University of Applied Sciences Hamburg
Faculty of Life Sciences
Department Biotechnology

Optimization of *Andes virus* cap-snatching endonuclease heterologous expression and purification

Bachelor Thesis
for the degree
B. Sc. in Biotechnology

by

Bianca Di Fabrizio
[redacted]
Matriculation-No. [redacted]

Hamburg
February 01st, 2021

1st Reviewer: Prof. Dr. Gesine Cornelissen (HAW Hamburg)

2nd Reviewer: Dr. Yaiza Fernández-García (BNITM)

Acknowledgements

First of all, special thanks to Dr. Yaiza Fernández-García for giving me the option to write a Bachelor Thesis under her supervision. With her, I had a great supervisor, teacher and motivator. I've always been able to consulate her with any questions or concerns and really appreciate the opportunity to develop myself.

Also, thanks to Prof. Dr. Gesine Cornelissen from the HAW Hamburg for the support and inspiration in the field of protein purification.

Finally, my thanks also belong to the Bernhard-Nocht-Institute for Tropical Medicine for me to be able to carry out the experiments in their laboratories and giving insights in good scientific practices.

In general, thank you for the great and interesting experiences during this period.

Eidesstattliche Erklärung

Ich versichere an Eides statt, dass die Bachelorarbeit mit dem Titel „Optimization of *Andes virus* cap-snatching endonuclease heterologous expression and purification“, ohne fremde Hilfe selbstständig verfasst und nur die im Literaturverzeichnis angegebenen Quellen und Hilfsmittel verwendet wurden. Ferner sind die verwendeten Tabellen, Grafiken, Bilder und andere Materialien richtig zitiert, oder selbstständig erstellt worden. Wörtliche oder dem Sinn nach aus anderen Werken entnommene Stellen sind unter Angabe der Quelle gemäß wissenschaftlichen Zitierregeln kenntlich gemacht. Die Arbeit wurde bisher in gleicher oder ähnlicher Form in keiner Prüfungsbehörde vorgelegt und auch nicht veröffentlicht.

Statutory Declaration

I hereby formally declare, that the Bachelor thesis with the title " Optimization of *Andes virus* cap-snatching endonuclease heterologous expression and purification" was written independently by myself. I did not use any outside support except for the quoted literature and other sources mentioned in the thesis. Passages and ideas originating from external sources are marked clearly, either literally or analogously. The thesis has not been submitted to any examining authority in the same or similar form and has not been published.

Hamburg, 01.02.2020 [redacted]

(Ort, Datum) (Unterschrift)

Abstract

The *Andes virus* (ANDV) is endemic in South America and causes hantavirus cardiopulmonary syndrome (HCPS) with a case fatality rate of up to 40 %. The N-terminal Cap-ENDO domain of the L-protein of ANDV avails a cap-snatching mechanism that is necessary for viral survival. As such, this domain can be used as a target to elucidate potential medical treatments.

The wild type ANDV Cap-ENDO shows toxic activity in expression hosts, making it difficult to perform small molecules screening campaigns. This work aims to produce a target that could be exploited for drug discovery. The attenuated mutant ($ANDV_{L1\text{-}200}$ N167A) was used for expression in Lemo21(DE3) *Eschericia coli (E. coli)* cells. To purify the $ANDV_{L1\text{-}200}$ N167A protein, various methods, such as affinity chromatography, tag-removal via digestion reaction and ion exchange chromatography were performed. Assessment of the purification process was carried out by 12 % SDS-PAGE. Improvement in $ANDV_{L1\text{-}200}$ N167A protein expression levels was achieved by expressing the protein with a N-terminal His-GST-3C tag. Off-column cleavage showed best results to cleave the fusion tag from the protein. To prevent the $ANDV_{L1\text{-}200}$ N167A protein from precipitation during cleavage at pH ranging its isoelectric point, the 3C cleavage site was successfully exchanged by a thrombin cleavage site via a two-step PCR mutagenesis followed by In-Fusion cloning. It is proposed that thrombin can work at acidic pH and this could increase the yield recovery of $ANDV_{L1\text{-}200}$ N167A after cleavage, thus improving the target´s production. In summary, this work has evaluated different steps in the purification process of a promising viral target and generated a new construct that may increase $ANDV_{L1\text{-}200}$ N167A yields.

Table of Contents

List of Figures

List of Tables

Abbreviations

A	Alanine
AA	Amino acid
AC	Affinity chromatography
Amp^R	Ampicillin resistance gene
ANDV	*Andes virus*
APS	Ammoniumperoxodisulfate
Bp	Base pair
BNITM	Bernhard-Nocht-Institute for Tropical Medicine
BSL	Biosafety level
Cam^R	Chloramphenicol resistance gene
cDNA	Complementary deoxyribonucleic acid
CEXC	Cation exchange chromatography
CHO	Chinese Hamster Ovary
Dam	DNA adenosine methylase
Dcm	DNA cytosine methylase
DTT	Dithiothreitol
E. coli	*Escherichia coli*
ENDO	Endonuclease
F	Forward
FT	Flow-through
GPC	Glycoprotein precursor
GST	Glutathione-S transferase
HCPS	Hantavirus cardiopulmonary syndrome
HRV	Human rhinovirus
IAV	Influenza A virus
IEXC	Ion exchange chromatography
IPTG	Isopropyl-β-D-thiogalactopyranoside
L	Large
LB	Lysogeny broth
M	Medium
MBP	Maltose binding protein
mRNA	messenger RNA
MWCO	Molecular weight cut-off

N	Asparagine
NP	Nucleoprotein
NSs	Non-structural protein
ON	Over night
PCR	Polymerase chain reaction
pI	Isoelectric point
PMSF	Phenylmethanesulfonylfluoride
POI	Protein of interest
R	Reverse
RdRp	RNA-dependent RNA-polymerase
RNase	Ribonuclease
rpm	Revolutions per minute
RT	Room temperature
S	Small
SDS-PAGE	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis
SEC	Size exclusion chromatography
SOC	Super optimal broth with catabolite repression
ssRNA	Segmented single-stranded RNA
Str^R	Streptomycine resistance gene
T	Thrombin
TB	Terrific broth
TEMED	Tetramethylethylendiamin
Tet^R	Tetracycline resistance gene
vRNA	Viral ribonucleic acid

1 Introduction

1.1 *Andes virus*

Andes orthohantavirus, also known as *Andes virus* (ANDV), is a potentially lethal human pathogen endemic in South America [1–3]. Its natural reservoir are rodents of the genus *Oligoryzomys* distributed primarily in Argentina and Chile [4 5]. Although ANDV (family: *Hantaviridae,* order: *Bunyavirales*) [6] is mainly spread via the aerosolized excreta (saliva, urine and feces) [7] of infected animals, it is the only member of the taxonomic family for which person-to-person transmission has been described [1 3 8–11].

In comparison to the nearly asymptomatic infection of its host, in humans, ANDV can cause hantavirus cardiopulmonary syndrome (HCPS) [1 9 12]. After 1 to 5 weeks of exposure, patients show flu-similar symptoms like headache, fever, muscle aches, diarrhea, nausea or vomiting. Severe symptoms appear typically 4 to 10 days after disease onset; these are characterized by hypotension, pulmonary edema and cardiac shock. As with other viral hemorrhagic fevers, a hallmark of HCPS is an increased capillary permeability that can provoke the extravasation of fluids and result in interstitial edemas [2 13].

There are no approved vaccines or therapeutic agents against ANDV infections, which unfortunately results in case fatality rates of up to 40 % [8 9]. Due to the high pathogenicity, ANDV ability to spread via aerosols and the lack of medical countermeasures, require handling of this virus in biosafety level (BSL) 3 and 4 facilities [9].

1.2 Molecular biology of hantaviruses

Hantaviruses are enveloped viruses that range between 120 - 160 nm in diameter [2] and contain a segmented single-stranded RNA genome with negative polarity ((-)ssRNA). The viral genome´s segments are denominated small (S), medium (M) and large (L) according to their size [7 8 12].

The S segment encodes for a nucleoprotein (NP), which encapsulates the viral genome segments, and a non-structural protein (NSs) involved in the host´s innate immune evasion [14]. The glycoproteins G_N and G_C, derived from a glycoprotein precursor (GPC), are encoded by the M segment. These glycoproteins are embedded in the lipid bilayer and built up the outer part of the virions as spikes [2]. ANDV virions consist of more than 50 % protein and 20 – 30 % lipids and are very stable regarding temperature changes, as they can survive more than 10 days at room temperature [2]. Meanwhile, the L-segment encodes the L-protein, which contains the RNA-dependent RNA-

polymerase (RdRp) domain and is in charge of transcription and replication of the viral genome [2 12].

Viral replication takes place in the host-cell cytoplasm (Figure. 1). Primarily, the virus targets endothelial cells and enters via β3 integrin receptors [8] that are recognize by the G_N and G_C spikes [7]. After binding to the cell surface receptor, hantaviruses can enter via clathrin-dependent endocytosis. That means that the virus is covered with a cellular membrane consisting of clathrin-protein and the clathrin-coated pit containing virions are invaginated into the cells. However, ANDV does not use clathrin-mediated endocytosis and other entry pathways have been suggested [2 7]. Hantaviruses spikes require low pH for the needed conformational changes that induce the fusion of the virion envelops to the endosomes releasing the vRNA into the cytoplasm. Fusion is reported to occur in the late endosomes (pH 5 - 6) and further transportation of the vRNA to replication sites is yet unknown.

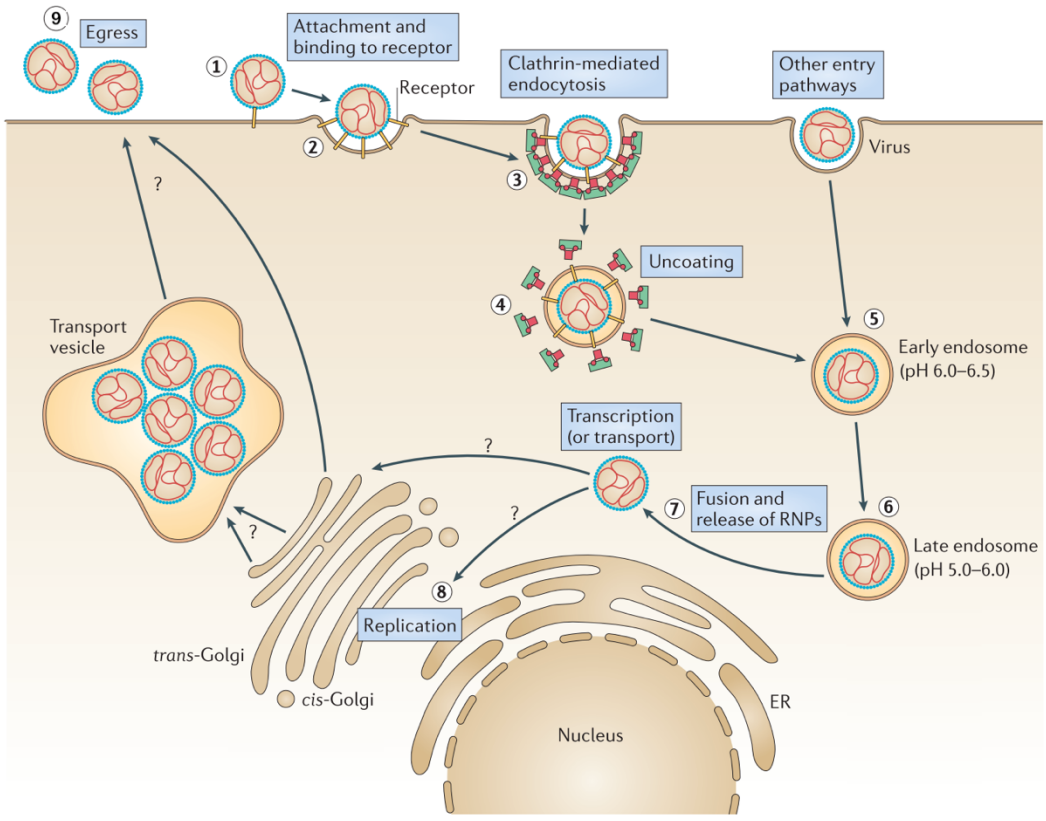

Figure 1. Schematic representation of the life cycle of hantaviruses. Modified from Vaheri et al., 2013.

It is hypothesized that viral mRNA transcription and translation precede the initiation of replication. Transcription and translation of viral mRNA would provide sufficient NP to mediate the switch from mRNA synthesis towards replication [7]. Transcription of the viral RNA occurs via a cap-snatching mechanism. The L-protein (250 kDa) possesses assisting enzymatic functions,

aside from the RdRp activity, that are necessary for transcription and replication of the vRNA genome. The endonuclease (Cap-ENDO), located in the N-terminal amino acid sequence of the L protein, is required for the virus mRNA synthesis [4 10]. This enzyme is responsible for the cleavage of short capped oligonucleotides from the cellular host mRNA that are later used as primers by the RdRp domain. Although the canonical cap-snatching mechanism has been described for influenza A virus (IAV) [15 16], it is believed that there are significant differences between IAV and hantaviruses mechanism for the initiation of transcription [4]. In comparison to IAV transcription where the nuclear process takes place without the regulatory action of additional viral proteins, hantaviruses NP has been reported to be a controlling element for a process that takes place in a cellular cytoplasm context [15].

1.3 Production of recombinant proteins

A variety of hosts have been used for the expression of heterologous proteins. The most common are bacteria, yeast, insect and mammalian cells, where each comes with its advantages and disadvantages. Starting with prokaryotes, *E. coli* is one of the most common expression hosts among bacteria. *E. coli* a well understood organism, cheap, shows high transformation efficiency, is easy to cultivate and fast-growing, which makes it attractive for large scale expression [17]. Thanks to progressive modifications of bacteria, the expression of toxic proteins can be enabled using a regulated T7 RNA bacteriophage system (Lemo system) [18 19]. Yet, bacteria lack some of the advantages of eukaryotic cells [20].

Eukaryotic cells are often used to express recombinant proteins that possess key posttranslational modifications and complex protein folding [21 22]. Yeasts, for example *Pichia pastoris* or *Saccharomyces cerevisiae* are mainly used for the production of biopharmaceuticals and industrial enzymes in large scale applications in high titers [23]. Yeasts allow expression of proteins either intra- or extracellularly. Despite the benefits, proteins expressed in yeast cause problems with hyperglycosylation and protein folding [24]. Insect cells (*Sf*9 [25]) are mostly used as a virus-based expression system (e. g. recombinant baculoviruses) [26]. The less complex system allows easy manipulation resulting in higher expression levels compared to mammalian cells [17], but vary in different lipids and are more cost-intensive compared to yeast or *E. coli*. As it has been reported that in some instances the proteins efficacy is closely linked to their processing host [27], mammalian cells and particularly the Chinese Hamster Ovary cell line (CHO) have become dominant in the production of recombinant proteins and antibodies for clinical applications [20 27]. However, this expression system is the most expensive, has low expression yields and a high cultivation complexity [20].

1.4 Scientific problem and Experimental approach

Up to this day, recombinant production of the soluble wild-type Cap-ENDO of ANDV is challenging due to the protein´s non-selective nuclease activity which is too toxic for high-level expression in both procaryotic and eukaryotic cells [4 10]. To bypass this problem, several attenuated mutants have been used for expression [4]. One of those mutants, $ANDV_{L1\text{-}200}$ N167A in which the amino acid (AA) asparagine (N) at position 167 is exchanged for Alanine (A), can be eligible for compound screening campaigns as it retains an intact active site and the mutated residue is found in great distance from the buried catalytic pocket. Thus, the mutated heterologous protein is very active and, in host cells, toxic in high concentrations. This is reflected in slower growth of *E. coli* BL21 (Gold) expressing $ANDV_{L1\text{-}200}$ N167A [4].

The aim of this thesis was to optimize the expression and purification of the Cap-ENDO domain of ANDV. Due to the toxic activity of $ANDV_{L1\text{-}200}$ N167A in expression hosts, its recombinant production was carried out in Lemo21(DE3) *E. coli*. This bacteria strain is suitable for the production of difficult proteins that benefit from a tight induction control of expression. The protein of interest (POI) is located in the *lac*-operon and expressed by the T7-polymerase after induction. Basal expression of the POI is suppressed by controlling the polymerase with T7-lysozyme. Moreover, different fusion tags were used to assess the expression levels of soluble $ANDV_{L1\text{-}200}$ N167A and to facilitate the purification of the respective recombinant proteins. To separate our POI from the tags employed, a HRV 3C cleavage site between the tag and the POI was exploited. Protease cleavage of the fusion proteins were performed under various buffers conditions and setups (ON- or Off-column). Yet, due to the unfavorable POI recovery after the HRV 3C protease reaction at pH´s too close to the isoelectric point (pI) of either the protease or POI, a new strategy for cleavage using thrombin (T) was proposed. In contrast to HRV 3C, thrombin is active in acidic buffers (pH 5 - 7) in which $ANDV_{L1\text{-}200}$ N167A is stable. Therefore, constructs were created exchanging the proteases recognition sites for further use.

2 Materials and Methods

2.1 Materials

2.1.1 Bacteria strains, plasmids, media and supplements

The following tables contain the bacteria, plasmids, media and chemicals used in this thesis.

Table 2. 1. *E. coli* bacteria strains

Name	Genotype	Manufacturer	Ref.-No.
NEB® 5-alpha	*fhuA2 Δ(argF-lacZ)U169 phoA glnV44 Φ80 Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17*	New England Biolabs	C2987
Lemo21(DE3)	*fhuA2 [lon] ompT gal (λ DE3) [dcm] ΔhsdS/ pLemo*(**Cam^R**) *λ DE3 = λ sBamHIo ΔEcoRI-B int::(lacI::PlacUV5::T7 gene1) i21 Δnin5* pLemo = pACYC184-*PrhaBAD-lysY*	New England Biolabs	C2528
dam$^-$/ *dcm*$^-$	*ara-14 leuB6 fhuA31 lacY1 tsx78 glnV44 galK2 galT22 mcrA dcm-6 hisG4 rfbD1 R(zgb210::Tn10)* **Tet^S** *endA1 rspL136* (**Str^R**) *dam13::Tn9* (**Cam^R**) *xylA-5 mtl-1 thi-1 mcrB1 hsdR2*	New England Biolabs	C2925

Table 2. 2. Plasmids

Name	Length (bp)	Manufacturer	Ref.-No.
pOPIN_M_ANDV$_{L1\text{-}200}$ N167A	6936	BNITM - in-house production	Backbone: pOPIN_M, Oxford Protein Production Facility, Addgene 26044
pOPIN_J_ANDV$_{L1\text{-}200}$ N167A	6489	BNITM - in-house production	Backbone: pOPIN_M, Oxford Protein Production Facility, Addgene 26045
pOPIN_J_T_ANDV$_{L1\text{-}200}$ N167A	6483	BNITM - in-house production	

Table 2. 3 LB- and TB-Media and supplements

Name	Manufacturer	Ref.-No.
Yeast-Extract	Roth	2904.3
Peptone	Roth	8986.2
NaCl	Roth	3957.2
Glycerol Rotipuran	Roth	3783.1
KH_2PO_4	Roth	3904.1
K_2HPO_4	Roth	P749.2
L-rhamnose	Sigma Aldrich	83650
Isopropyl-β-D-thiogalactopyranoside	Roth	CN08.3
Chloramphenicol	Roth	3886.2
SOC Outgrowth Medium	New England Biolabs	B9020
LB-Agar n. Lennox	Roth	X965.1
Ethanol	Roth	9065.4
Carbenicillin	Roth	6344.3

2.1.2 In-Fusion Cloning

The following tables contain enzymes, buffers and supplements used in this thesis.

Table 2. 4. Enzymes, buffers and supplements

Name	Manufacturer	Ref.-No.
SexAI (A↓CCWGGT)	New England Biolabs	R0605
BclI-HF (T↓GATCA)	New England Biolabs	R3160
HRV-3C Protease (LEVLFQ↓GP)	in-house recombinant production of GST-tagged cysteine protease from human rhinovirus	
Lysozyme	PanReac AppliChem	9001-63-2
Q5 High Fidelity DNA Polymerase	New England Biolabs	M0491
Q5® Reaction Buffer	New England Biolabs	B9027
dNTP Set 100 mM Solutions	Thermo Fisher Scientific	10297-018
CutSmart® Buffer	New England Biolabs	B7204
NEBuilder® HiFi DNA Assembly Master Mix	New England Biolabs	E2621

Table 2. 5. Primers

Name	Length (bp)	Sequence 5' → 3'
J-Thrombin-ANDV 1F	33	CATATTTAAGTGATCATGTAACCCATCCTG
J-Thrombin-ANDV 1R	35	ACTGCCGCGTGGCACCAGACCGCTGCTCA GATCCG
J-Thrombin-ANDV 2F	55	CTGGTGCCACGCGGCAGTATGGAAAAGTA TAGAGAGATTCATCAGAGAGTTAGGG
J-Thrombin-ANDV 2R	27	CATCTGGTCAACCAGGTCATGTCTGAC

Table 2. 6. Agarose-Gel

Name	Manufacturer	Ref.-No.
Biozym LE Agarose	Biozym	840004
Gel Loading Dye, Purple (6X)	New England Biolabs	B7024
Quick load 100 bp DNA ladder	New England Biolabs	N0467
Quick load 1 kbp DNA ladder	new England Biolabs	N0468
Ethidium bromide 0.5 %	Roth	HP46.1

2.1.3 Kits

The following table provides the kits used in this thesis.

Table 2. 7. Kits

Name	Manufacturer	Ref.-No.
NucleoSpin Plasmid, Mini kit for plasmid DNA	Macherey Nagel	740588250
Macherey-Nagel™ NucleoSpin™ Gel and PCR Clean-up Kit	Macherey Nagel	11992242

2.1.4 Protein Purification

The following tables contain the material and chemicals used for purification in this thesis.

Table 2. 8. Reagents and columns for chromatographic protein separation

Name	Manufacturer	Ref.-No.
Disposable Gravity-flow Columns	QIAGEN	1018597
Amylose-resin	New England Biolabs	E8021
Glutathione Sepharose® 4B	Sigma Aldrich	GE17-0756-01
IEXC: HiTrap SP XL 1 mL	GE Healthcare	17-5160-01

Table 2. 9. Reagents for buffers and solutions

Name	Manufacturer	Ref.-No.
EDTA Disodium salt dihydrate	Roth	8043.1
NaCl	Roth	3957.2
Tris	Roth	2449.3
Na_2HPO_4	Roth	T876.2
NaH_2PO_4 x 2 H_2O	Merck	1063421000
MES	Roth	4259.3
reduced Glutathione	Roth	6382.1
Phenylmethanesulfonylfluoride	Roth	6367.1
Dithiothreitol	Roth	6908.2
Isopropanol	Roth	9866.5
$MnSO_4$	Merck	5963
Glycerol Rotipuran	Roth	3783.1
SDS, ultra-pure	Roth	2326.2
Glycine	Roth	3908.2
Bromophenol Blue Sodium salt	Roth	A512.1
β-Mercaptoethanol	Roth	4227.1
Brilliant blue G-250 C.I. 42655	Roth	R9598.1
Hydrochloric acid	Roth	4625.1
TEMED	Roth	2367.3
ROTIPHORESE® NF-Acrylamid/ Bis-Lösung 30 (29:1)	Roth	A124.2
Ammonium peroxydisulphate (APS)	Roth	9592.2
2-Propanol	Roth	6752.4
PAGE Ruler	Thermo Fisher Scientific	26616

2.2 Methods

2.2.1 Transformation of chemically competent bacteria

Competent cells were transformed, to amplify plasmids for sequencing or cloning reasons and for protein expression.

Aliquots were thawed on ice for maximum 10 min before 500 - 1000 ng of plasmid DNA was added. The mixtures were incubated for 30 min on ice followed by a heat-shock for 10 - 45 s at 42 °C in a heat-block and placed back on ice for 5 min. Pre-warmed SOC-Media (37 °C) was added to the cells to a final volume of 1 mL for recovery and was left shaking for 1 h at 37 °C. Then, 100 μL of the suspension as well as 1 x 10-fold dilution was plated onto pre-warmed LB agar-plates containing the selective antibiotics in accordance to the combination of transformed bacteria strain and plasmid. The selective LB-agar plates were incubated overnight (ON) at 37 °C. Glycerol stocks of all bacteria strain-constructs combinations were prepared from 5 ml LB media supplemented with corresponding antibiotics and stored at -80 °C as 1 mL aliquots.

2.2.2 Protein expression

The $ANDV_{L1\text{-}200}$ N167A protein was expressed in Lemo21(DE3) *E. coli* with plasmids containing the encoding sequence. The POI was already cloned into plasmids pOPIN_M and pOPIN_J. Proteins that were expressed from pOPIN_M contained the N-terminal His-MBP-3C tag, whereas expression from pOPIN_J generated His-GST-3C-$ANDV_{L1\text{-}200}$ N167A. The used plasmids containing the antibiotic resistance and the POI located in the *lac*-operon are shown in Figure 2.

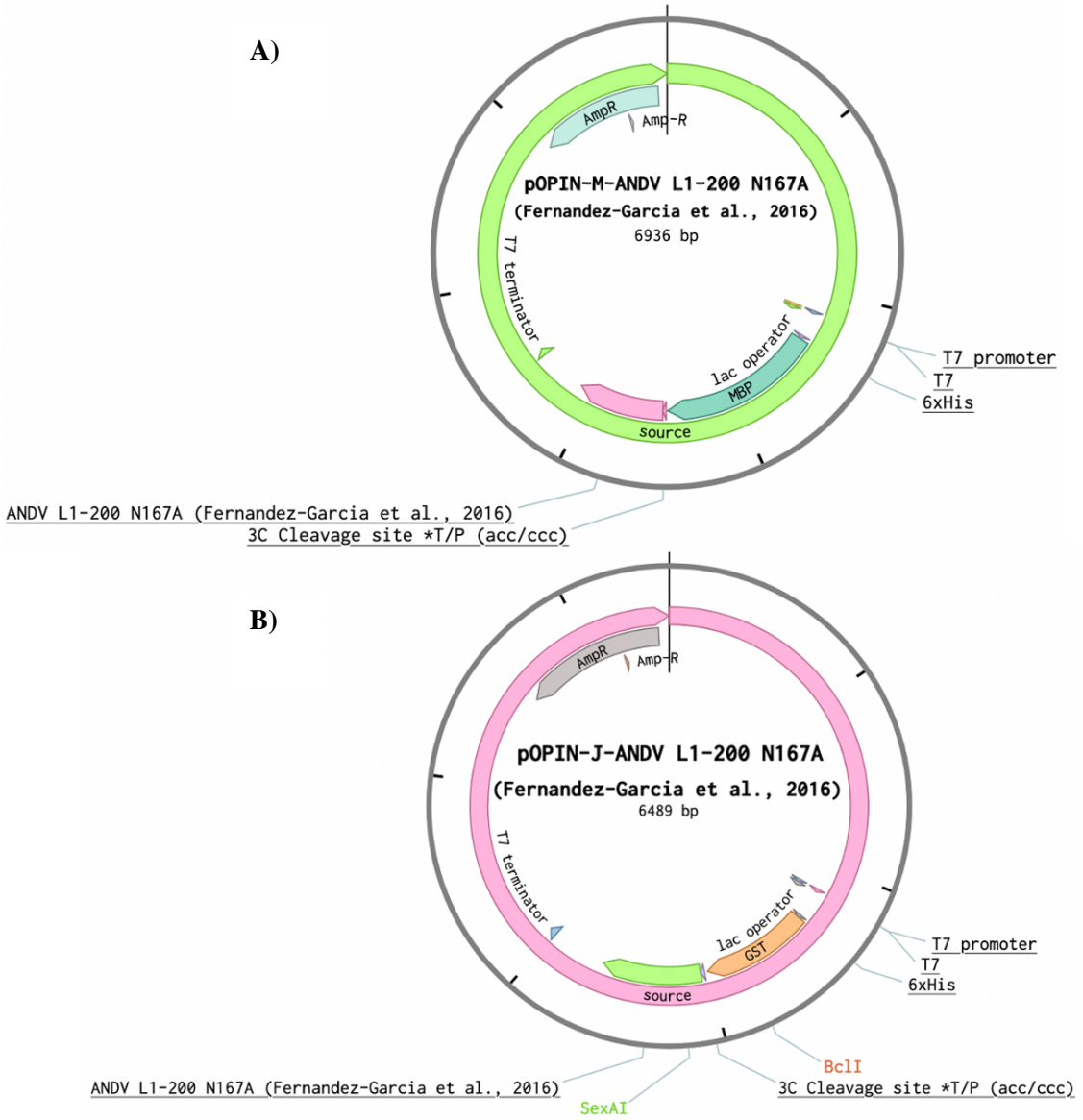

Figure 2. Schematic illustration of the used plasmids for protein expression. **A)** Plasmid pOPIN_M_ANDV$_{L1\text{-}200}$ N167A with 6xHis-MBP-3C tag (6936 bp) and **B)** shows plasmid pOPIN_J_ANDV$_{L1\text{-}200}$ N167A with 6xHis-GST-3C tag and restriction enzymes *Bcl*I and *Sex*AI. In both plasmids, the tag N-terminal tag and 3C cleavage site of the POI, lac-operon, T7 promoter and T7 terminator are shown.

Expression of fusion proteins was executed in *E. coli* strain Lemo21 (DE3). In these bacteria, the overexpression of proteins is regulated by T7 RNA Polymerase. Additionally, the activity of the T7 RNA Polymerase can be regulated by its natural inhibitor T7 lysozyme, which is modulated by the very well titratable rhamnose promoter (PrhaBAD).

A single colony 5 mL LB preculture was used to inoculate a 200 mL LB starter culture, in both cases the media was supplemented with the corresponding antibiotics. For large-scale expression,

10 mL of the starter culture was added to 900 mL TB media supplemented with 100 mL of 10X TB additives and corresponding antibiotics in 2.5 L baffling flasks. To sample the influence of L-rhamnose over the POI expression levels, different concentrations (0 μM, 250 μM, 500 μM and 1 mM) were added to the expression media as well. Incubation of the 1 L cultures at 37 °C, shaking at 180 rpm was carried out until OD_{600} reached 0.6 – 0.8. After OD_{600} was in the range, baffling flasks were pre-cooled in iced water and expression of $ANDV_{L1\text{-}200}$ N167A fusion proteins were induced with isopropyl-β-D-thiogalactopyranoside (IPTG) at a final concentration of 400 μM. After addition of IPTG, all 1 L cultures were incubated ON at 17 °C, shaking at 180 rpm. The next day, cultures were centrifuged in a pre-cooled Sorvall Evolution RC Superspeed Centrifuge for 15 min, 6000 x g, 4 °C in 250 mL bottles to recover the bacteria pellets. The weighted pellets were stored at -20 °C until purification started.

2.2.3 Protein purification

Variations to an overall purification process were based on the protein's properties. Buffers (Table I. III to Table I. V) were prepared at least one day pre purification and stored at 4 °C. Prior starting the purification, the pH of all buffers was checked and corrected if necessary. Whenever possible, steps were carried out on ice or at 4 °C to prevent the protein from degradation by proteases. All subsequent steps for purification are shown as flow-chart in Figure 3.

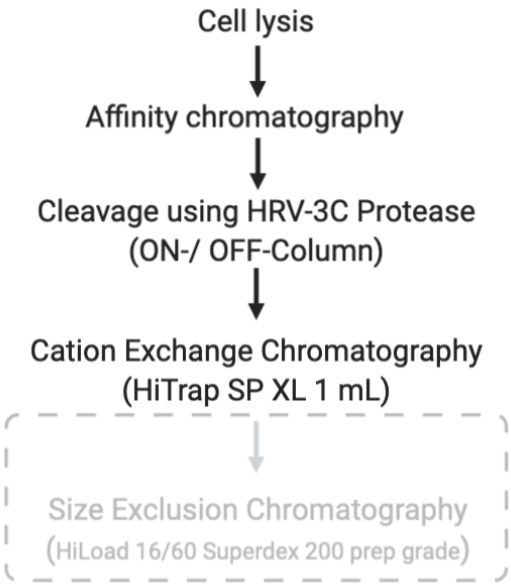

Figure 3. Overall purification process of $ANDV_{L1\text{-}200}$ N167A.

2.2.3.1 Cell lysis

The bacteria pellet was resuspended in lysis-buffer freshly supplemented with 1 mM Phenylmethanesulfonylfluoride (PMSF) and 0.1 mg/mL Lysozyme to a 1:5 ratio (w/v). On ice, cells were lysed via sonication with a Branson Ultrasonics™ Sonifier Modell 250 CE with a 13 mm disintegrator sonotrode. Three rounds of 7 min with a 40 % frequency and an amplitude setting of 4 – 5 were performed including a 10 min rest between each sonication step.

2.2.3.2 Affinity chromatography

After sonication, the lysed material was centrifuged in pre-cooled 40 mL tubes for 30 min at 15000 rpm in a pre-cooled Sorvall Evolution RC Superspeed Centrifuge at 4 °C. In the meantime, 1 mL of the corresponding 50 % (w/v) slurry affinity beads per 1 L of induced bacteria pellet was equilibrated according to the affinity tag by adding the resin onto a disposable gravity flow column and letting the ethanol flow through. Then, 10 mL of injection water was added to get rid of the remaining ethanol, followed by 5 min incubation with 10 mL of the supplemented lysis-buffer.

After centrifugation, the soluble fraction was transferred to a pre-cooled flask and incubated with the equilibrated beads for 30 to 60 min stirring at 4 °C. While one pellet from the insoluble fraction was resuspended in its original volume with lysis-buffer. After incubation of the soluble fraction with the beads, the suspension was passed through the gravity flow column using a peristaltic pump and the flow-through (FT) was stored at 4 °C. The beads were washed with 60 mL of cleavage buffer (see Table I. IV and Table I. V) supplemented with 1 mM Dithiothreitol (DTT). If the removal of the tag was performed Off-column, the fusion protein was eluted via 3 incubations of 5 min with 5 mL of freshly prepared elution buffer.

2.2.3.3 Removal of tags

Constructs available for the expression of MBP and GST fusion $ANDV_{L1\text{-}200}$ N167A proteins contained an HRV 3C protease recognition site at the N-terminal region of the POI. The optimal conditions for protease activity limited the pH range for the cleavage buffers. Therefore, to purify $ANDV_{L1\text{-}200}$ N167A, buffers with a pH ranging from 7 – 7.5 were chosen albeit being 0.5 – 1 units proximal to its pI.

2.2.3.3.1 On-column cleavage

For performing On-column cleavage, the beads containing the respective fusion $ANDV_{L1\text{-}200}$ N167A protein, were incubated in 5 mL of cleavage buffer (see Table I. IV) supplemented with 1 mM DTT and 1 mL of HRV 3C protease (1000 U) ON at 4 °C in rotation. The HRV 3C protease cleavage efficiency was assessed according to the amount of fusion protein still present in the beads after incubation.

If at least 75 % cleavage of the fusion protein was visible, the POI was recovered in a tube, by letting the suspension flow through the column. To recover any leftover protein, the beads were washed twice with 5 mL of HiTrap SP XL Buffer A (Table I. IV) supplemented with fresh 1 mM DTT. If less cleavage was visible, the reaction was incubated for another 24 h at 4 °C in rotation.

2.2.3.3.2 Off-column cleavage

Off-column or in solution cleavage, was carried out in a dialysis bag. The eluates from the affinity chromatography were pooled in a 50 mL falcon and the equivalent amount of HRV 3C cleavage buffer (Table I. V) supplemented with 1 mM DTT and 1 mL of HRV 3C protease (1000 U) was added. The reaction mix was transferred to a pre-wet dialysis bag (MWCO 6-8 kDa) and incubated ON at 4 °C, stirring against 1 L of HRV 3C cleavage buffer supplemented with 1 mM DTT. The efficiency of cleavage was assessed according to the amount of fusion protein still present after incubation.

If more than 75 % of the fusion protein was digested, the reaction mixture was transferred to a beaker and 125 μL of the equilibrated Glutathione Sepharose resin was added to the mixture to sequester the GST-HRV 3C protease. The suspension was incubated for 30 min, stirring at 4 °C. $ANDV_{L1\text{-}200}$ N167A was recovered in a beaker on ice, by passing the suspension through a gravity-flow column and any leftover protein was recovered by washing the beads three times with 5 mL of HiTrap SP XL Buffer A (Table I. V).

2.2.3.4 Ion exchange chromatography

The ion exchange chromatography step was used for increasing the purity of $ANDV_{L1\text{-}200}$ N167A by removing the remaining tag or fusion protein. Therefore, a strong cation exchanger (CEXC) column HiTrap SP XL 1 mL was connected to an ÄKTA pure 25 (GE Healthcare). The conductivity of the input sample was checked with a GLF 100 (GHM GREISINGER) conductivity meter and adjusted to a maximum of κ = 10 mS/cm using the HiTrap SP XL Buffer A supplemented with 1 mM DTT. The ÄKTA pure 25 program used (Table 2. 10) was run at a flow rate setting of f_R = 1 mL/min with a delta column pressure control of Δp_{column} = 0,3 MPa. When loading the sample onto the column, the flow-through containing unbound sample was collected in a beaker. Later elution of the protein was conducted with a NaCl gradient using 20 column volumes (CV) of the HiTrap SP XL Buffer B (Table I. V) supplemented with 5 mM $MnSO_4$.

Table 2. 10. ÄKTA pure 25 program used for CEXC using a HiTrap SP XL 1 mL column. Detection of proteins at λ = 280 nm, delta column pressure Δp_{column} = 0,3 MPa, f_R = 1 mL/min, CV = column volumes.

Step	CV	Fractionation
Equilibration: Water	5	/
Equilibration: HiTrap SP XL Buffer A	5	/
Sample application with airsensor on	-	Outlet valve
Column wash	5	Fraction collector
Elution (linear gradient)	20	Fraction collector
Equilibration: HiTrap SP XL Buffer B	5	/
Equilibration: Water	5	/
Equilibration: Ethanol	5	/

2.2.4 One-dimensional SDS-PAGE

To track the purification process, samples were taken from each of the relevant steps and run in 12 % Tris-Glycine SDS-PAGE where proteins in a range of 10 to 70 kDa are separated. The fusion-proteins vary in size according to their tag from 51 kDa (His-GST-3C) to 66 kDa (His-MBP-3C), while $ANDV_{L1\text{-}200}$ N167A has a molecular weight of 23 kDa.

The gels were hand-cast following instructions shown in Table 2. 11. Briefly, all ingredients except for TEMED, were mixed in a 50 mL falcon. To start the polymerization TEMED was added immediately before pouring. The gels were stored wet at 4 °C until further use.

Table 2. 11. Recipe for preparing 6 gels (8.6 cm x 6.8 cm) for 12 % Tris-Glycine SDS-PAGE.

Solutions	12 % Resolving Gel [mL]	6 % Stacking Gel [mL]
Agua dest.	10,50	10,60
Rotiphorese 30 % (29:1)	12,80	4,00
Tris (1,5 M, pH 8,8)	8,00	
Tris (0,5 M, pH 6,8)		5,00
SDS 10 %	0,32	0,20
APS 10 %	0,32	0,20
TEMED	0,032	0,02
Total Volume	32,00	20,00

The samples from the purification process were prepared by adding 4X SDS-PAGE loading buffer and incubating them for 1 min at 95 °C in a heat-block. After adding the samples to the mounted gel-chamber, the run was started at 100 V to allow all samples to enter the resolving gel at the same time at which point the run was set to 150 V until the front dye left the gel. Visualization of the immobilized proteins was carried out after washing the gel with boiling water to eliminate the SDS, by incubating it with Coomassie Safe Stain (see Table I. I) on an orbital shaker for 10 min and destaining the gel with water.

2.2.5 Modification of the pOPIN_J backbone

The HRV 3C protease recognition site (LEVLFQ↓GP) within the pOPIN_J-ANDV$_{L1\text{-}200}$ N167A vector was exchanged for the recognition site of thrombin (LVPR↓GS) via two-step PCR mutagenesis and subsequent In-Fusion cloning.

To generate the insert J_T-ANDV$_{L1\text{-}200}$ N167A, 2 sets of primers were designed. A set of primers (J-Thrombin ANDV 1F and 2R) comprised a minimum of 15 nucleotides complementary to the vector´s sequence after its linearization with single cutter restriction enzymes *Bcl*I and *Sex*AI. The second set of primers (J-Thrombin ANDV 2F and 1R) are degenerated, containing the coding sequence for the thrombin recognition site at the 5'-end aside from a minimum of 15 nucleotides complementary to the vector´s sequence (see Figure 4).

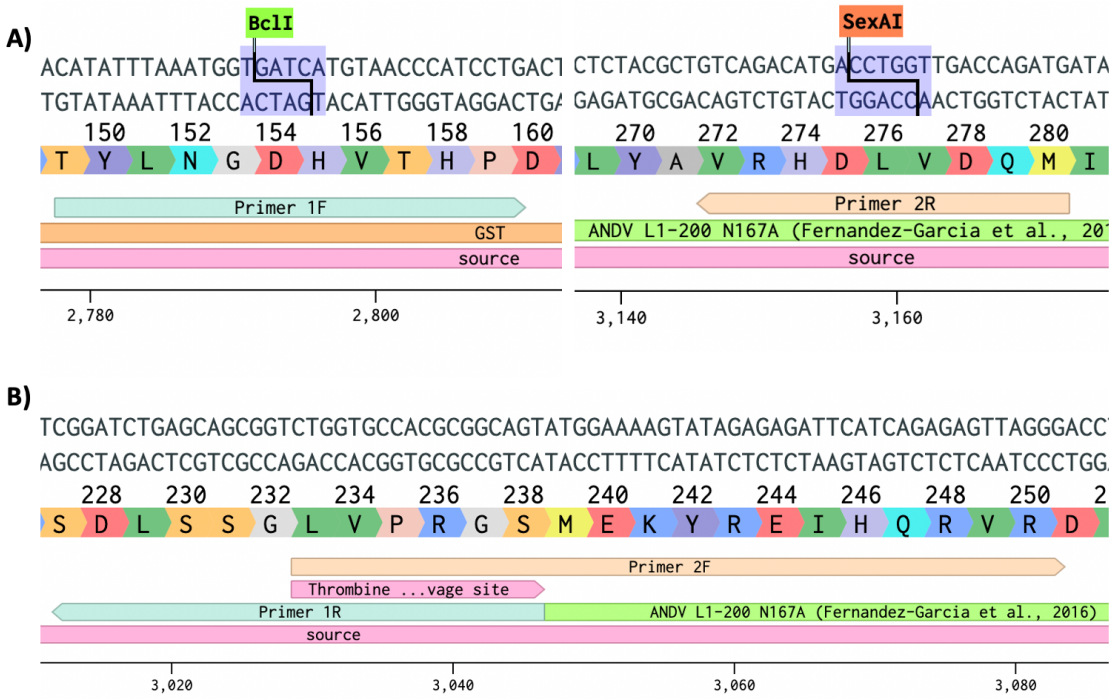

Figure 4. Sets of primers for insert generation of J_T-ANDV$_{L1-200}$ N167A, to exchange the 3C cleavage site for thrombin's. **A**) Primer 1F and 2R at restriction sites for BclI and SexAI, respectively. **B**) Primer 1R and 2F, containing the encoding sequence for the thrombin cleavage site.

The insert J_T-ANDV$_{L1-200}$ N167A was created by two-step PCR mutagenesis. The first step used pOPIN_J_ANDV$_{L1-200}$ N167A as template and included the separate amplification of two fragments with an overlapping region comprising the desired mutagenic site. The second step used the purified PCR products from the previous reactions as overlapping templates together with the primers for the 5´ and 3´-ends of the pOPIN_J-ANDV$_{L1-200}$ N167A vector for the new amplification reaction. The PCR mixes for the first and second steps of the PCR mutagenesis were prepared as described in Table 2. 12 A) and B) followed by the corresponding amplifications using the thermal cycling program in Table 2. 12 C). PCR products were purified from 2 % agarose gels utilizing the NucleoSpin Gel and PCR Clean-up Kit according to the manufacturer's instructions and DNA yields were determined using a NanoDrop™ 2000/2000c Spectrophotometer.

Table 2. 12. Tables A – C show the reaction setups used for the insert generation via two-step PCR mutagenesis. **A**) PCR reaction setup for generation of Amplicon 1 and Amplicon 2. **B**) PCR reaction setup for final insert J_T_ANDV$_{L1\text{-}200}$ N167A via PCR. **C**) Thermocycling conditions for the two-step PCR mutagenesis using a T100 Thermal Cycler (Bio-rad).

A)

Components	**Amplicon 1 [μL]**	**Amplicon 2 [μL]**
5X Q5 Buffer	5,00	5,00
10 mM dNTP´s (2,5 mM each)	0,50	0,50
10 µM J-Thrombin-ANDV 1F:	1,25	
10 µM J-Thrombin-ANDV 1R:	1,25	
10 µM J-Thrombin-ANDV 2F:		1,25
10 µM J-Thrombin-ANDV 2R:		1,25
10 ng/μL pOPIN_J_ANDVL1-200 N167A (Fernandez-Garcia et al., 2016)	1,00	1,00
H_2O	15,75	15,75
2 U/μL Q5 HF DNA pol	0,25	0,25
Total Volume	**50,00**	**50,00**

B)

Components	**Volume [μL]**
5X Q5 Buffer	10,00
10 mM dNTP´s (2,5 mM each)	1,00
10 µM J-Thrombin-ANDV 1F:	2,50
10 µM J-Thrombin-ANDV 2R:	2,50
10 ng/μL Amplicon 1	3,08
10 ng/μL Amplicon 2	1,39
H_2O	29,03
2 U/μL Q5 HF DNA pol	0,50
Total Volume	**50,00**

C)

Step	**cycles [n]**	**T [°C]**	**t [s]**
Initial Denaturation	1	98	30
Denaturation	35	98	30
Annealing		55	30
Elongation		72	15
Final Extension	1	72	120
Hold	**1**	**4**	∞

The resulting insert was cloned into *Bcl*I/*Sex*AI linearized pOPIN_J-ANDV$_{L1\text{-}200}$ N167A. Since *Bcl*I and *Sex*AI enzymatic activities are blocked by DNA dam and dcm methylations respectively, their substrate had to be produced in the methyltransferase deficient dam$^-$/dcm$^-$ *E. coli* strain. A 10-fold double digestion was set up according to Table 2. 13 during 1 h at 37 °C.

Table 2. 13. Reaction setup for the linearization of pOPIN_J_ANDV$_{L1\text{-}200}$ N167A (Fernández-García et al., 2016) with BclI HF and SexAI.

Components	**Volume [μL]**
10 X CutSmart Buffer	30,0
*Sex*AI (5 U/μL)	20,0
*Bcl*I HF (20 U/μL)	5,0
10 μg pOPIN_J_ANDV$_{L1\text{-}200}$ N167A	13,2
H_2O	231,8
Total Volume	**300,0**

In-Fusion cloning was performed using the NEBuilder HiFi DNA Assembly Master Mix with a 2:1 insert to plasmid ratio during 30 min at 50 °C. The whole of the In-Fusion cloning process is displayed in Figure 5. Half of the In-Fusion reaction was used to transform NEB® 5-alpha cells by incubations for 45 s at 42 °C and 2 min on ice followed with 1 h of recovery in SOC media at 37 °C. Plasmid DNA was purified via NucleoSpin Plasmid Kit from colonies grown under antibiotic selective pressure. The integrity of the construct was confirmed by sequencing.

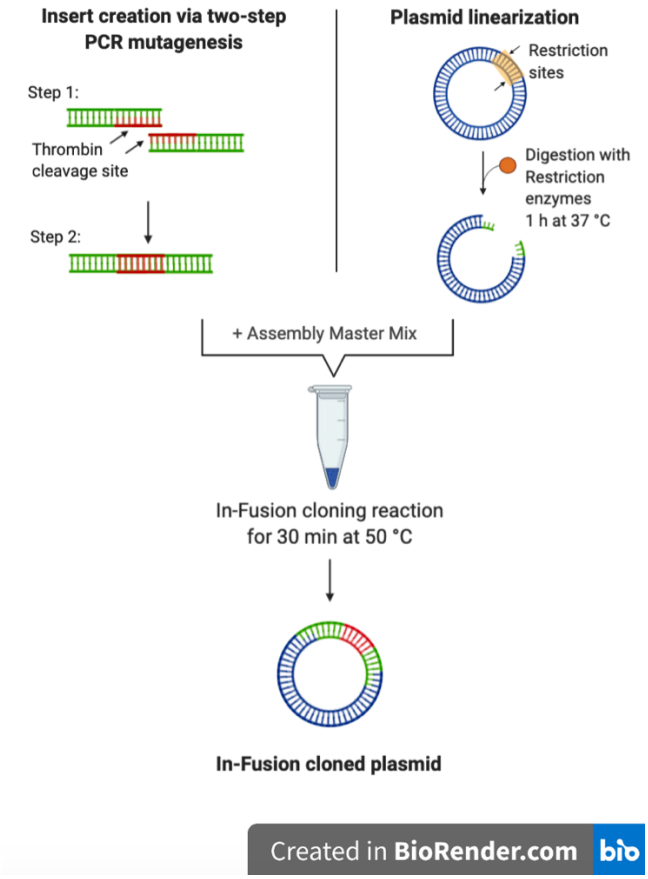

Figure 5. Schematic overview of the pOPIN_J_T-ANDV$_{L1\text{-}200}$ N167A generation. Insert creation via two-step PCR mutagenesis (in red is shown the coding sequence for the thrombin cleavage site, green shows the PCR generated insert), plasmid linearization and In-Fusion cloning. Image was created using BioRender.com.

3 Results

3.1 Protein expression

Due to the toxicity of $ANDV_{L1\text{-}200}$ N167A in high concentrations, expression of the POI in *E. coli* BL21 Gold (DE3) and mammalian cells was difficult so far [4 28].

The tight control of the T7 RNA polymerase by the Lemo system, in which the titratable L-rhamnose promoter (PrhaBAD) modulates T7-lysozyme levels in Lemo21(DE3) bacteria [18 29], allowed expression of the $ANDV_{L1\text{-}200}$ N167A domain as an MBP and GST fusion protein [18]. To express these fusion proteins, the plasmids pOPIN_M-$ANDV_{L1\text{-}200}$ N167A and pOPIN_J-$ANDV_{L1\text{-}200}$ N167A were used (Figure 2). pOPIN_M-$ANDV_{L1\text{-}200}$ N167A provides an N-terminal 6xHis-Maltose Binding Protein (MBP), whereas pOPIN_J-$ANDV_{L1\text{-}200}$ N167A contains a 6xHis-Glutathione S-Transferase (GST), in both cases a HRV 3C cleavage site is present between the fusion tag and the POI [25].

To determine the best conditions for large-scale expression, different L-rhamnose concentrations were tested. As shown in Figure 6, inductions of His-MBP-3C-$ANDV_{L1\text{-}200}$ N167A (66 kDa) showed low expression levels as bound to the Amylose beads, independently of the used L-rhamnose concentration. Meanwhile, Figure **7** shows high expression levels of His-GST-3C-$ANDV_{L1\text{-}200}$ N167A (51 kDa) looking at the fusion protein bound to the Glutathione Sepharose resin.

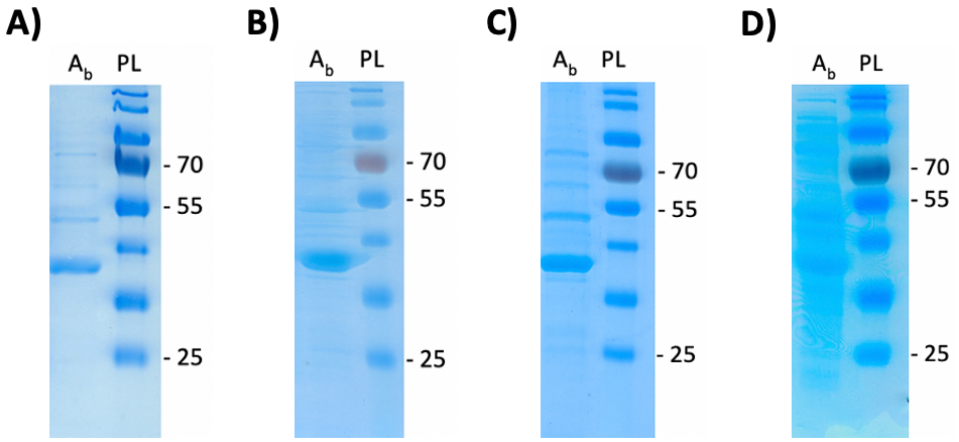

Figure 6. Assessment of expression levels of His-MBP-3C-$ANDV_{L1\text{-}200}$ N167A (66 kDa) expressed with **A)** 0 μM, **B)** 250 μM **C)** 500 μM and **D)** 1 mM L-rhamnose, respectively. Lane $\mathbf{A_b}$ presents proteins bound to the Amylose-resin and **PL** is the MW PAGE Ruler protein ladder (kDa).

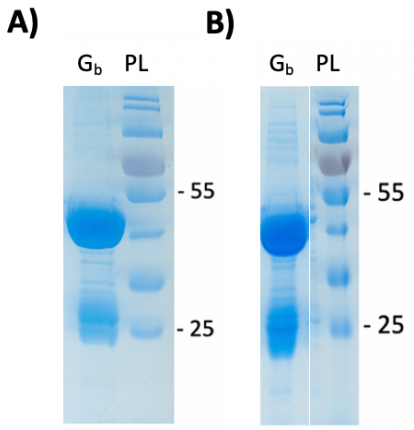

Figure 7. Evaluation of expressed His-GST-3C-ANDV$_{L1\text{-}200}$ N167A (51 kDa). Comparison of bound fusion protein to the Glutathione Sepharose resin (**G$_b$**), expressed with **A)** 0 μM and **B)** 250 μM L-rhamnose. **PL:** MW PAGE Ruler protein ladder (kDa).

Both MBP and GST are known to enhance the solubility and stability of fused proteins and can also protect the POI from proteolytic degradation [30 31]. In the case of ANDV$_{L1\text{-}200}$ N167A, a clear contrast between the effects of MBP and GST over the POI can be observed (Figure 6 and Figure **7**). While the GST fused ANDV$_{L1\text{-}200}$ N167A is well expressed, MBP fusion shows little protein being expressed.

To confirm the low expression levels of the MBP fused POI, Lemo21(DE3) *E. coli* was transformed again with pOPIN_M_ANDV$_{L1\text{-}200}$ N167A. Expression of the fusion protein (66 kDa) was performed only with 500 μM L-rhamnose, since it displayed the highest protein level among MBP fused to ANDV$_{L1\text{-}200}$ N167A (see Figure 8).

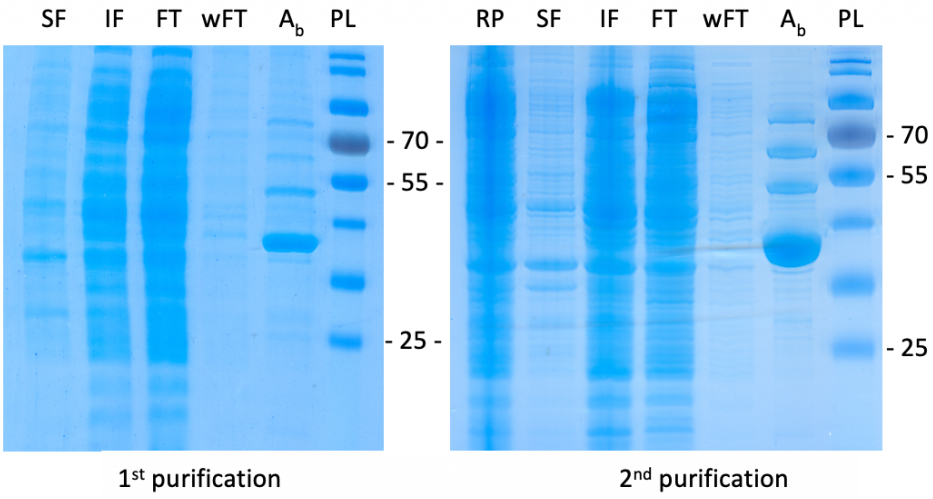

Figure 8. Assessment of the reproducibility of the expression levels of His-MBP-ANDV$_{L1\text{-}200}$ N167A. 12 % SDS-PAGE of His-MBP-3C-ANDV$_{L1\text{-}200}$ N167A (66 kDa) after AC of protein expressed in the presence of 500 μM L-rhamnose. **RP:** resuspended pellet, **IF:** insoluble fraction, **SF:** soluble fraction, **FT:** flow-through (FT), **wFT:** washed FT, **A$_b$:** Amylose-resin bound, **PL:** MW PAGE Ruler protein ladder (kDa)

3.2 Protein purification

Fusion tags can facilitate the purification of a POI while using affinity chromatography (AC). Considering the cell lysis and AC of all purifications, the highest amount of protein was detected in the soluble fraction (Figure 9 and Figure 10). For ANDV$_{L1\text{-}200}$ N167A fused to MBP, Amylose-beads were used to bind the POI. To bind the GST tagged protein, the Glutathione Sepharose resin was used. Most of the impurities did not bind to the respective affinity beads and were eliminated in the flow-through. The performed washes showed little loss of the POI while removing contaminants. As little amounts of His-MBP-ANDV$_{L1\text{-}200}$ N167A (66 kDa) is visible at the amylose resin, the purification process was stopped. In contrast, high amounts of GST tagged protein (51 kDa) was bound to the Glutathione Sepharose resin and purification continued with the tag-removal. Beads were left in the gravity flow column for On-column cleavage (Figure 10, A), while the protein was eluted for Off-column cleavage (B).

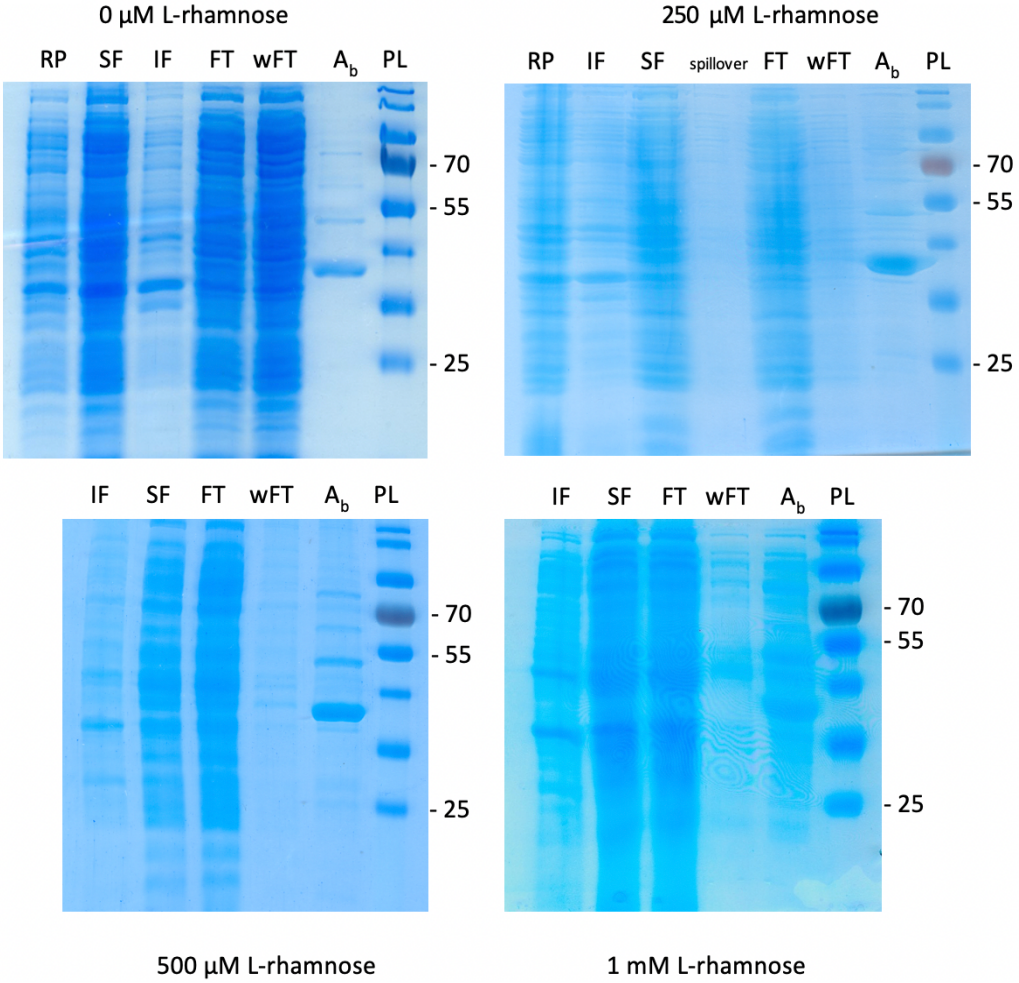

Figure 9. Assessment of His-MBP-ANDV$_{L1\text{-}200}$ N167A affinity chromatography. 12 % SDS-PAGE of the AC step for His-MBP-3C-ANDV$_{L1\text{-}200}$ N167A (66 kDa) expressed in the absence or presence of 250 μM, 500 μM and 1000 μM of L-rhamnose. **RP:** resuspended pellet, **IF:** insoluble fraction, **SF:** soluble fraction, **FT:** flow-through (FT), **wFT:** washed FT, **A$_b$:** Amylose-resin bound, **PL:** MW PAGE Ruler protein ladder (kDa), His-MBP (43 kDa).

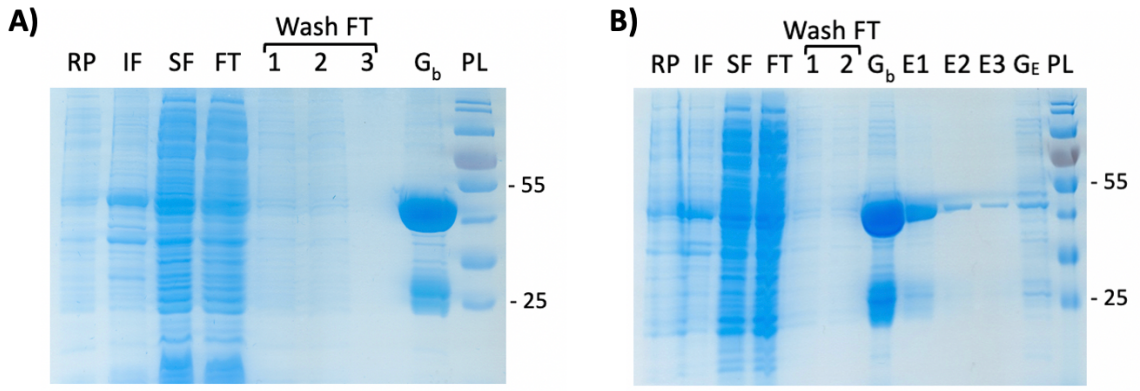

Figure 10. Assessment of His-GST-ANDV$_{L1\text{-}200}$ N167A affinity chromatography via 12 % SDS-PAGE of protein expressed in the **A)** absence or **B)** presence of 250 μM L-rhamnose. **RP:** resuspended pellet, **IF:** insoluble fraction, **SF:** soluble fraction, **FT:** flow-through (FT), **Wash FT:** washed FT of 2-3 washes, **Gb:** Glutathione Sepharose-resin bound, **E1-E3:** Eluted fractions 1-3, **PL:** MW PAGE Ruler protein ladder (kDa), His-GST (28 kDa).

Cleavage of the tag is required for downstream applications like crystallization experiments and biochemical endonuclease assays. Taking advantage of the HRV 3C cleavage site adjacent to the N-terminal side of ANDV$_{L1\text{-}200}$ N167A in the GST fused protein (51 kDa), the efficiency of proteolysis in On-column and Off-column set ups with an in-house produced GST-HRV 3C protease (46 kDa) was evaluated. As shown in Figure 11 A, after an overnight On-column reaction approximately 50 % cleavage efficacy was observed along with the low recovery of the POI (23 kDa, lane SN). Meanwhile, after an overnight Off-column reaction (Figure 11, B) results in more than 75 % cleavage of the fusion protein and also in higher recovery (see lane SN in B). The optimal pH for GST-HRV 3C protease activity ranges from 7.5 to 8.5. Unfortunately, the theoretical pI of ANDV$_{L1\text{-}200}$ N167A at 8 is impeding its soluble recovery.

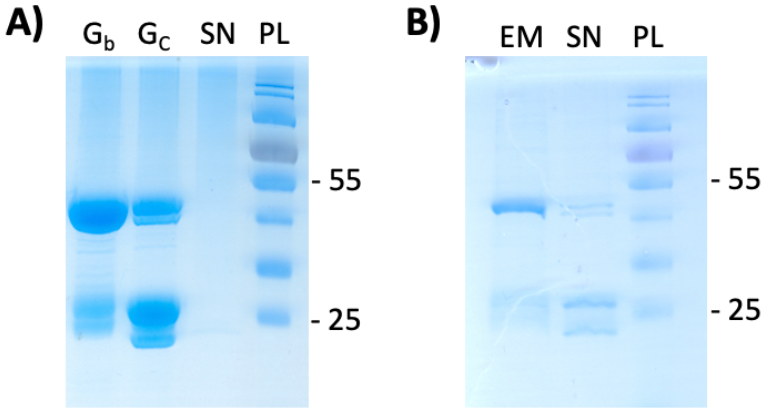

Figure 11. Assessment of the proteolysis efficiency in **A)** On-column and **B)** Off-column cleavage of His-GST-ANDV$_{L1\text{-}200}$ N167A (51 kDa) by HRV 3C protease (46 kDa) in 12 % SDS-PAGE. After cleavage His-GST appears at 28 kDa and ANDV$_{L1\text{-}200}$ N167A at 23 kDa. **G_b:** Glutathione Sepharose-resin bound, **G_C:** Glutathione Sepharose-resin after ON cleavage, **EM:** Eluate-Mix, **SN:** Supernatant after ON cleavage **PL:** MW PAGE Ruler protein ladder (kDa).

To further separate the 6xHis-GST tag from the POI the step of a CEXC was evaluated. The representative chromatogram (Figure 12) shows the elution profile after 170 mL of sample was loaded to the HiTrap SP XL 1 mL cation exchanger column; x-axis in mL, left y-axis (blue) indicating the absorbance at 280 nm and right y-axis (green) the percentage of HiTrap SP XL Buffer B in a linear NaCl gradient. Several peaks, indicating eluted protein (fractions 7, 8, 11, 14, 16, 17, 18, 22) were checked in 12 % SDS-PAGE, see Figure 13. Fractions from 177 – 179.5 mL contained the GST-tag (28 kDa) together with some amount of impurities above 50 kDa (Figure **13**, fractions 7 and 8). Elution of $ANDV_{L1\text{-}200}$ N167A (23 kDa) is visible at 183 mL (Fraction 11). However, eluted fractions starting at 184 mL (fractions 14-22) contained the POI, GST-tag and the fusion-protein.

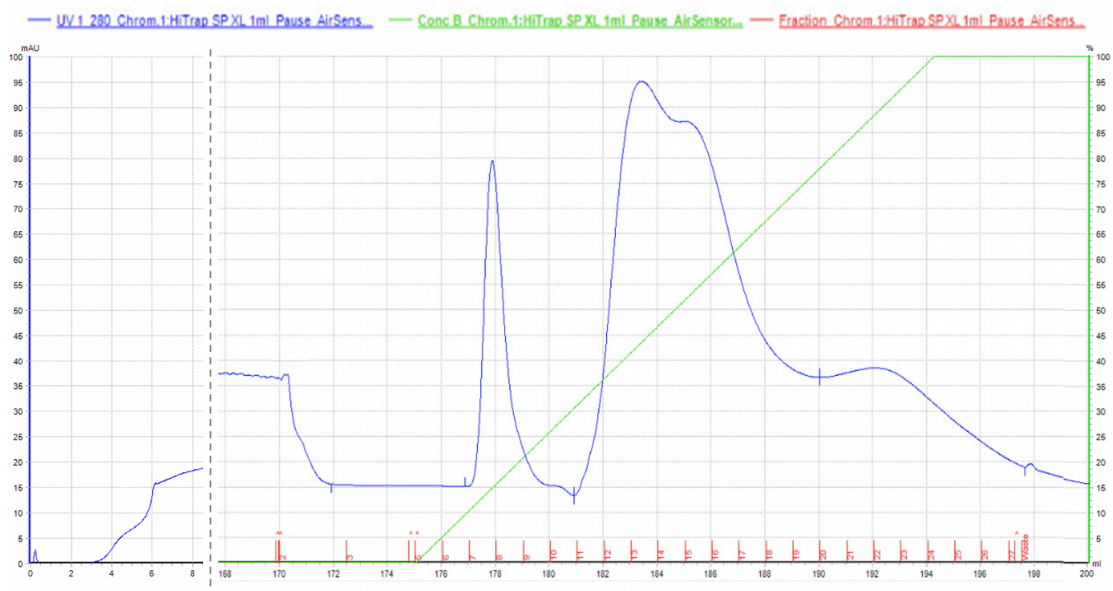

Figure 12. Chromatogram showing the CEXC-run of $ANDV_{L1\text{-}200}$ N167A. The x-axis shows the run in mL, the left y-axis the detected absorbance at λ = 280 nm and the right y-axis the %-concentration of HiTrap SP XL Buffer B during elution. After loading 170 mL of the sample, the column was washed with 5 CV of HiTrap SP XL Buffer A, followed by a 20 CV elution in a linear NaCl gradient (green line, Conc.B [%]) with HiTrap SP XL Buffer B (5 mM $MnSO_4$) in 20 CV. The blue line gives information of detected proteins measured at UV_{280} (mAU).

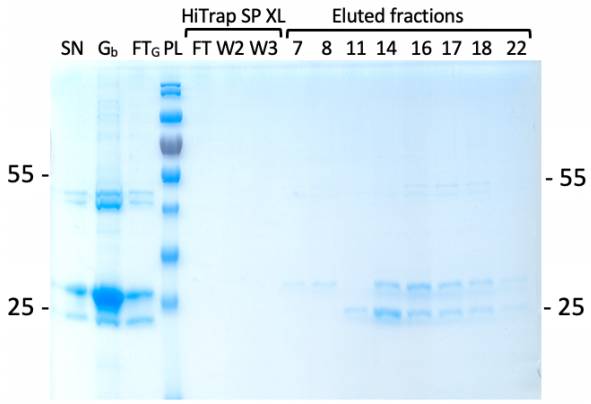

Figure 13. Assessment of His-GST-ANDV$_{L1\text{-}200}$ N167A CEXC. Fractions showing a peak in CEXC were run in 12 % SDS-PAGE. **SN:** Supernatant after cleavage, **G_b:** Glutathione Sepharose-resin, **FT_G:** Flow-through after Off-column cleavage, **PL:** MW PAGE Ruler protein ladder (kDa). **FT:** FT of HiTrap SP XL during loading (unbound sample), HiTrap SP XL wash fractions (**W2, W3)** and HiTrap SP XL eluted fractions (7, 8, 11, 14, 16, 17, 18, 22).

3.3 Generation of pOPIN_J_T-ANDV$_{L1\text{-}200}$ N167A

To improve recovery of ANDV$_{L1\text{-}200}$ N167A after cleavage of the fusion tag, a protease with a unique recognition site having a wider pH range of action could be used. thrombin cleaves the LVPR↓GS sequences and acts at a slightly acidic pH. Therefore, we proceeded to modify the pOPIN_J_ANDV$_{L1\text{-}200}$ N167A expression vector by exchanging the HRV-3C cleavage site by thrombin´s using a two-step PCR mutagenesis followed by In-Fusion cloning. The method is based on overlapping regions between the vector and the insert. For doing so, the vector is linearized with restriction enzymes and the insert is generated via PCR with complementary ends to the vector. Assembly takes place in the In-Fusion cloning reaction without the need of DNA ligase [32].

The first step to create the insert J_T-ANDV$_{L1\text{-}200}$ N167A was achieved by generating two constructs containing the coding sequence for the thrombin cleavage site (Figure 14 A). These were used as templates for the second step of the PCR mutagenesis for the final insert (395 bp, Figure **14**, B). The free dam and dcm methylated pOPIN_J_ANDV$_{L1\text{-}200}$ N167A was linearized with *Bcl*I and *Sex*AI, resulting in a 6119 bp vector (Figure 14, C). A successful In-Fusion reaction resulted in clones with the correct sequence that will be used for further expression and purification of the ANDV$_{L1\text{-}200}$ N167A.

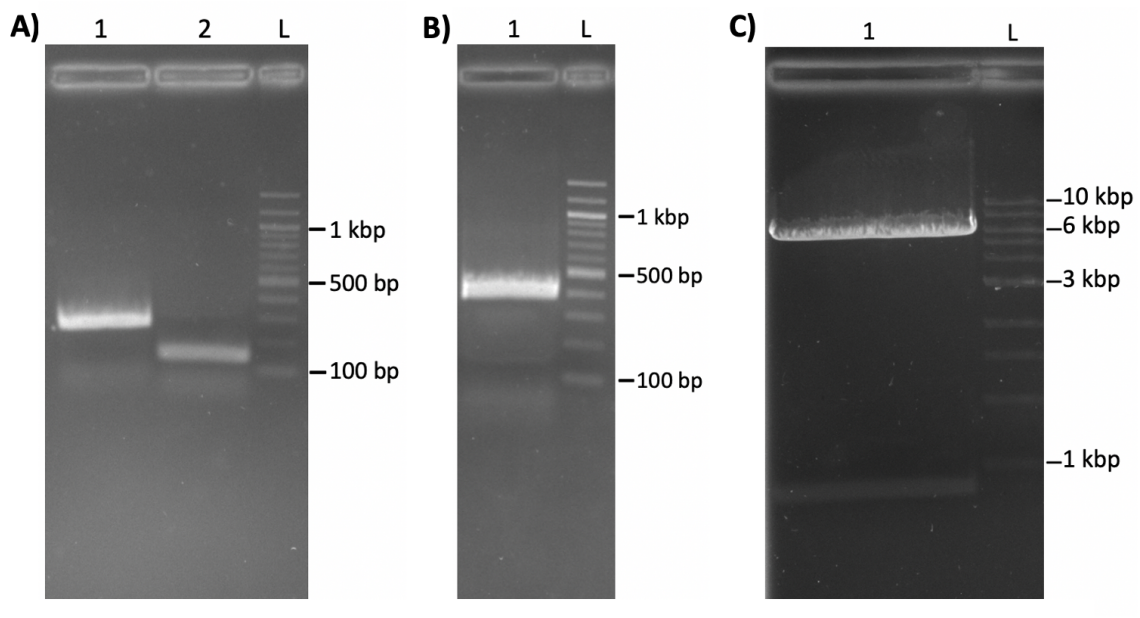

Figure 14. Agarose gels from insert generation via two-step PCR mutagenesis and plasmid linearization for In-Fusion cloning. Lane L denotes in A) and B) the 100 bp DNA ladder. Lane L in C) contains the 1 kbp DNA ladder. **A)** amplicons of 269 bp (lane 1) and 144 bp (lane 2) and in **B)** the final insert containing the thrombin cleavage site (lane 1, 395 bp), each separated in 2 % agarose gel electrophoresis is shown. **C)** 1 % agarose gel of linearized plasmid pOPIN_J_ANDV$_{L1\text{-}200}$ N167A (lane 1, 6119 bp) after digestion with *Sex*AI and *Bcl*I HF at 37 °C for 1 h.

4 Discussion

Considering the expression of $ANDV_{L1\text{-}200}$ N167A, so far, overexpression of the N-terminal domain of ANDV was difficult due to its high nuclease activity [4 10]. Lemo21(DE3) *E. coli* cells allowed expression of the POI by the T7 polymerase, which is controlled by T7 lysozyme to prevent leakage expression [18]. Moreover, when comparing the results of the respective affinity chromatographies of His-MBP-3C-$ANDV_{L1\text{-}200}$ N167A with His-GST-3C-$ANDV_{L1\text{-}200}$ N167A (see Figure 6 and 7) the GST tagged $ANDV_{L1\text{-}200}$ N167A protein gave an evident benefit to the POI expression.
Since His-GST-3C-$ANDV_{L1\text{-}200}$ N167A was purified from protein expressions containing no or 250 µM L-rhamnose, no conclusions can be drawn about the optimal L-rhamnose concentration. Low expression levels of His-MBP-3C-$ANDV_{L1\text{-}200}$ N167A were reproducible, see Figure 8. Further purification of His-MBP-$ANDV_{L1\text{-}200}$ N167A was not considered due to low yields. Difficulties in expressing the ANDV domain fused to MBP in another *E. coli* strain was observed in a previous study as well [4].

Checking on the amount of soluble protein during affinity chromatography, a clear difference of insoluble (IF) to soluble fraction (SF) is visible in 12 % SDS-PAGE (affinity chromatography). Most proteins are visible in the soluble fraction, which assumes well set parameters (temperature, IPTG concentration, agitation and aeration e.g.) for protein expression in Lemo21(DE3) *E. coli*. The advantage of affinity tags as an efficient first purification step was exploited, to isolate the fusion-protein by using affinity chromatography as [33]. As seen in Figure 10, most impurities remain in the flow-through (FT, wash FT) which results in a purer sample. High amounts of the GST-fused POI (51 kDa) are bound to the Glutathione Sepharose resin and allowed further purification by digestion with HRV-3C protease.

Even though fusion tags can facilitate the expression and purification of proteins [34], (big) affinity tags have to be removed in downstream processes, since they could alter the function or structure of the POI [35], especially with regard to endonuclease assays or target-studies.
Despite the advantage of retaining the GST-tag in the Glutathione Sepharose resin and the recombinant GST-HRV 3C protease while performing an On-column protease digestion, when comparing the beads before and after cleavage (G_b and G_C, Figure 11, A), only approximately 50 % of the fusion protein was cleaved from the tag. This suggested, that the cleavage site was not well accessible to the HRV 3C protease. Less activity of On-column digestion is a downside of this method [36]. Additionally, most of the cleaved protein (23 kDa) remained in the beads, see G_C and SN, (Figure 11 A) which could be explained that the POI becomes insoluble once it is close to its

pI in the cleavage buffer. In contrast, the Off-column cleavage of the fusion-protein resulted in more than 75 % of the digestion product (see EM and SN, Figure 11, B). To minimize the carryover of GST from the cleavage recovery in the sample, an additional incubation with the Glutathione Sepharose resin or an increased number of beads for cleavage cleanup could improve the step. Yet, this holds the risk to lose protein by unspecific binding to the beads (see Figure 13, G_b). However, the optimal pH-range for the protease activity is between pH 7.5 – 8.5 and interferes with the calculated isoelectric point of $ANDV_{L1\text{-}200}$ N167A at pH 8 [4].

The CEXC step did not achieve a further separation of contaminants as shown in Figure 13. Impurities at higher molecular weight (>40 kDa) are not considered as critical because they would be separated in SEC. However, $ANDV_{L1\text{-}200}$ N167A and the GST-tag are similar in size and therefore, removal of the tag is crucial before entering into SEC. The net charge of His-GST-3C (1.43) is close to $ANDV_{L1\text{-}200}$ N167A (4.41) at pH 6 and explains close elution to the POI using an increasing NaCl gradient. By expanding the gradient to more CVs, one might be able to achieve a better separation of $ANDV_{L1\text{-}200}$ N167A and impurities. Otherwise, a step-gradient with a certain NaCl concentration (350 mM e.g., see Figure 12) can be considered in order to reach higher purity.

Since the tag cleavage-reaction is a crucial step to be improved, the exploitation by a different cleavage site was considered. Using thrombin for tag-removal allows cleavage at slight acidic pH [34], more fitting for $ANDV_{L1\text{-}200}$ N167A as it results in a higher protein stability. By avoiding precipitation of proteins, a better recovery after cleavage and separation in CEXC is expected. Therefore, the generation of pOPIN_J_T-$ANDV_{L1\text{-}200}$ N167A construct confirmed by sequencing may allow the expression of a new fusion protein that could improve the purification process.

In summary, high expression levels of $ANDV_{L1\text{-}200}$ N167A fused to GST were achieved in Lemo21(DE3) *E. coli*. Based on this perception, the HRV-3C recognition site was successfully substituted by thrombin. The new plasmid pOPIN_J_T-$ANDV_{L1\text{-}200}$ N167A gets transformed into Lemo21(DE3) *E. coli* and His-GST-T-$ANDV_{L1\text{-}200}$ N167A will be expressed at different L-rhamnose concentrations. By taking the advantage of using the thrombin cleavage-site, it is expected to recover the protein from the tag-removal reaction more stable and in high amounts. Further, CEXC would profit from cleavage showing better separation of the POI and impurities and minimize the time for concentration for the final SEC. Gel filtration would finish the purification procedure of $ANDV_{L1\text{-}200}$ N167A and results in a pure sample useful for crystallization and functional studies.

5 References

1. Martinez VP, Bellomo C, San Juan J, et al. Person-to-person transmission of Andes virus. *Emerg Infect Dis*. 2005;11(12):1848-1853. doi:10.3201/eid1112.050501
2. Vaheri A, Strandin T, Hepojoki J, et al. Uncovering the mysteries of hantavirus infections. *Nat Rev Microbiol*. 2013;11(8):539-550. doi:10.1038/nrmicro3066
3. Alonso DO, Pérez-Sautu U, Bellomo CM, et al. Person-to-person transmission of Andes virus in hantavirus pulmonary syndrome, Argentina, 2014. *Emerg Infect Dis*. 2020;26(4):756-759. doi:10.3201/eid2604.190799
4. Fernández-García Y, Reguera J, Busch C, et al. Atomic Structure and Biochemical Characterization of an RNA Endonuclease in the N Terminus of Andes Virus L Protein. *PLoS Pathog*. 2016;12(6):1-18. doi:10.1371/journal.ppat.1005635
5. Bi Z, Formenty PBH, Roth CE. Hantavirus Infection: a review and global update. *J Infect Dev Ctries*. 2008;2(01):3-23. doi:10.3855/jidc.317
6. Walker PJ, Siddell SG, Lefkowitz EJ, et al. Changes to virus taxonomy and the Statutes ratified by the International Committee on Taxonomy of Viruses (2020). *Arch Virol*. 2020;165(11):2737-2748. doi:10.1007/s00705-020-04752-x
7. Hepojoki J, Strandin T, Lankinen H, Vaheri A. Hantavirus structure - Molecular interactions behind the scene. *J Gen Virol*. 2012;93(8):1631-1644. doi:10.1099/vir.0.042218-0
8. Wang Z, Mir MA. Andes Virus Nucleocapsid Protein Interrupts Protein Kinase R Dimerization To Counteract Host Interference in Viral Protein Synthesis. *J Virol*. 2015;89(3):1628-1639. doi:10.1128/jvi.02347-14
9. Garrido JL, Presscott J, Calvo M, et al. Two recombinant human monoclonal antibodies that protect against lethal Andes hantavirus infection in vivo. *Sci Transl Med*. 2019;11(475):1-12. doi:10.1126/scitranslmed.aaw4903
10. Heinemann P, Schmidt-Chanasit J, Gunther S. The N Terminus of Andes Virus L Protein Suppresses mRNA and Protein Expression in Mammalian Cells. *J Virol*. 2013;87(12):6975-6985. doi:10.1128/jvi.00043-13
11. Martínez VP, Di Paola N, Alonso DO, et al. "Super-Spreaders" and Person-to-Person Transmission of Andes Virus in Argentina. *N Engl J Med*. 2020;383(23):2230-2241. http://www.nejm.org/doi/10.1056/NEJMoa2009040
12. Cheng E, Mir MA. Signatures of Host mRNA 5′ Terminus for Efficient Hantavirus Cap Snatching. *J Virol*. 2012;86(18):10173-10185. doi:10.1128/jvi.05560-11
13. Baluna R, Vitetta ES. Vascular leak syndrome: A side effect of immunotherapy. *Immunopharmacology*. Published online 1997. doi:10.1016/S0162-3109(97)00041-6
14. Vera-Otarola J, Solis L, Lowy F, et al. The Andes Orthohantavirus NSs Protein Antagonizes the Type I Interferon Response by Inhibiting MAVS Signaling. *J Virol*. 2020;94(13):1-22. doi:10.1128/jvi.00454-20
15. Hagen M, Chung TD, Butcher JA, Krystal M. Recombinant influenza virus polymerase:

requirement of both 5' and 3' viral ends for endonuclease activity. *J Virol*. 1994;68(3):1509-1515. doi:10.1128/jvi.68.3.1509-1515.1994

16. Mir MA, Duran WA, Hjelle BL, Ye C, Panganiban AT. Storage of cellular 5′ mRNA caps in P bodies for viral cap-snatching. *Proc Natl Acad Sci U S A*. 2008;105(49):19294-19299. doi:10.1073/pnas.0807211105
17. Saida F, Uzan M, Odaert B, Bontems F. Expression of Highly Toxic Genes in E. coli: Special Strategies and Genetic Tools. *Curr Protein Pept Sci*. 2006;7(1):47-56. doi:10.2174/138920306775474095
18. Vikström D, Klepsch M, Wagner S, Samuelson JC, De Gier J-W. *Escherichia coli* Lemo21 (DE3) : A T7 RNA polymerase-based protein overexpression platform for the routine and difficult targets. *New Engl Biolabs, Inc*. 2018;98(1):0-2.
19. Dumon-Seignovert L, Cariot G, Vuillard L. The toxicity of recombinant proteins in Escherichia coli: A comparison of overexpression in BL21(DE3), C41(DE3), and C43(DE3). *Protein Expr Purif*. 2004;37(1):203-206. doi:10.1016/j.pep.2004.04.025
20. Hunter M, Yuan P, Vavilala D, Fox M. Optimization of Protein Expression in Mammalian Cells. *Curr Protoc Protein Sci*. 2019;95(1):1-28. doi:10.1002/cpps.77
21. Leis B, Angelov A, Liebl W. Screening and Expression of Genes from Metagenomes. In: *Advances in Applied Microbiology*. ; 2013:1-68. doi:10.1016/B978-0-12-407678-5.00001-5
22. Mattanovich D, Branduardi P, Dato L, Gasser B, Sauer M, Porro D. Recombinant protein production in yeasts. *Methods Mol Biol*. Published online 2012. doi:10.1007/978-1-61779-433-9_17
23. Ahmad M, Hirz M, Pichler H, Schwab H. Protein expression in Pichia pastoris: Recent achievements and perspectives for heterologous protein production. *Appl Microbiol Biotechnol*. Published online 2014. doi:10.1007/s00253-014-5732-5
24. Cereghino JL, Cregg JM. Heterologous protein expression in the methylotrophic yeast Pichia pastoris. *FEMS Microbiol Rev*. 2000;24(1):45-66. doi:10.1016/S0168-6445(99)00029-7
25. Berrow NS, Alderton D, Sainsbury S, et al. A versatile ligation-independent cloning method suitable for high-throughput expression screening applications. *Nucleic Acids Res*. 2007;35(6):e45-e45. doi:10.1093/nar/gkm047
26. Geisse S, Gram H, Kleuser B, Kocher HP. Eukaryotic expression systems: A comparison. *Protein Expr Purif*. Published online 1996. doi:10.1006/prep.1996.0101
27. Wurm FM. Production of recombinant protein therapeutics in cultivated mammalian cells. *Nat Biotechnol*. 2004;22(11):1393-1398. doi:10.1038/nbt1026
28. Cheng E, Wang Z, Mir MA. Interaction between Hantavirus Nucleocapsid Protein (N) and RNA-Dependent RNA Polymerase (RdRp) Mutants Reveals the Requirement of an N-RdRp Interaction for Viral RNA Synthesis. *J Virol*. 2014;88(15):8706-8712. doi:10.1128/jvi.00405-14
29. Variables TP, Assays QC, Efficiency H, Protocol T. C3013I_Datasheet. *New Engl Biolabs, Inc*. 21. https://www.nebiolabs.com.au/-/media/catalog/datacards-or-manuals/c3013idatasheet-lot0091412.pdf

30. Costa S, Almeida A, Castro A, Domingues L. Fusion tags for protein solubility, purification, and immunogenicity in Escherichia coli: The novel Fh8 system. *Front Microbiol*. 2014;5(FEB):1-20. doi:10.3389/fmicb.2014.00063
31. Young CL, Britton ZT, Robinson AS. Recombinant protein expression and purification: A comprehensive review of affinity tags and microbial applications. *Biotechnol J*. 2012;7(5):620-634. doi:10.1002/biot.201100155
32. Enzymes DNAM. DNA MODIFYING ENZYMES Instruction Manual NEBuilder ® HiFi DNA Assembly Master Mix/NEBuilder HiFi DNA Assembly Cloning Kit. Published online 2017.
33. Jensen DM. Simple and Efficient Purification of Recombinant Proteins using the Heparin-Binding Affinity Tag. *Physiol Behav*. 2018;176(1):1570-1573. doi:10.1038/s41395-018-0061-4.
34. Waugh DS. An overview of enzymatic reagents for the removal of affinity tags. *Protein Expr Purif*. 2011;80(2):283-293. doi:10.1016/j.pep.2011.08.005
35. Ullah R, Ali Shah M, Tufail S, et al. Activity of the Human Rhinovirus 3C Protease Studied in Various Buffers, Additives and Detergents Solutions for Recombinant Protein Production. *PLoS One*. 2016;11(4):1-13. doi:10.1371/journal.pone.0153436
36. Harper S, Speicher DW. Expression and purification of GST fusion proteins. *Curr Protoc Protein Sci*. 2008;2008(May):1-26. doi:10.1002/0471140864.ps0606s52

Appendix

Solutions and buffers

This section contains tables showing the composition of media, buffers and additives used in this thesis.

Table I. I. Buffers and Solutions for electrophoresis

Name	Composition
4X SDS PAGE Loading buffer	200 mM Tris pH 6,8 8 % SDS 0.1 % Bromophenolblue 40 % Glycerol 100 mM DTT
10X Tris-Glycine-SDS buffer	35 mM SDS 250 mM Tris 1.9 M Glycin
50X TAE Buffer	2 M Tris 50 mM Na_2 EDTA pH 8 5.7 % pure acetic acid
Coomassie Safe Stain	0.008 % Brilliant blue G-250 dissolved in 500 ml MilliQ water 0.34 % HCl (37 %, fuming)

Table I. II. Cultivation media and additives

Name	Composition
LB-Media	5 g/L Yeast-Extract 10 g/L Peptone 10 g/L NaCl
TB-Media	12 g/L Peptone 24 g/L Yeast-Extract 0,4 % Glycerol (100 %)
10 X TB Additives	0,17 M KH_2PO_4 0,72 M K_2HPO_4

Table I. III. Buffers for purification of His-MBP-3C-$ANDV_{L1\text{-}200}$ N167A.

Name	Composition
Lysis-Buffer, pH 7	50 mM Tris 300 mM NaCl 1 mM EDTA 5 % Glycerol (100 %)

Table I. IV. Buffer for purification of His-GST-3C-ANDV$_{L1\text{-}200}$ N167A expressed with 0 μM L-rhamnose.

Name	Composition
Lysis-Buffer/ HRV 3C Cleavage Buffer, pH 7	21,3 mM NaH_2PO_4 x H_2O 28,7 mM Na_2HPO_4 300 mM NaCl 1 mM EDTA 5 % Glycerol (100 %)
Wash-Buffer, pH 7	21,3 mM NaH_2PO_4 x H_2O 28,7 mM Na_2HPO_4 1000 mM NaCl 1 mM EDTA 5 % Glycerol (100 %)
HiTrap SP XL Buffer A, pH 7	21,3 mM NaH_2PO_4 x H_2O 28,7 mM Na_2HPO_4
HiTrap SP XL Buffer B, pH 7	21,3 mM NaH_2PO_4 x H_2O 28,7 mM Na_2HPO_4 1000 mM NaCl

Table I. V. Buffer for purification of His-GST-3C-ANDV$_{L1\text{-}200}$ N167A expressed with 250 μM L-rhamnose

Name	Composition
Lysis-Buffer/ HRV 3C Cleavage Buffer, pH 7,5	50 mM Tris 500 mM NaCl 1 mM EDTA 5 % Glycerol (100 %)
Wash-Buffer, pH 7,5	50 mM Tris 1000 mM NaCl 1 mM EDTA 5 % Glycerol (100 %)
Elution-Buffer, pH 7,5	50 mM Tris 500 mM NaCl 50 mM reduced Glutathione 1 mM EDTA 5 % Glycerol (100 %)
HiTrap SP XL Buffer A, pH 7	50 mM MES
HiTrap SP XL Buffer B, pH 7	50 mM MES 1000 mM NaCl