal of Medicinal Chemistry* **2010**, 53, 5523-5535.

- [213] Galkin, A. V.; Mullen, L.; Fox, W. D.; Brown, J.; Duncan, D.; Moreno, O.; Madison, E. L.; Agus, D. B. CVS-3983, a selective matriptase inhibitor, suppresses the growth of androgen independent prostate tumor xenografts. *The Prostate* **2004**, 61, 228-235.
- [214] Han, Z.; Harris, P. K. W.; Jones, D. E.; Chugani, R.; Kim, T.; Agarwal, M.; Shen, W.; Wildman, S. A.; Janetka, J. W. Inhibitors of HGFA, Matriptase, and Hepsin Serine Proteases: A Nonkinase Strategy to Block Cell Signaling in Cancer. *ACS medicinal chemistry letters* **2014**, 5, 1219-1224.
- [215] Stürzebecher, J.; Vieweg, H.; Steinmetzer, T.; Schweinitz, A.; Stubbs, M. T.; Renatus, M.; Wikström, P. 3-Amidinophenylalanine-based inhibitors of urokinase. *Bioorganic & medicinal chemistry letters* **1999**, 9, 3147-3152.
- [216] Kotthaus, J.; Steinmetzer, T.; Loch, A.; Clement, B. Analysis of highly potent amidine containing inhibitors of serine proteases and their N -hydroxylated prodrugs (amidoximes). *Journal of enzyme inhibition and medicinal chemistry* **2011**, 26, 115-22.
- [217] Steinmetzer, T.; Dönnecke, D.; Korsonewski, M.; Neuwirth, C.; Steinmetzer, P.; Schulze, A.; Saupe, S. M.; Schweinitz, A. Modification of the N-terminal sulfonyl residue in 3-amidinophenylalanine-based matriptase inhibitors. *Bioorganic & medicinal chemistry letters* **2009**, 19, 67-73.
- [218] Hammami, M.; Rühmann, E.; Maurer, E.; Heine, A.; Gütschow, M.; Klebe, G.; Steinmetzer, T. New 3-amidinophenylalanine-derived inhibitors of matriptase. *MedChemComm* **2012**, 3, 807-813.
- [219] Zuo, K.; Qi, Y.; Yuan, C.; Jiang, L.; Xu, P.; Hu, J.; Huang, M.; Li, J. Specifically targeting cancer proliferation and metastasis processes: the development of matriptase inhibitors. *Cancer metastasis reviews* **2019**.
- [220] Setyono-Han, B.; Stürzebecher, J.; Schmalix, W.; Muehlenweg, B.; Sieuwerts, A.; Timmermans, A.; Magdolen, V.; Schmitt, M.; Klijn, J.; Foekens, J. Suppression of rat breast cancer metastasis and reduction of primary tumour growth by the small synthetic urokinase inhibitor WX-UKI. *Thrombosis and haemostasis* **2005**, 93, 779-86.
- [221] Bergner, A.; Bauer, M.; Brandstetter, H.; Stürzebecher, J.; Bode, W. The X-Ray Crystal Structure of Thrombin in Complex with N $\alpha$ -2-Naphthylsulfonyl-L-3-Amidino-Phenylalanyl-4-Methylpiperidide: The Beneficial Effect of Filling Out an Empty Cavity. *Journal of enzyme inhibition* **1995**, 9, 101-110.
- [222] Renatus, M.; Bode, W.; Huber, R.; Stürzebecher, J.; Stubbs, M. T. Structural and Functional Analyses of Benzamidine-Based Inhibitors in Complex with Trypsin: Implications for the Inhibition of Factor Xa, tPA, and Urokinase. *Journal of Medicinal Chemistry* **1998**, 41, 5445-5456.
- [223] Pilgram, O. Synthese und Charakterisierung neuer Matriptase-Hemmstoffe vom 3-Amidinophenylalanin-Typ. Master Thesis, Philipps University Marburg, **2017**.
- [224] Schweinitz, A.; Dönnecke, D.; Ludwig, A.; Steinmetzer, P.; Schulze, A.; Kotthaus, J.; Wein, S.; Clement, B.; Steinmetzer, T. Incorporation of neutral C-terminal residues in 3-amidinophenylalanine-derived matriptase inhibitors. *Bioorganic & medicinal chemistry letters* **2009**, 19, 1960-1965.
- [225] Rautio, J.; Meanwell, N. A.; Di, L.; Hageman, M. J. The expanding role of prodrugs in contemporary drug design and development. *Nature Reviews Drug Discovery* **2018**, 17, 559-587.

- [226] Katz, B. A.; Sprengeler, P. A.; Luong, C.; Verner, E.; Elrod, K.; Kirtley, M.; Janc, J.; Spencer, J. R.; Breitenbucher, J. G.; Hui, H.; McGee, D.; Allen, D.; Martelli, A.; Mackman, R. L. Engineering inhibitors highly selective for the S1 sites of Ser190 trypsin-like serine protease drug targets. *Chem Biol* **2001**, 8, 1107-1121.
- [227] Brandstetter, H.; Kuhne, A.; Bode, W.; Huber, R.; von der Saal, W.; Wirthensohn, K.; Engh, R. A. X-ray structure of active site-inhibited clotting factor Xa. Implications for drug design and substrate recognition. *The Journal of biological chemistry* **1996**, 271, 29988-92.
- [228] Kumar, N.; Hendriks, B. S.; Janes, K. A.; de Graaf, D.; Lauffenburger, D. A. Applying computational modeling to drug discovery and development. *Drug Discovery Today* **2006**, 11, 806-811.
- [229] Fortin, S.; Moreau, E.; Lacroix, J.; Teulade, J.-C.; Patenaude, A.; C-Gaudreault, R. N-Phenyl-N'-(2-chloroethyl)urea analogues of combretastatin A-4: Is the N-phenyl-N'-(2-chloroethyl)urea pharmacophore mimicking the trimethoxy phenyl moiety? *Bioorganic & medicinal chemistry letters* **2007**, 17, 2000-2004.
- [230] Sperl, S.; Bergner, A.; Stürzebecher, J.; Magdolen, V.; Bode, W.; Moroder, L. Urethanyl-3-Amidinophenylalanine Derivatives as Inhibitors of Factor Xa. X-Ray Crystal Structure of a Trypsin/Inhibitor Complex and Modeling Studies. *Biological Chemistry* **2000**, 381, 321-329.
- [231] Zürcher, M.; Diederich, F. Structure-Based Drug Design: Exploring the Proper Filling of Apolar Pockets at Enzyme Active Sites. *The Journal of Organic Chemistry* **2008**, 73, 4345-4361.
- [232] Politzer, P.; Lane, P.; Concha, M. C.; Ma, Y.; Murray, J. S. An overview of halogen bonding. *Journal of Molecular Modeling* **2007**, 13, 305-311.
- [233] Cramer, J.; Sager, C. P.; Ernst, B. Hydroxyl Groups in Synthetic and Natural-Product-Derived Therapeutics: A Perspective on a Common Functional Group. *J Med Chem* **2019**, 62, 8915-8930.
- [234] Paoloni-Giacobino, A.; Chen, H.; Peitsch, M. C.; Rossier, C.; Antonarakis, S. E. Cloning of the TMPRSS2 Gene, Which Encodes a Novel Serine Protease with Transmembrane, LDLRA, and SRCR Domains and Maps to 21q22.3. *Genomics* **1997**, 44, 309-320.
- [235] Kurachi, K.; Torres-Rosado, A.; Tsuji, A. Hepsin. *Methods in enzymology* **1994**, 244, 100-14.
- [236] Böttcher-Friebertshäuser, E.; Stein, D. A.; Klenk, H.-D.; Garten, W. Inhibition of influenza virus infection in human airway cell cultures by an antisense peptide-conjugated morpholino oligomer targeting the hemagglutinin-activating protease TMPRSS2. *Journal of virology* **2011**, 85, 1554-1562.
- [237] Ko, C.-J.; Huang, C.-C.; Lin, H.-Y.; Juan, C.-P.; Lan, S.-W.; Shyu, H.-Y.; Wu, S.-R.; Hsiao, P.-W.; Huang, h.-P.; Shun, C.; Lee, M.-S. Androgen-Induced TMPRSS2 Activates Matriptase and Promotes Extracellular Matrix Degradation, Prostate Cancer Cell Invasion, Tumor Growth, and Metastasis. *Cancer research* **2015**, 75, 2949-2960.
- [238] Esumi, M.; Ishibashi, M.; Yamaguchi, H.; Nakajima, S.; Tai, Y.; Kikuta, S.; Sugitani, M.; Takayama, T.; Tahara, M.; Takeda, M.; Wakita, T. Transmembrane serine protease TMPRSS2 activates hepatitis C virus infection. *Hepatology* **2015**, 61, 437-46.

- [239] Chen, M.; Chen, L.-M.; Lin, C.-Y.; Chai, K. X. Hepsin activates prostaticin and cleaves the extracellular domain of the epidermal growth factor receptor. *Molecular and Cellular Biochemistry* **2010**, 337, 259-266.
- [240] Lin, B.; Ferguson, C.; White, J. T.; Wang, S.; Vessella, R.; True, L. D.; Hood, L.; Nelson, P. S. Prostate-localized and Androgen-regulated Expression of the Membrane-bound Serine Protease TMPRSS2. *Cancer research* **1999**, 59, 4180-4184.
- [241] Szabo, R.; Bugge, T. H. Type II transmembrane serine proteases in development and disease. *The international journal of biochemistry & cell biology* **2008**, 40, 1297-1316.
- [242] Vaarala, M. H.; Porvari, K. S.; Kellokumpu, S.; Kyllönen, A. P.; Vihko, P. T. Expression of transmembrane serine protease TMPRSS2 in mouse and human tissues. *The Journal of Pathology* **2001**, 193, 134-140.
- [243] Kim, T. S.; Heinlein, C.; Hackman, R. C.; Nelson, P. S. Phenotypic analysis of mice lacking the Tmprss2-encoded protease. *Mol Cell Biol* **2006**, 26, 965-975.
- [244] Donaldson, S.; Hirsh, A.; Chen Li, D.; Holloway, G.; Chao, J.; Boucher, R.; E Gabriel, S. Regulation of the Epithelial Sodium Channel by Serine Proteases in Human Airways. *The Journal of biological chemistry* **2002**, 277, 8338-45.
- [245] Bertram, S.; Heurich, A.; Lavender, H.; Gierer, S.; Danisch, S.; Perin, P.; Lucas, J. M.; Nelson, P. S.; Pöhlmann, S.; Soilleux, E. J. Influenza and SARS-coronavirus activating proteases TMPRSS2 and HAT are expressed at multiple sites in human respiratory and gastrointestinal tracts. *PLoS One* **2012**, 7, e35876-e35876.
- [246] Lucas, J. M.; Heinlein, C.; Kim, T.; Hernandez, S. A.; Malik, M. S.; True, L. D.; Morrissey, C.; Corey, E.; Montgomery, B.; Mostaghel, E.; Clegg, N.; Coleman, I.; Brown, C. M.; Schneider, E. L.; Craik, C.; Simon, J. A.; Bedalov, A.; Nelson, P. S. The androgen-regulated protease TMPRSS2 activates a proteolytic cascade involving components of the tumor microenvironment and promotes prostate cancer metastasis. *Cancer discovery* **2014**, 4, 1310-25.
- [247] Tomlins, S. A.; Rhodes, D. R.; Perner, S.; Dhanasekaran, S. M.; Mehra, R.; Sun, X. W.; Varambally, S.; Cao, X.; Tchinda, J.; Kuefer, R.; Lee, C.; Montie, J. E.; Shah, R. B.; Pienta, K. J.; Rubin, M. A.; Chinnaiyan, A. M. Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. *Science (New York, N.Y.)* **2005**, 310, 644-8.
- [248] Hossain, D.; Bostwick, D. G. Significance of the TMPRSS2:ERG gene fusion in prostate cancer. *BJU International* **2013**, 111, 834-835.
- [249] Vaarala, M. H.; Porvari, K.; Kyllönen, A.; Vihko, P. Differentially Expressed Genes in Two LNCaP Prostate Cancer Cell Lines Reflecting Changes during Prostate Cancer Progression. *Laboratory Investigation* **2000**, 80, 1259-1268.
- [250] Kumar-Sinha, C.; Tomlins, S. A.; Chinnaiyan, A. M. Recurrent gene fusions in prostate cancer. *Nature reviews. Cancer* **2008**, 8, 497-511.
- [251] Wilson, S.; Greer, B.; Hooper, J.; Zijlstra, A.; Walker, B.; Quigley, J.; Hawthorne, S. The membrane-anchored serine protease, TMPRSS2, activates PAR-2 in prostate cancer cells. *The Biochemical journal* **2005**, 388, 967-972.
- [252] Fan, B.; Brennan, J.; Grant, D.; Peale, F.; Rangell, L.; Kirchhofer, D. Hepatocyte growth factor activator inhibitor-1 (HAI-1) is essential for the integrity of basement membranes in the developing placental labyrinth. *Developmental Biology* **2007**, 303, 222-230.

- [253] Jiang, W.; Hiscox, S.; Matsumoto, K.; Nakamura, T. Hepatocyte growth factor/scatter factor, its molecular, cellular and clinical implications in cancer. *Critical reviews in oncology/hematology* **1999**, 29, 209-48.
- [254] Limburg, H.; Harbig, A.; Bestle, D.; A. Stein, D.; M. Moulton, H.; Jaeger, J.; Janga, H.; Hards, K.; Koepke, J.; Schulte, L.; Koczulla, R.; Schmeck, B.; Klenk, H.-D.; Böttcher-Friebertshäuser, E. TMPRSS2 is the major activating protease of influenza A virus in primary human airway cells and influenza B virus in human type II pneumocytes. *Journal of virology* **2019**, 93, e00649-19.
- [255] Glowacka, I.; Bertram, S.; Müller, M. A.; Allen, P.; Soilleux, E.; Pfefferle, S.; Steffen, I.; Tsegaye, T. S.; He, Y.; Gnirss, K.; Niemeyer, D.; Schneider, H.; Drosten, C.; Pöhlmann, S. Evidence that TMPRSS2 activates the severe acute respiratory syndrome coronavirus spike protein for membrane fusion and reduces viral control by the humoral immune response. *Journal of virology* **2011**, 85, 4122-4134.
- [256] Shulla, A.; Heald-Sargent, T.; Subramanya, G.; Zhao, J.; Perlman, S.; Gallagher, T. A transmembrane serine protease is linked to the severe acute respiratory syndrome coronavirus receptor and activates virus entry. *Journal of virology* **2011**, 85, 873-882.
- [257] Abe, M.; Tahara, M.; Sakai, K.; Yamaguchi, H.; Kanou, K.; Shirato, K.; Kawase, M.; Noda, M.; Kimura, H.; Matsuyama, S.; Fukuhara, H.; Mizuta, K.; Maenaka, K.; Ami, Y.; Esumi, M.; Kato, A.; Takeda, M. TMPRSS2 is an activating protease for respiratory parainfluenza viruses. *Journal of virology* **2013**, 87, 11930-11935.
- [258] Shirogane, Y.; Takeda, M.; Iwasaki, M.; Ishiguro, N.; Takeuchi, H.; Nakatsu, Y.; Tahara, M.; Kikuta, H.; Yanagi, Y. Efficient multiplication of human metapneumovirus in Vero cells expressing the transmembrane serine protease TMPRSS2. *Journal of virology* **2008**, 82, 8942-8946.
- [259] Esumi, M.; Ishibashi, M.; Yamaguchi, H.; Nakajima, S.; Tai, Y.; Kikuta, S.; Sugitani, M.; Takayama, T.; Tahara, M.; Takeda, M.; Wakita, T. Transmembrane serine protease TMPRSS2 activates hepatitis C virus infection. *Hepatology* **2015**, 61, 437-446.
- [260] Bertram, S.; Glowacka, I.; Blazejewska, P.; Soilleux, E.; Allen, P.; Danisch, S.; Steffen, I.; Choi, S. O. Y.; Park, Y.; Schneider, H.; Schughart, K.; Pöhlmann, S. TMPRSS2 and TMPRSS4 facilitate trypsin-independent spread of influenza virus in Caco-2 cells. *Journal of virology* **2010**, 84, 10016-25.
- [261] Stieneke-Gröber, A.; Vey, M.; Angliker, H.; Shaw, E.; Thomas, G.; Roberts, C.; Klenk, H. D.; Garten, W. Influenza virus hemagglutinin with multibasic cleavage site is activated by furin, a subtilisin-like endoprotease. *EMBO J* **1992**, 11, 2407-2414.
- [262] Böttcher-Friebertshäuser, E.; Freuer, C.; Sielaff, F.; Schmidt, S.; Eickmann, M.; Uhlenendorff, J.; Steinmetzer, T.; Klenk, H.-D.; Garten, W. Cleavage of influenza virus hemagglutinin by airway proteases TMPRSS2 and HAT differs in subcellular localization and susceptibility to protease inhibitors. *Journal of virology* **2010**, 84, 5605-5614.
- [263] Zhirnov, O. P.; Vorobjeva, I. V.; Ovcharenko, A. V.; Klenk, H. D. Intracellular cleavage of human influenza A virus hemagglutinin and its inhibition. *Biochemistry. Biokhimiia* **2003**, 68, 1020-6.
- [264] Ma, M.-J.; Liu, C.; Wu, M.-N.; Zhao, T.; Wang, G.-L.; Yang, Y.; Gu, H.-J.; Cui, P.-W.; Pang, Y.-Y.; Tan, Y.-Y.; Huang, H.; Lin, B.; Qin, J.-C.; Fang, L.-Q.; Cao, W.-C.; Chen, L.-L. Influenza A(H7N9) Virus Antibody Responses in Survivors 1 Year after Infection, China, 2017. *Emerg Infect Dis* **2018**, 24, 663-672.

- [265] Trombetta, C.; Piccirella, S.; Perini, D.; Kistner, O.; Montomoli, E. Emerging Influenza Strains in the Last Two Decades: A Threat of a New Pandemic? *Vaccines (Basel)* **2015**, *3*, 172-185.
- [266] Chen, H.; Yuan, H.; Gao, R.; Zhang, J.; Wang, D.; Xiong, Y.; Fan, G.; Yang, F.; Li, X.; Zhou, J.; Zou, S.; Yang, L.; Chen, T.; Dong, L.; Bo, H.; Zhao, X.; Zhang, Y.; Lan, Y.; Bai, T.; Dong, J.; Li, Q.; Wang, S.; Zhang, Y.; Li, H.; Gong, T.; Shi, Y.; Ni, X.; Li, J.; Zhou, J.; Fan, J.; Wu, J.; Zhou, X.; Hu, M.; Wan, J.; Yang, W.; Li, D.; Wu, G.; Feng, Z.; Gao, G. F.; Wang, Y.; Jin, Q.; Liu, M.; Shu, Y. Clinical and epidemiological characteristics of a fatal case of avian influenza A H10N8 virus infection: a descriptive study. *The Lancet* **2014**, *383*, 714-721.
- [267] Straus, M. R.; Whittaker, G. R. A peptide-based approach to evaluate the adaptability of influenza A virus to humans based on its hemagglutinin proteolytic cleavage site. *PLoS One* **2017**, *12*, e0174827.
- [268] Bertram, S.; Glowacka, I.; Steffen, I.; Kuhl, A.; Pohlmann, S. Novel insights into proteolytic cleavage of influenza virus hemagglutinin. *Reviews in medical virology* **2010**, *20*, 298-310.
- [269] Böttcher-Friebertshäuser, E.; Garten, W.; Matrosovich, M.; Dieter Klenk, H. *The Hemagglutinin: A Determinant of Pathogenicity*. Springer, Cham: **2014**; Vol. 385.
- [270] Lucas, J.; True, L.; Hawley, S.; Matsumura, M.; Morrissey, C.; Vessella, R.; Nelson, P. The androgen-regulated type II serine protease TMPRSS2 is differentially expressed and mislocalized in prostate adenocarcinoma. *The Journal of Pathology* **2008**, *215*, 118-125.
- [271] Bottcher-Friebertshauser, E.; Klenk, H. D.; Garten, W. Activation of influenza viruses by proteases from host cells and bacteria in the human airway epithelium. *Pathog Dis* **2013**, *69*, 87-100.
- [272] Coloma, M. J.; Hastings, A.; Wims, L. A.; Morrison, S. L. Novel vectors for the expression of antibody molecules using variable regions generated by polymerase chain reaction. *Journal of Immunological Methods* **1992**, *152*, 89-104.
- [273] Chen, Y.-W.; Lee, M.-S.; Lucht, A.; Chou, F.-P.; Huang, W.; Havighurst, T. C.; Kim, K.; Wang, J.-K.; Antalis, T. M.; Johnson, M. D.; Lin, C.-Y. TMPRSS2, a serine protease expressed in the prostate on the apical surface of luminal epithelial cells and released into semen in prostasomes, is misregulated in prostate cancer cells. *Am J Pathol* **2010**, *176*, 2986-2996.
- [274] Knappe, S.; Wu, F.; Rose Masikat, M.; Morser, J.; Wu, Q. Functional analysis of the transmembrane domain and activation cleavage of human corin - Design and characterization of a soluble corin. *The Journal of biological chemistry* **2004**, *278*, 52363-70.
- [275] Baghirova, S.; G Hughes, B.; Hendzel, M.; Schulz, R. Sequential fractionation and isolation of subcellular proteins from tissue or cultured cells. *MethodsX* **2015**, *2*, 440-445.
- [276] Niwa, H.; Yamamura, K.; Miyazaki, J. Efficient selection for high-expression transfectants with a novel eukaryotic vector. *Gene* **1991**, *108*, 193-9.
- [277] Gasteiger, E.; Hoogland, C.; Gattiker, A.; Duvaud, S. e.; R. Wilkins, M.; Appel, R.; Bairoch, A. Protein Identification and Analysis Tool on the ExPASy Server. In *The Proteomics Protocols Handbook*, Walker, J. M., Ed. Humana Press, **2007**; Vol. 112, pp 571-607.

- [278] Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res* **2000**, 28, 235-42.
- [279] Chevillard, F.; Rimmer, H.; Betti, C.; Pardon, E.; Ballet, S.; van Hilten, N.; Steyaert, J.; Diederich, W. E.; Kolb, P. Binding-Site Compatible Fragment Growing Applied to the Design of  $\beta$ 2-Adrenergic Receptor Ligands. *Journal of Medicinal Chemistry* **2018**, 61, 1118-1129.
- [280] Sambrook, J.; Fritsch, E. F.; Maniatis, T. *Molecular cloning: a laboratory manual*. Cold Spring Harbor, N.Y Cold Spring Harbor Laboratory, **1989**; p 1546 pp.
- [281] Kyhse-Andersen, J. Electrophoretic transfer of multiple gels: a simple apparatus without buffer tank for rapid transfer of proteins from polyacrylamide to nitrocellulose. *Journal of Biochemical and Biophysical Methods* **1984**, 10, 203-209.
- [282] Schulz, I. Permeabilizing cells: Some methods and applications for the study of intracellular processes. *Methods in enzymology* **1990**, 192, 280-300.
- [283] Dixon, M. The determination of enzyme inhibitor constants. *The Biochemical journal* **1953**, 55, 170-171.
- [284] Mueller, U.; Darowski, N.; Fuchs, M. R.; Förster, R.; Hellmig, M.; Paithankar, K. S.; Pühringer, S.; Steffien, M.; Zocher, G.; Weiss, M. S. Facilities for macromolecular crystallography at the Helmholtz-Zentrum Berlin. *Journal of Synchrotron Radiation* **2012**, 19, 442-449.
- [285] McCoy, A.; Grosse-Kunstleve, R.; Adams, P.; Winn, M.; Storoni, L.; Read, R. PHASER crystallographic software. *Journal of applied crystallography* **2007**, 40, 658-674.
- [286] The CCP4 suite: programs for protein crystallography. *Acta crystallographica. Section D, Biological crystallography* **1994**, 50, 760-3.
- [287] Brandt, T.; Holzmann, N.; Muley, L.; Khayat, M.; Wegscheid-Gerlach, C.; Baum, B.; Heine, A.; Hangauer, D.; Klebe, G. Congeneric but Still Distinct: How Closely Related Trypsin Ligands Exhibit Different Thermodynamic and Structural Properties. *Journal of molecular biology* **2011**, 405, 1170-1187.









## Danksagung

Mein erster Dank geht an meine Prüfungskommission. Prof. Dr. Torsten Steinmetzer danke ich für die Übernahme der Betreuung und den stetigen kreativen Freiraum den ich bei der Gestaltung meiner Projekte inklusive der Kooperationen hatte, aber trotzdem jederzeit auf Unterstützung bauen konnte. Besonders dankbar bin ich für die Geduld während des mühsamen Korrekturlesens und für den guten Kontakt zu der Arbeitsgruppe von Prof. Dr. Eva Friebertshäuser. Ich bin ich sehr dankbar, dass ich über einen langen Zeitraum Teil deiner AG sein durfte und du mir immer mit Rat und Ideen geholfen hast, wo du nur konntest. Prof. Dr. Cornelia Keck und Prof. Dr. Maike Petersen möchte ich herzlich für ihre Zeit danken meine Prüfungskommission zu vervollständigen. Ein gesonderter Dank geht an Prof. Dr. Maike Petersen für die Übernahme des Prüfungsvorsitzes. Dr. Stefan Merkl danke ich an dieser Stelle für die Bereitschaft als Protokollant an meiner Prüfung teilzunehmen, sowie Thuy für die Unterstützung bei der Organisation.

Prof. Dr. Viktor Magdolen danke ich für seine Unterstützung aus der Ferne und für das Bereitstellen des Matriptase Konstruktes und der Enterokinase sowie seinen eigenen Expressionsversuchen zur TMPRSS2. Prof. Dr. Gerhard Klebe und Prof. Dr. Klaus Reuther danke ich für die Erlaubnis zur Nutzung des S1 Bereiches und damit dem Zugang zu der AG Klebe. Durch die Expertise und Ausstattung der drei Arbeitsgruppen konnte ich meine Arbeit methodisch sehr breit gestalten und zudem viele verschiedene wunderbare Menschen aus unterschiedlichen Flecken der Erde kennen lernen. Besonders dankbar bin ich für meine kleine italienische principessa Nicole, ohne die ich die maximal harten Phasen nicht überstanden hätte. Ich freue mich, wenn unser Langzeitplan aufgeht und uns kein Teich mehr trennt. Wahnsinnig dankbar bin ich für die Friebertshäuser-Mädels Anne, Aybike, Doro, Hannah, Konni und Ruth. Durch euren Zusammenhalt und euer Engagement seid ihr das stärkste Team in dem ich bisher arbeiten durfte. Danke, dass ihr mich adoptiert habt und für den Support bei enttäuschenden Western Blot Freitagen, Pommes, Bier oder an der Trash-TV Front. Ich freue mich auf viele weitere heitere Treffen im Sudhaus, Köln, Basel oder auch wo immer es die Möglichkeit gibt. Leider erst viel zu spät kam ich in den Genuss der Zusammenarbeit mit Oliver und den Inhibitoren. Du warst einer der angenehmsten und geduldigsten Menschen am Institut und auch wenn ich deine Leidenschaft zur Chemie nicht teilen kann, danke ich dir sehr, dass du die Hoffnung nicht aufgegeben hast und mir alles, wenn nötig, auch mehrfach erklärst. Danke für die gute Zusammenarbeit, das Korrekturlesen und den rheinischen Humor. Komplettiert wurde unser kleines drug discovery Team durch Florents Computergeflüster und durch Stefan und Janis den Kristallexperten. Ich bin euch für die Zeit, das Engagement und die Leidenschaft mit

der ihr bei der Sache wart sehr dankbar. Besonders Zeit ist etwas sehr wertvolles was ich sehr zu schätzen weiß. Ohne euren Support wäre diese Arbeit so für mich nicht möglich gewesen. Allen Anderen aus der AG Steinmetzer und AG Klebe/Kolb, die mir während der letzten Jahre positive Momente beschert haben möchte ich dafür ebenfalls danken.

Natürlich gibt es auch Menschen außerhalb der Pharmazie/Virologie-PhD Bubble denen ich für den Support in den letzten Jahren dankbar bin. Dazu gehört die chillige Trierer Urbesetzung der Ketzterbach-WG mit den vielen wunderbaren Fußballtagen, Filmabenden und diversen cornereien. Niköln danke ich besonders für die zeitweise hochfrequentierte Bereitstellung eines Schlafplatz in Köln, wenn ich mal wieder dem Marburger Kleinstadt Driss entfliehen musste. Ich freue mich jetzt auf die nächsten Monate, in denen wir nach vielen Jahren an getrennten Orten zusammen in unserer Herzensstadt wohnen. Das wird grandios! Meiner treuesten Lebensbegleiterin Mari bin für die stetige mentale Unterstützung und die vielen ermutigenden Worte aber auch für die vielen gemeinsamen Erlebnisse und Erinnerungen der vergangenen 13 Jahre dankbar. Egal was die Zukunft bringt ich bin ich mir sicher, dass du ein Teil davon sein wirst. Meiner Familie danke ich für ihre Bodenständigkeit, Ehrlichkeit und bedingungslose Akzeptanz. Wir sind gewiss keine Menschen der großen emotionalen Worte und ich bin froh, dass wir sie nicht brauchen um uns des gegenseitigen Rückhalts sicher zu sein. Ich bin dankbar diesen einen safety place zu haben. Als letztes möchte ich (dem privaten) Florent inklusive seinem Humor und positiven Art für die Unterstützung in den letzten Jahren danken. Du hast kein Aufgeben zugelassen und mich durch deine eigene Zielstrebigkeit und Disziplin motiviert das Ding durchzuziehen. Ich bin sehr gespannt, wie ein Leben mit „nur“ Arbeiten sein wird, und egal welche Stadt, welches Land oder welche Challenge als nächstes kommt...

**...et hätt noch immer jot jejange.**

(§3 Kölsches Grundgestz)

## Publikationen

### JOURNALS

---

Badran, M. J.; Bertolotti, N.; Keils, A.; Heine, A.; Klebe, G.; Marchais-Oberwinkler, S. Mutational and structural studies uncover crucial amino acids determining activity and stability of 17 $\beta$ -HSD14. *The Journal of Steroid Biochemistry and Molecular Biology* **2019**, 189, 135-144.

Häußler, D.; Schulz-Fincke, A.-C.; Beckmann, A.-M.; Keils, A.; Gilberg, E.; Mangold, M.; Bajorath, J.; Stirnberg, M.; Steinmetzer, T.; Gütschow, M. A Fluorescent-Labeled Phosphono Bisbenzguanidine As an Activity-Based Probe for Matriptase. *Chemistry – A European Journal* **2017**, 23, 5205-5209.

### POSTERS

---

Keils A., Magdolen V., Böttcher-Friebertshäuser E., Steinmetzer T. Characterization of hemagglutinin cleaving proteases as potential targets for influenza treatment. 28th Annual Meeting of the Society for Virology **2018**

Keils A., Böttcher-Friebertshäuser E., Magdolen V., Steinmetzer T. Investigation of the hemagglutinin cleaving type-II transmembrane proteases TMPRSS2 and matriptase as potential targets for influenza treatment. DPhG Annual Meeting **2016**



**Lebenslauf**





## Eidesstattliche Erklärung

Ich versichere, dass ich meine Dissertation

### **„Characterization of the Hemagglutinin Cleaving Transmembrane Serine Proteases Matriptase and TMPRSS2”**

selbständig ohne unerlaubte Hilfe angefertigt und mich dabei keiner anderen als der von mir ausdrücklich bezeichneten Quellen bedient habe. Alle vollständig oder sinngemäß übernommenen Zitate sind als solche gekennzeichnet.

Die Dissertation wurde in der jetzigen oder einer ähnlichen Form noch bei keiner anderen Hochschule eingereicht und hat noch keinen sonstigen Prüfungszwecken gedient.

Köln, den.....

.....  
(Unterschrift mit Vor- und Zuname)