University of South Bohemia in České Budějovice
Faculty of Science

Mechanism of dsRNA virus replication: cloning, production and structural characterization of N-terminal domain of σNS

Bachelor thesis

Eliška Kotounová

Supervisor: RNDr. Zdeněk Franta, Ph.D.

Co-supervisor: prof. RNDr. Ivana Kutá-Smatanová, Ph.D.

České Budějovice 2019

Kotounová E., 2019: Mechanism of dsRNA virus replication: cloning, production and structural characterization of N-terminal domain of σNS. Bc. Thesis, in English – 53 p., Faculty of Science, University of South Bohemia, České Budějovice, Czech Republic.

Annotation:

Avian reovirus is an important pathogen, which causes substantial losses in poultry industry. σNS protein is essential for the virus replication, however, the mechanism and assortment of genomic segments during virus replication is yet not well understood. The structure of σNS is not known as previous attempts at crystallization yielded non-diffracting crystals due to the flexible regions in the molecule. This thesis deals with the production and purification of the N-terminal domain of σNS, which lacks the exposed flexible region.

Declaration:

České Budějovice, 13.5.2019

…………………………………

Eliška Kotounová

Table of contents

List of abbreviations

aa – amino acid

AIEC - anion exchange chromatography

APS – ammonium persulfate

ARV – avian reovirus

bp – base pair

CCB - Coomassie R-250 Brilliant Blue

ddH_2O – double-distilled water

dNTP - deoxynucleotide triphosphate

dsRNA – double-stranded RNA

EDTA – ethylenediaminetetraacetic acid

His – histidine

HRP – horseradish peroxidase

IMAC - immobilized metal affinity chromatography

IPTG – isopropyl β-D-1-thiogalactopyranoside

kb - kilobase

kDa - kilodalton

LB – lysogeny broth

MM – master mix

mRNA – messenger RNA

MRV – mammalian reovirus

OD_{600} – optical density measured at 600 nm

ORF – open reading frame

PAGE – polyacrylamide gel electrophoresis

PCR - polymerase chain reaction

PCT – pre-crystallization test

SDS – sodium dodecyl sulfate

SEC – size exclusion chromatography

ssDNA – single-stranded DNA

ssRNA – single-stranded RNA

TAE – Tris-acetate-EDTA

TB – terrific broth

TBS – Tris-buffered saline

TEMED – tetramethylethylenediamine

UV - ultraviolet

1 Introduction

1.1 *Reoviridae* family

Family *Reoviridae* is the largest and the most diverse family of double-stranded RNA (dsRNA) viruses containing important human and animal pathogens (eg. Rotaviruses). The name *Reoviridae* was derived as an acronym from "respiratory, enteric and orphan viruses" as the virus was originally isolated from respiratory and enteric samples, but with no associated symptoms of the clinical disease (Markey et al., 2013). These viruses enclose their genomes, consisting of 9-12 segments of linear dsRNA, in icosahedral, non-enveloped, double layered protein capsids with a size of 60-80 nm. Host organisms of this family extend from fungi, plants, and insects to fish, reptiles, birds, and mammals. Family *Reoviridae* can be divided into two groups based on the presence or the absence of large protein spikes ("turrets") situated on the surface of the virus particle as showed in Figure 1. The genus *Orthoreovirus* with its two principal members avian reovirus (ARV) and mammalian reovirus (MRV) is a part of the "turreted" group while globally important genus *Rotavirus* belongs to the "non-turreted" group (Day, 2009; King et al., 2011).
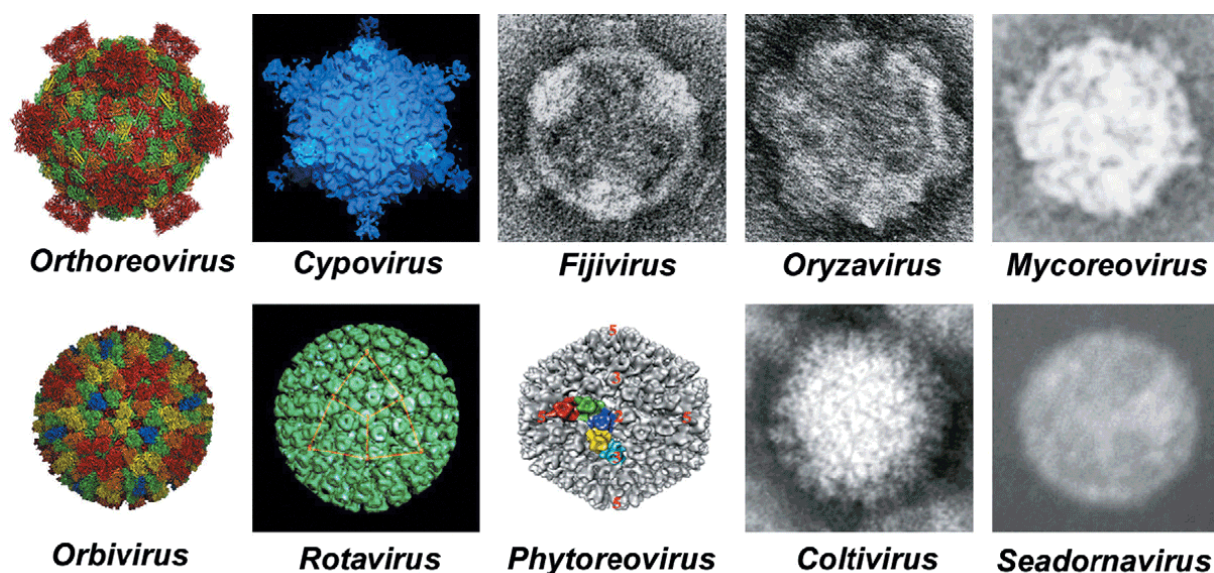

Figure 1: A comparison of two particle types present amongst the genera of the *Reoviridae* family (King et al., 2011).

Biological properties diverge among the genera. Orthoreoviruses and rotaviruses replicate exclusively in certain vertebrate species, and they are transmitted by respiratory or fecal-oral routes. Other vertebrate viruses, including orbiviruses, coltiviruses,

and seadornaviruses can replicate in both vertebrates and arthropods while plant viruses as phytoreoviruses, fijiviruses, and oryzaviruses can replicate in plants and arthropods. Cypoviruses, which infect insects, are spread by contact or fecal-oral routes (King et al., 2011).

1.2 Avian reovirus: pathogenicity

ARV cause significant losses in the commercial poultry industry. Although in 85-90% cases, the isolated reoviruses are non-pathogenic, the infection may result into development of a disease, whose nature is dependent on host age, immune system or virus pathotype. The infection may cause several disease conditions such as viral arthritis (sometimes referred to as tenosynovitis), chronic respiratory disease, enteric disease, stunting and malabsorption syndromes. Viral arthritis is the most readily diagnosed and the best defined reovirus-associated disease. It predominantly affects meat type chickens (broilers), but it has been reported to affect turkeys and light meat-type chickens. The main disease symptom is a swelling of hock joints and inflammation causing acute lameness. In the most severe cases the disease leads to tendon rupture and erosion of articular cartilage. The economic losses are mostly due to the poor growth, inability to reach feed and water, and death by trampling by healthy birds. Except for chickens, avian reoviruses have been also isolated from other bird species such as turkeys, Muscovy ducks, pigeons, geese or African green parrots (Jones, 2000; Jones, 2013). The ARV infections have been also diagnosed in the wild avian species and potential transmission to farmed birds could occur (Styś-Fijoł et al., 2017). The virus transmits primarily via fecal-oral route, but also a transmission via respiratory tract and transovarial transmission may occur. The virus stays in the organism for several weeks and the bird resistance to the infection is age dependent. It has been recorded that birds in the post-hatch period are more susceptible to evolve arthritis than the birds two weeks old or older (Jones, 2000). In addition to the viral arthritis, reoviruses have been also associated with other disease conditions such as growth retardation, pericarditis, myocarditis, hepatitis or osteoporosis. The virus itself can be the pathogenic agent or can act in combination with other pathogenic agents such as *Mycoplasma synoviae* or *Staphylococcus sp.*, which can exacerbate the disease (Jones, 2000; Jones, 2013).

Considering that the ARVs infections are widespread and can be transmitted vertically, maintaining the flocks free from the infection is, in fact, impossible. Attenuated and/or inactivated virus vaccines represent the main approach to control the reovirus infections. However, all the commercial vaccines belong to the same serotype and do not provide a

sufficient protection against many strains of circulating reoviruses that are antigenically distinct (Jones, 2013; Sellers, 2017).

1.3 Avian reovirus: characteristics

Avian reoviruses have been considered similar to the well-studied mammalian reoviruses, therefore, they have not been properly characterized in the past. Nevertheless, studies over last 30 years have reported that ARV possess many unique properties and activities that are different to those found in MRV. Both ARV and MRV share many morphological and physicochemical characteristics, however, they vary in pathogenicity, host range, and further biological properties (Day, 2009). For example, ARV, in the contrast to MRV, causes fusion of infected cells, which results in the formation of multinucleated cellular syncytia but lacks the ability to agglutinate red blood cells (Benavente and Martínez-Costas, 2007).

Nonenveloped ARV reovirions have 85.7 nm in diameter and are composed of the central compartment, called the inner core (Figure 2), containing the genome segments and enclosed by inner and outer capsid with icosahedral symmetry (Zhang et al., 2005; King et al., 2011). As shown in Figure 2, the shell of the central compartment is formed by λA protein and protein σA seated on the top of λA to stabilize the shell. Turrets, extending from the inner core through the outer capsid, are composed of pentamers of λC protein. λC protein is accompanied by trimers of σC protein which serves for cell attachment. The particle is protected by the outer capsid shell made of proteins μB and σB (Benavente and Martínez-Costas, 2007).
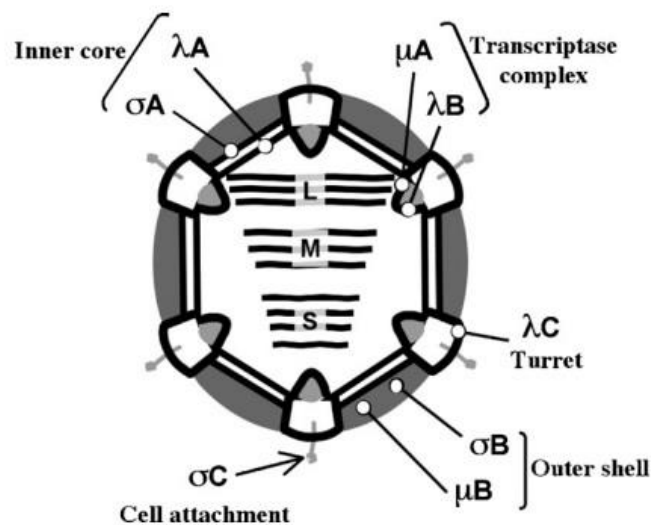

Figure 2: Schematic representation of ARV virion (Benavente and Martínez-Costas, 2007).

The genome of ARV consists of 10 linear dsRNA segments with a total length of 23,593 bp (King et al., 2011; Benavente and Martínez-Costas, 2007; NCBI). According to the electrophoretic mobility, the dsRNA segments are divided into 3 classes: large (L), medium (M), and small (S). Both the L-class and M-class consist of 3 segments each (L1, L2, L3, and M1, M2, M3), while there are 4 segments of the S-class (S1, S2, S3, S4) (King et al., 2011; Benavente and Martínez-Costas, 2007). All of the 10 ARV segments are unique and the sum of their length corresponds to the total size of the viral genome, therefore, an infectious virus has to contain one copy of each segment. There are no single-stranded overhangs at the end of the molecule, each segment is fully base paired (Dimmock, 2007). All of the genomic segments are monocistronic except for the S1 segment which encodes 3 translation products (Bodelón et al., 2001). The negative strand of each segment serves as a template for synthesis of mRNA and contains a pyrophosphate group at the 5' end, while the positive strand contains type-1 cap at this position. The genomic dsRNAs do not contain poly(A)tails and covalently linked proteins. In addition to the dsRNA genome segments, ARV virions contain many small adenine-rich, single-stranded oligonucleotides of unknown function (Benavente and Martínez-Costas, 2007; King et al., 2011).

There are at least 12 primary translation products expressed from the ARV genome. 8 of them represent structural proteins which are integrated into progeny virions, and remaining 4 are non-structural proteins found only in the infected cells and not in the mature reovirions. The proteins are named based on the class of genomic segment that encodes them. All proteins are encoded on only one strand of the dsRNA segment. L-class genes encoded proteins are denoted as lambda (λ), M-class genes as mu (µ), and S-class genes as sigma (σ). An alphabetical letter (λA, λB, λC, µA, µB, σA, σB, and σC) has been assigned to each structural protein within its class to distinguish them from the MRV structural proteins, which have been assigned a number (eg. λ1, λ2). Eight of the ARV structural proteins above mentioned are primary translation products. There are two more proteins µBN and µBC, formed by post-translational cleavage of µB (Varela et al., 1996). Except for the structural proteins, ARV expresses several non-structural proteins (Table 1). The two major non-structural proteins are µNS and σNS encoded by M3 and S4 genes respectively. Both of these proteins appear promptly in the cytoplasm of infected cells together with cleavage products of µNS, termed µNSC and µNSN (Rodríguez-Grille et al., 2014). Apart from µNS and σNS, there are two other non-structural proteins, named p10 and p17, which are encoded by the first and

the second cistron of the S1 genomic segment, and indeed found in the ARV infected cells (Benavente and Martínez-Costas, 2007; Bodelón et al., 2001).

Table 1: Summary of ARV expressed proteins and their functions (Benavente and Martínez-Costas, 2007).

Genome segment	Protein	Distribution	Function
L1	λA	Inner core	Core-shell scaffold
L2	λB	Inner core	RNA dependent RNA polymerase
L3	λC	Turrets	Capping enzyme
M1	μA	Inner core	Putative transcriptase co-factor
M2	μB, μBN, μBC	Outer capsid	Penetration
M3	μNS, μNSC	Nonstructural	Formation of viral factories and protein recruitment
S1	σC	Outer capsid	Cell attachment
	p10	Nonstructural	Permeabilizing/fusogenic
	p17	Nonstructural	Unknown
S2	σA	Inner core	dsRNA binding, anti-interferon activity
S3	σB	Outer capsid	Unknown
S4	σNS	Nonstructural	ssRNA binding

1.4 Avian reovirus: replication cycle

The attachment of ARV to the host cell occurs via a specific interaction between viral cell-attachment protein σC and cell surface receptors. As ARV can attach and replicate also in most mammalian cells, it is assumed that the ARV receptor is an omnipresent cell surface protein (Benavente and Martínez-Costas, 2007). It has been shown that protein σC is also associated with the neutralization of virus infectivity and with the reovirus-induced formation of syncytia (Grande et al., 2002). While the enveloped viruses penetrate into the cell by fusion of their lipidic envelopes with the cytoplasmic membrane, the nonenveloped viruses cross the membrane by receptor-mediated endocytosis. Once inside the cell, the parental virion

undergoes a proteolytic process called uncoating in which the outer capsid proteins are removed and viral cores are released into the cytoplasm and activated for transcription (Benavente and Martínez-Costas, 2007).

The gene expression starts with transcription of viral mRNA from each of the dsRNA segment under the catalysis od virus-encoded dsRNA-dependent RNA polymerase. The viral mRNAs are synthesized inside the core and leave the particle through the turrets. The viral mRNAs exhibit dual function in the infected cell: they are used for protein synthesis by the ribosomes as well as they function as a genome segment precursor. Protein synthesis is detected at an early stage of the infection and although all transcripts are produced in a similar amount, the most abundant viral proteins are μBC, σB and σNS. Later on during the infection, majority of the proteins, which are synthesized in the infected cells, are those of viral origin (Benavente and Martínez-Costas, 2007).

ARV virions assemble, as all members of *Reoviridae* family, within cytoplasmic phase-dense inclusion bodies known as viral factories or viroplasms. The viral factories fulfil several important functions as (1) protection of the viral core from recognition by the host immune system, (2) keeping the concentration of viral proteins within separate locations, (3) regulation of viral replicative cycle, and (4) providing a scaffold in which the progeny virions are assembled (Shah et al., 2017). Viral factories are primarily made of association of non-structural proteins μNS and σNS and lack any membranes and cellular organelles. Moreover, μNS is the only protein which was shown to form viroplasms even when produced alone in the cell and was shown to be able to recruit σNS and λA in viral factories (Benavente and Martínez-Costas, 2007; Shah et al., 2017) (Figure 3).

It has been shown by Tourís-Otero et al. (2004) that the ARV morphogenesis occurs exclusively within viral factories of globular shape which are not associated with microtubules. Further results implied that after the synthesis of protein components of the particle, the viral cores are assembled within 30 minutes, and the morphogenesis is completed by the addition of outer capsid proteins (Figure 4). Genome packaging is a crucial step in the life cycle of the virus as all dsRNA segments are essential for replication and assembly of the progeny virions. Therefore, it is necessary to ensure that progeny virions contain one copy of each segment. This process is facilitated by specific RNA-RNA interactions mediated by non-structural protein σNS. However, the exact mechanism of dsRNA segment packaging into virions remains unknown (Borodavka et al., 2015).
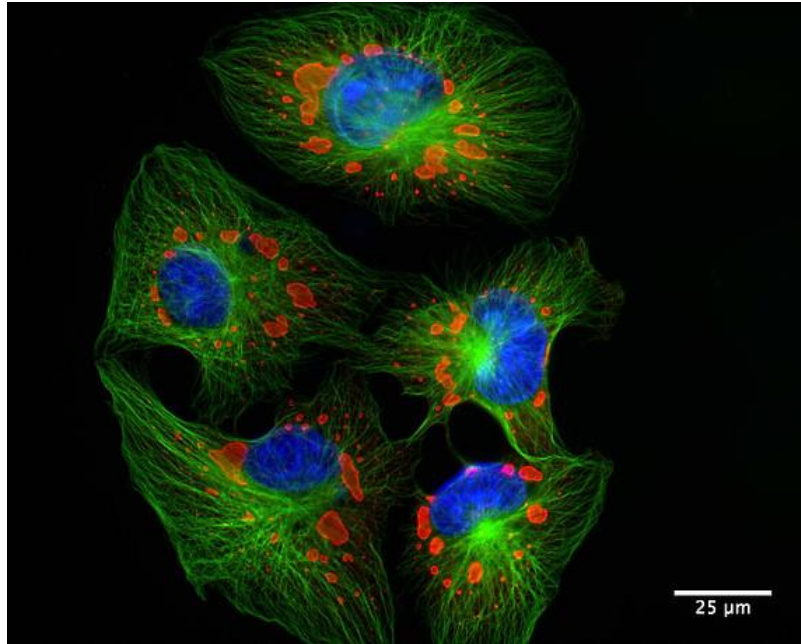
Figure 3: Immunofluorescent imaging of reovirus viral factories of Mammalian orthoreovirus type 3 (T3D). Viral factories and microtubules stained with anti-µNS (red) and an anti-tubulin (green) antibody, cell nuclei were visualised with 4,6-diamidino-2-phenylindole (blue) (Shah et al., 2017).
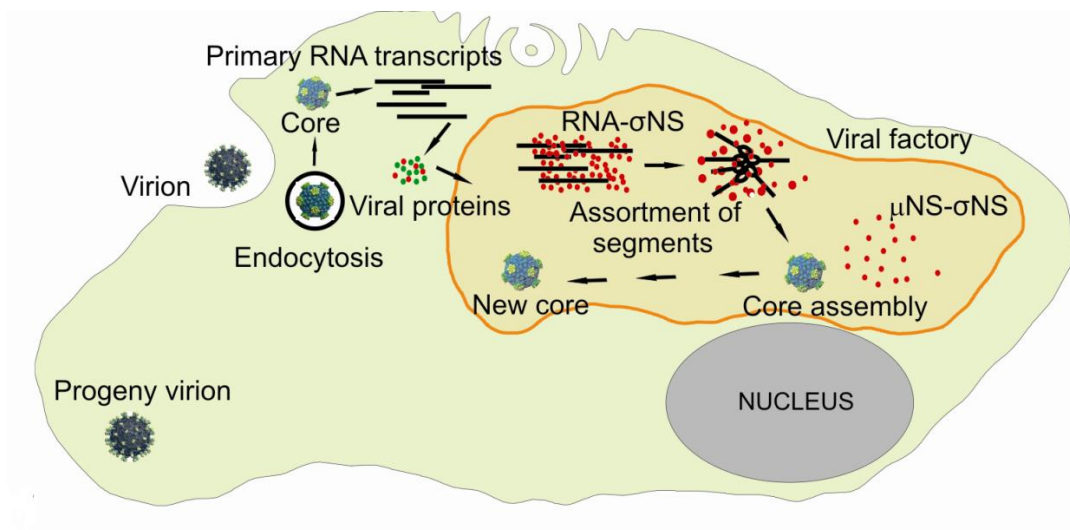
Figure 4: Scheme of the avian reovirus life cycle (Borodavka et al., 2015).

1.5 Avian reovirus: σNS

The non-structural protein σNS, which is encoded by the S4 genome segment, consist of 367 amino acids with a molecular weight of 40.64 kDa. The protein contains three epitopes which are highly conserved among virus strains (Huang et al., 2005). σNS is readily detectable in the cytoplasm of infected cells and after the synthesis, it is promptly recruited into viral factories via association with µNS (NCBI, Benavente and Martínez-Costas, 2007). σNS forms

oligomers that bind ssRNA, as well as ssDNA, but have a very low affinity for double-stranded nucleic acids. The binding of ssRNA occurs in a sequence-independent manner and the minimal ssRNA length for binding is between 10-20 nucleotides (Tourís-Otero et al., 2005), nonetheless, it has been recently shown by Bravo et.al (2018), that the affinity is dependent on length. σNS exists as a stable elongated hexamer, which assembles as a trimer of dimers. σNS hexamers bind ssRNAs with high affinity and single hexamer was shown to bind one RNA strand at a time while it can further oligomerize to an octamer when bound to two RNAs (Bravo et al., 2018). On the top of that, σNS destabilizes the RNA secondary structure by binding to partially single-stranded regions which leads to the helix unwinding at low micromolar and above concentrations. Binding of several ssRNAs to σNS mediates the annealing. As σNS has low affinity for dsRNA, it dissociates from newly formed duplexes to be able to bind new ssRNA (Figure 5) (Borodavka et al., 2015). Upon the RNA binding, σNS undergoes transition from hexamer to octamer to increase the efficiency of RNA unwinding. However, even though σNS is able to anneal short RNA oligonucleotides, it fails to promote interactions between full-length genomic ssRNA. This might be caused by the seclusion of complementary sequences in the secondary structure of RNA (Bravo et al., 2018).

σNS does not display ATPase activity, therefore, together with non-specific ssRNA binding it exhibits characteristics of RNA chaperone (Borodavka et al., 2015). As an RNA-binding protein, it stabilizes and increases viral RNA half-life (Zamora et al., 2018).
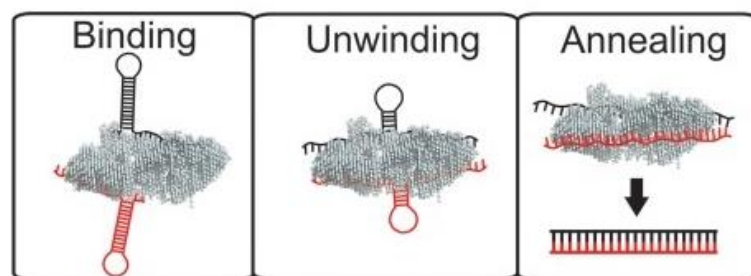
Figure 5: Scheme of σNS from ARV acting as an RNA chaperone (Borodavka et al., 2015).

It has been reported by Xie et al. (2016) that σNS can also activate phosphatidylinositol 3kinase-dependent Akt signalling pathway which contributes to virus-induced inflammation and anti-apoptotic response.

2 Goals

Despite the importance of σNS during the virus assembly, its structure remains unknown. As previous attempts to crystallize the full-length protein yielded non-diffracting crystals, limited proteolysis with elastase was performed to generate smaller more stable fragments. Elastase has low specificity and cleaves at small, hydrophobic residues, therefore, it should cleave the exposed, flexible regions. The proteolysis yielded two fragments of size 15 kDa and 25 kDa potentially lacking the disordered or flexible regions. These regions may have been the cause of non-diffracting crystals (Jack Bravo, manuscript in preparation). The aim of this project was to clone, produce and crystallize the N-terminal domain of σNS.

3 Material and methods

3.1 Material

The gene encoding σNS derived from Avian orthoreovirus strain 1733 cloned into pET 15 was provided by Roman Tůma, University of Leeds.

3.2 Design of PCR primers

Amplification of the σNS N-terminal domain was done using gene specific primers designed for each type of vector separately. All primers were synthetized by Generi Biotech, s.r.o. The list of all primers used in this study is presented in Table 2.

Table 2: Specific primers used for amplification of individual fragments.

Primer name	Sequence	Cloning vector	Restriction enzyme
sNS_wh_pASK37_F	ATGGTA**GGTCTC**AGCGCGACAACACCGTGCGTGTT	pASK-IBA37 plus	BsaI
sNS_Nterm_ pASK37_R	ATGGTA**GGTCTC**ATATCTTAAATGCGAGAAATAGCATTTGCGATT	pASK-IBA37 plus	BsaI
sNS_Nterm_wo11AA_F	ATGGTA**GGTCTC**AGCGCAACACATCCGGCGCA	pASK-IBA37 plus	BsaI
sNS_Nterm_pASK33_F	ATGGTA**GGTCTC**AAATGGACAACACCGTGCGTGTT	pASK-IBA33 plus	BsaI
sNS_Nterm_pASK33_R	ATGGTA**GGTCTC**AGCGCTAATGCGAGAAATAGCATTTGCGAT T	pASK-IBA33 plus	BsaI
sNS_Nterm_ORF1_F	ATGGTA**GAATTC**GGACAACACCGTGCGTGTT	pET Duet 1	EcoRI
sNS_Nterm_ORF1_R	ATGGTA**GCGGCCGC**TTAAATGCGAGAAATAGCATTTGCGATT	pET Duet 1	NotI
sNS_Cterm_ORF2_F	ATGGTA**CATATG**CCTATTGACTACCTCGCTGCT	pET Duet 1	NdeI
sNS_Cterm_ORF2_R	ATGGTA**CTCGAG**CGCCATCCTAGCTGGAGA	pET Duet 1	XhoI
sNS_wh_SUMO_F	CACAGAGAACAGATTGGTGGAGACAACACCGTGCGTGTT	pET SUMO	-
sNS_Nterm_SUMO_R	TGTCTCCTGAGTTCTAGAGTACTTTATTAAATGCGAGAAATAGCATTTGCGAT T	pET SUMO	-

Note: (1) The restriction site for each enzyme is highlighted in bold and underlined.

(2) Primers for SUMO vector do not contain any restriction site since the cloning was done using Gibson assembly method.

3.3 PCR amplification

3.3.1 Gradient PCR

Each pair of designed primers was tested for optimal annealing temperature by gradient PCR experiment. The primers for amplification of N-terminal fragment for cloning into pASK-IBA37plus and the N-terminal fragment without the first 11 amino acids to pASK-IBA37plus were tested using TaKaRa Ex Taq DNA polymerase (TaKaRa Bio). The PCR reaction composed of initial denaturation at 94°C for 1 min and 30 cycles of: denaturation at 98°C for 10 sec, annealing of primers at gradient from 50-60°C for 30 sec and elongation at 72°C for 1 min. Concentrations and amounts of PCR components in the reaction mixture is shown in Table 3.

The primers designed for cloning into pASK-IBA33plus, pET Duet 1 and pET SUMO vectors were tested using PPP Master Mix (Top-Bio), which contained a suitable buffer, dNTPs and Taq DNA polymerase, following the protocol from Top-Bio. Composition of this PCR reaction is given in Table 3.

The results were visualized on 1% agarose gel prepared by dissolving agarose (SERVA Electrophoresis) in 1x TAE buffer (prepared from 50x stock solution: 242 g Tris, 57.1 mL acetic acid, and 100 mL 0.5M EDTA in H_2O) and stained with Serva DNA Stain Clear G (SERVA Electrophoresis). Based on the results the optimal annealing temperature for all of the primer pairs was set at 56°C. All PCR reactions were conducted in the Biometra TOne Thermal Cycler (Analytik Jena).

3.3.2 PCR using Q5 High-Fidelity DNA Polymerase

The fragment for each cloning was amplified utilizing Q5 High-Fidelity DNA Polymerase (New England Biolabs). A standard PCR profile based on following steps was used for amplification: initial denaturation at 98°C for 30 sec followed by 30 cycles of 3 recurrent steps (denaturation at 98°C for 10 sec, primer annealing at 56°C for 15 sec, and elongation at 72°C for 30 sec) and finished with final extension at 72°C for 2 min. The mixture was kept at 16°C after the process was done. Table 3 shows the concentrations and the amounts of all components for the reaction.

Table 3: Composition of reaction mixtures for all PCR reactions.

	Ex Taq [μL]	Q5 [μL]	PPP Master Mix [μL]
Buffer (10x Ex Taq Buffer/5x Q5 Reaction Buffer)	2	10	
dNTPs (10 mM)	0.4	1	
DNA polymerase (TaKaRa Ex Taq (5U/μL)/ Q5 High-Fidelity DNA Polymerase (2U/μL)	0.1	0.5	12,5
Primers (forward/reverse) (10 μM)	2/2	2.5/2.5	2/2
5X Q5 High GC Enhancer	-	10	-
Template DNA	0.5	1	0.5
ddH$_2$O	13	22.5	8
Total amount	20	50	25

The PCR amplicon was purified from the reaction mixture with NucleoSpin Gel and PCR Clean-Up (Macherey-Nagel) according to the PCR clean-up protocol except it was eluted in water instead of elution buffer, and the DNA concentration was measured with NanoPhotometer Pearl (Implen).

3.4 Restriction enzyme digestion

Both DNA fragments and cloning vectors were digested with respective enzymes (all purchased from New England Biolabs) as shown in Table 2. The digestion was carried out for 10 min at 37°C followed by the enzyme deactivation for 20 min at 65°C. The composition of reaction mixtures is presented in Table 4. Digested plasmids were further dephosphorylated by 2.5 μL Shrimp Alkaline Phosphatase (New England Biolabs) which was added into the original reaction mixture. The mixture was incubated at 37°C for 30 min and heat inactivated at 65°C for 5 min.

The successful digestion of cloning vectors was verified on 1% agarose gel stained with Serva DNA Stain Clear G (SERVA Electrophoresis).

Table 4: Set up of digestion reaction mixtures.

Component	50 μL reaction
DNA	1 μg
10X CutSmart Buffer	5 μL
BsaI-HF/EcoRI/NotI/NdeI/XhoI	1/0.2/1/1/1 μL
ddH$_2$O	up to 50 μL

3.5 Ligation

3.5.1 Ligation using Instant Sticky-End Ligase Master Mix

Digested fragments and vectors were ligated using Instant Sticky-End Ligase Master Mix (New England Biolabs). The molar ratio of the DNA fragment to the vector was set to 3:1, and the precise calculations were done with NEBio Ligation Calculator (https://nebiocalculator.neb.com/#!/ligation). DNA insert and vector were combined with nuclease-free water based on the calculations to the final volume of 5 µL and mixed with 5 µL of Instant Sticky-End Ligase Master Mix.

3.5.2 Ligation using T4 ligase

For the ligation of the C-terminal domain of σNS to the pET Duet 1 vector, which already contained cloned N-terminal domain of σNS, T4 ligase (Invitrogen by Thermo Fischer Scientific) was used following the given protocol by Invitrogen – for further details see Table 5. The reaction took place at 14°C for 24h. For the purpose of optimal transformation, the ligation mixture was diluted 5-fold before addition to the competent cells.

Table 5: Reaction setup of the ligation reaction.

5x Ligase Reaction Buffer	4 µL
Insert:Vector Molar Ratio	3:1
Vector	4 µL (45.2 fmol)
Insert	0.6 µL (132.7 fmol)
T4 DNA Ligase (5U/µL)	0.2 µL
ddH$_2$O	up to 20 µL

3.6 Gibson Assembly

Gibson Assembly provides a way of seamless DNA assembly, which was used to insert the N-terminal domain of σNS to pET SUMO vector. Linearized pET SUMO vector and the gene fragment with specific overhangs were generated by PCR, isolated with NucleoSpin Gel and PCR Clean-Up (Macherey-Nagel) and measured. Both amplicons were then mixed in of 1:2 molar ratio and the volume of the mixture was adjusted with nuclease-free water to 10 µL, which was followed by the addition of 10 µL of NEBuilder HiFi DNA Assembly Master Mix

(New England Biolabs). The mixture was incubated in a thermocycler at 50°C for 15 min and placed on the ice prior further work.

3.7 Transformation to TOP10 Competent *E. coli* cells

One Shot TOP10 Chemically Competent *E. coli* cells (Invitrogen by Thermo Fischer Scientific) were used for transformation of the ligation products. Vials with the competent cells were thawed on the ice for about half an hour. Care was taken to avoid stressing the cells as much as possible. 1-5 µL of the ligated DNA product was added to the cells, depending on each sample, and mixed gently by finger flicking. The mixture was incubated on ice for 30 min, and subsequently heat-shocked in a water bath at 42°C for the 40 sec. After the heat-shock, cells were placed on the ice for 2 min, and 250 µL of pre-warmed S.O.C. medium (New England Biolabs) was added to each vial. Vials were shaken at 37°C for 1 hour at 220 rpm in a horizontal shaker. At the end, 200 µL and 50 µL from the transformation reaction were separately spread on pre-warmed plates containing appropriate antibiotics and incubated overnight at 37°C.

3.8 Colony PCR

Colony PCR was carried out to validate the presence of inserted plasmids in TOP10 *E. coli* cells. Individual colonies were carefully picked one by one with a sterile pipet tip, and each one was placed in 20 µL of nuclease-free water. The tip with the colony was used to stir the mixture well, so the colony gets into water, and then the remnants of the mixture on the tip were placed on a new plate marked with numbers for each of the colonies. The PCR reaction mixture was prepared according to the Top-Bio PPP Master Mix protocol with 1 µL of colony template. The thermocycler was set on following steps: initial denaturation at 94°C for 6 min followed by repetition of 3 steps (denaturation at 94°C for 15 sec, annealing of primers at 56°C for 15 sec and elongation at 72°C for 1 min) for 35 times and finished with final elongation step at 72°C for 7 min.

The presence or absence of the cloned DNA fragment in the cells was verified with gel electrophoresis on 1% agarose gel together with a positive control containing whole σNS template that was used in the beginning as a template for PCR amplification.

3.9 Isolation of plasmid and sequence verification

Positive clones from the colony PCR were placed into 4 mL LB media with appropriate antibiotic diluted to final concentration of 0.1 mg/mL for ampicillin, 0.05 mg/mL for kanamycin and incubated at 37°C overnight. Next day, the plasmids were isolated utilizing Nucleo-Spin Plasmid kit (Macherey-Nagel) following Low copy plasmid isolation protocol with elution into water instead of elution buffer. The concentration of purified plasmids was measured using NanoPhotometer Pearl (Implen), and the samples were sent for sequencing together with respective sequencing primers to the SEQme company.

3.10 Transformation to BL21(DE3) and Rosetta-gami 2(DE3) Competent *E. coli* cells

Plasmids with confirmed correct sequence of σNS were used to transform the BL21(DE3) (Invitrogen) and Rosetta-gami 2(DE3) (Novagen) competent *E. coli* cells used for the protein production. The transformation was carried out according to the protocol mentioned earlier for the transformation of TOP10 cells with one exception – only 1 μL of the purified plasmid was added to the thawed cells.

3.11 Pilot expression

In order to optimize the production of the recombinant protein, 10 mL of LB media (or TB media), supplemented with the appropriate antibiotic diluted to final concentration of 0.1 mg/mL for ampicillin, or 0.05 mg/mL for kanamycin, was inoculated with the single colony of BL21(DE3) cells containing the plasmid with the gene of interest (Table 6). This culture was incubated overnight at 37°C and 220 rpm in a horizontal shaker. Next day, 0.5 mL of the overnight culture was added to the 10 mL of fresh LB medium containing appropriate antibiotic and incubated at 37°C and 220 rpm until the OD_{600} reached 0.6-0.8. Once the bacterial culture reached the OD level, protein production was induced with respective inducer at the concentration presented in Table 6. Exceptions were the cases when the effect of the amount of inducer was tested, then 2-fold to 10-fold amount of inducer required for induction was added. Uninduced culture was kept at the same conditions as a negative control. Afterwards, all cultures were incubated at 37°C, 30°C, and 18°C, and 1 mL sample was taken from both induced and uninduced cultures at specific time points. The samples were centrifuged at maximum speed in a microcentrifuge for 30 sec and pelleted cells were frozen at -20°C.

Table 6: Basic features of expression in each vector.

Vector	Antibiotic resistance	Inducer	Required conc. of the inducer
pASK-IBA37plus	Ampicillin	anhydrotetracycline	200 µg/1 L
pASK-IBA33plus	Ampicillin	anhydrotetracycline	200 µg/1 L
pET Duet 1	Ampicillin	IPTG	1mM
pET SUMO	Kanamycin	IPTG	1mM

3.12 SDS-PAGE

The harvested cells from pilot expression experiment were thawed and resuspended in 500 µL of lysis buffer (50mM KH_2PO_4, 400mM NaCl, 100mM KCl, 10% glycerol, 0.5% TritonX-100, 10mM imidazole, pH = 7.8). The samples were lysed by consecutive freezing in liquid nitrogen and thawing at 42°C in a heating block 3 times. Then the samples were centrifuged at maximum speed at 4°C for 1 min to pellet the insoluble protein fraction. The supernatant was separated from the pellet, transferred to a fresh tube and stored on ice prior to the addition of SDS-PAGE sample buffer. Insoluble proteins were mixed with 500 µL of 1X SDS-PAGE sample buffer in 1X PBS (prepared from 4X Laemmli sample buffer stock: 20mL of 0.5M Tris (pH = 6.8), 4g SDS, 20 mL glycerol, 0.1 g of bromphenol blue, H_2O up to 45mL; SDS-PAGE sample buffer prepared by addition of 100 µL β-mercaptoethanol to 900 µL of 4X Laemmli sample buffer), and boiled for 5 min. Soluble proteins in the supernatant were combined with 125 µL of 4X SDS-PAGE sample buffer and boiled for 5 min.

For the purpose of SDS-PAGE, 12.5% and later 15% gels were cast. The detailed gel composition is described in Table 7. To each gel, 10 µL of PageRuler Protein Prestained Ladder (Thermo Fischer Scientific) was added. Samples were loaded to the gels in a consecutive manner, uninduced samples followed by induced ones for each time point. As soon as SDS-PAGE was done, gels were disposed of the stacking gel, and resolving gel was stained in Coomassie R-250 Brilliant Blue (CBB) (0.625g Coomassie R-250 Brilliant Blue, 112.5 mL methanol, 112.5 mL ddH$_2$O, 25 mL acetic acid) for 1 hour. De-staining solution (60% H_2O, 30% methanol, 10% acetic acid) was used to wash off the remnants of CBB from the gels.

Table 7: Composition of SDS-PAGE gels (A – resolving gel, B – stacking gel).

A

	12.5%	15%
30% Acrylamide (37.5:1)	2.08 mL	2.5 mL
H_2O	1.57 mL	1.15 mL
1.5M Tris (pH = 8.8)	1.25 mL	1.25 mL
10% SDS	50 μL	50 μL
10% APS	50 μL	50 μL
TEMED	5 μL	5 μL

B

	4%
30% Acrylamide (37.5:1)	340 μL
H_2O	1.36 mL
0.5M Tris (pH = 6.8)	250 μL
10% SDS	20 μL
10% APS	20 μL
TEMED	2 μL

3.13 Western Blot

The SDS-PAGE gel was run as described above, but instead CBB staining, the gel was blotted on a membrane using precut blotting Trans-Blot Turbo Transfer pack (Bio-Rad) following the Trans-Blot Turbo Transfer Pack Quick Start Guide. The assembled sandwich was wetted with SDS-PAGE transfer buffer (10x SDS-PAGE transfer buffer stock: 58.15g Tris, 29.3g glycine, 3.75g SDS; 1L of working solution is prepared by dilution to 1x and addition of 100 mL methanol). The blotting run on 25V, 1.0 A for 30 min using Trans Blot Turbo (Biorad). Once the transfer was finished, the membrane was washed 3x10 min in TBS-T (10x TBS stock buffer: 60.5g Tris, 87.66g NaCl, pH = 7.6; TBS-T prepared by dilution of stock solution to 1x and addition of 1 mL Tween 20) on a rocking platform. The membrane was blocked in blocking solution prepared from 10x Blocking Reagent Buffer (Qiagen), Blocking Reagent Powder (Qiagen), and Tween 20 according to the manufacturer's instructions. The membrane was blocked for 1 hour at room temperature on a rocking platform. Next, the membrane was washed 3x10 min in TBS-T on a rocking platform and incubated for 1 hour at room temperature on the rocking platform with primary antibody Penta His HRP Conjugate (Qiagen) diluted 1:1000 in a blocking solution. Afterwards, the membrane was washed 2x10 min in TBS-T and 10 min in TBS. The detection was done using Opti-4CN Substrate kit (Bio-Rad) until the bands were clearly visible and the reaction was stopped by washing in H_2O.

3.14 Large-scale protein production

Once the conditions for the protein production were verified by SDS-PAGE and western blot, four 35 mL overnight cultures with an appropriate antibiotic (Table 6), were prepared for large-scale protein production. Samples were incubated at 37°C and 220 rpm overnight. Next day, each overnight culture was transferred to separate 500 mL of LB medium (in total 2L of media) with antibiotics and incubated in a horizontal shaker until OD_{600} reached values between 0.6 to 0.8. At that point, cultures were induced with the appropriate inducer and incubated at specific temperature and time, which were previously shown to gain the highest protein yield. Finally, the cultures were spun down in a centrifuge at 10°C and 4000 rpm for 20 min. The supernatant was discarded and the harvested cells were frozen at -80°C.

3.15 Protein purification

Harvested cells were thawed on ice and resuspended in 10 mL of resuspension buffer (20mM Tris, 400mM NaCl, 7.5mM $MgCl_2$, pH = 8.5, supplemented with protease inhibitor cocktail (SIGMAFAST Protease Inhibitor Cocktail tablets, EDTA-free by Sigma-Aldrich)). Cell lysis was performed by passing the cell mixture twice through the Cell disruptor Stansted Press (Stansted Fluid Power). The lysate was centrifuged in an ultracentrifuge at 4°C, 25 000 rpm for 1 hour.

3.15.1 Immobilized metal affinity chromatography

ÄKTA Pure system (GE Healthcare), fitted with the HisTrap HP 5 mL column (GE Healthcare), was washed with double-distilled and de-gased H_2O for 25 min (5x volume of the column) with a flow rate of 1 mL per min. Afterwards, the system was equilibrated in equilibration buffer (20mM Tris, 50mM NaCl, pH = 8.5) for 25 min, elution buffer (20mM Tris, 50mM NaCl, 1M imidazole, pH = 8.5) for 25 min, and again with equilibration buffer. The sample – supernatant from the centrifugation – was loaded on the column with the sample pump, at a flow rate of 1 mL per min. The flow-through fraction was collected when UV started to rise. When the sample was completely loaded, the system was equilibrated in the equilibration buffer until the UV dropped to the baseline. Next, the equilibration buffer was changed to 5% elution buffer, and the fraction was collected. Subsequently, the gradient of elution buffer was set to increase to 100% in 20 min. 1 mL fractions were collected once the UV started to rise using the fraction collector. Following the purification, the system was

equilibrated with equilibration buffer, washed with H_2O and kept in 20% ethanol. Collected fractions were stored in the fridge at 4°C. Purified fractions were analysed on an SDS-PAGE gel and those containing the protein of interest were pooled prior next purification step.

3.15.2 Anion exchange chromatography

Anion exchange chromatography was done using ÄKTA Pure system (GE Healthcare) fitted with the HiTrap Q HP column (GE Healthcare). The system was equilibrated as described above with the exception of equilibration buffer (20mM Tris, 50mM NaCl, pH = 8.5) and elution buffer (20mM Tris, 1M NaCl, pH = 8.5). The gradient was set directly from 0-100% of elution buffer in 20 mins, and fractions were collected the same way as described above. Purification result was checked using SDS-PAGE of the fractions. Fractions were kept in the fridge and eventually pooled based on the impurities still present in the fractions.

3.15.3 Size exclusion chromatography

ÄKTA Pure system (GE Healthcare) fitted with Superdex 200 Increase 10/350 GL column (GE Healthcare) was washed with H_2O and equilibration buffer (20mM Tris, 50mM NaCl, pH = 8.5). The sample was loaded manually, 0.5 mL for one round of purification, and flow rate was set at 1 mL per min. ÄKTA Pure system was connected to multiangle laser light scattering *Dawn8+* (Wyatt Technology) and refractive index detector Optilab T-rEX (Wyatt Technology) to reckon molecular weight and purity of the proteins. 0.5 mL fractions were collected as soon as the protein peak appeared. Finally, the purity of the protein was validated using SDS-PAGE. The concentration of protein was measured with NanoPhotometer Pearl (Implen) from the fractions with pure protein.

3.16 Crystallization

3.16.1 Pre-crystallization test

Pre-crystallization test was performed utilizing PCT kit by Hampton Research, and all four of the kit supplied reagents (A1, A2, B1, B2) were tested to evaluate the optimal protein concentration. 40 μL of the reagent was put into each reservoir. From the reservoir, 1 μL of reagent was mixed with 1 μL of the protein. The plate was sealed and left on the table for 30 min.

As there was no precipitation, samples were concentrated with Amicon Ultra – 0.5 Centrifugal Filter Devices (10K) (Merck Millipore). 500 µL of the sample was loaded and centrifuged at 4°C for 15 min at 14 000 rpm. Filter with enriched fraction was turned around, placed in a new tube and spun for 4 min at 1000 rpm. Then the PCT was performed again, with smaller amounts – 20 µL of the reagent in the reservoir and 0.5/0.5 µL drop.

3.16.2 Crystallization screening

For crystallization, 3 plates (SwissDrop2D) with different screens were prepared. For detailed information see Table 8.

Table 8: Preparation of crystallization screening.

Name of the screen	1:1 ratio	1:2 ratio	Reservoir volume
Morpheus 2 (Molecular Dimensions)	0.2/0.2 µL	0.1/0.2 µL	50 µL
PEGRx (Hampton Research)	0.2/0.2 µL	0.1/0.2 µL	50 µL
MemGold2 (Molecular Dimensions)	0.4/0.4 µL	0.2/0.4 µL	50 µL

Note: (1) Ratios are given as protein/reagent.

Crystals which grew in the drops were dyed with Izit Crystal Dye (Hampton Research) in order to distinguish protein crystals from the salt crystals. Izit Crystal Dye was diluted 1:1 in H_2O and each drop was dyed with 0.1 µL of the diluted Izit.

4 Results

Primers amplifying the N-terminal domain of σNS were designed based the results from full-length σNS limited proteolysis with elastase (performed by Jack Bravo: Astbury Centre for Structural Molecular Biology, University in Leeds, UK). Amplified N-terminal fragment of approximately 432 bp was cloned into the various expression vectors and its production in *E. coli* expression system was investigated.

4.1 Cloning and production of the N-terminal domain of σNS in pASK-IBA37plus

The estimated annealing temperature of both primers based on the computed analysis was 56°C. To identify the best annealing temperature the gradient PCR with the temperature range spanning from 45-65°C was performed and the optimal annealing temperature was set at 56°C (data not shown).

Next, the N-terminal domain of σNS was amplified with Q5 High-Fidelity Polymerase which possesses $3´ \rightarrow 5´$ exonuclease activity to ensure ultra-low error rate. Presence of the 432 bp DNA insert length was verified by 1% agarose gel electrophoresis (Figure 6). PCR amplicon was purified from the gel yielding the DNA concentration of 197 ng/µL. Both the DNA insert and pASK-IBA37plus plasmid were digested by BsaI-HF restriction enzyme. The digested plasmid (size of empty pASK-IBA37plus: 3270 bp) was visualized on 1% agarose gel to verify successful cleavage. We detected clear band for linearized plasmid at approximately 3000 bp that did not migrate as far as the supercoiled, non-linearized plasmid (Figure 6). Since the digestion of DNA insert involved only a couple of bases, the difference in length was negligible and gel electrophoresis of the cleaved DNA insert was not done.
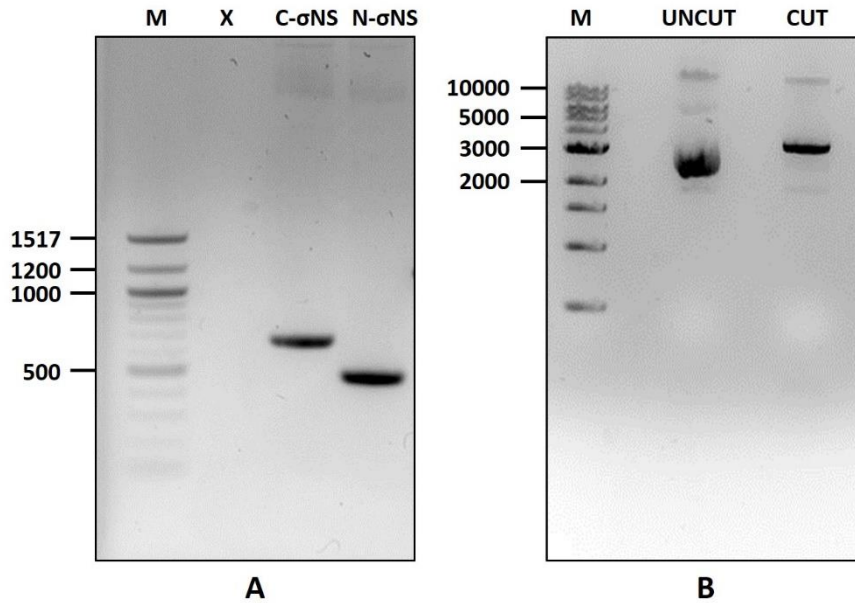
Figure 6 - **Gel electrophoresis of PCR product (A) and pASK-IBA37plus digestion (B)**. M (A): 100 bp DNA Ladder (New England Biolabs). M (B): 1 kb DNA Ladder (New England Biolabs). X – negative control containing no DNA template.

Partial gene fragment encoding the N-terminal domain of σNS was inserted into pASK-IBA37plus by ligation using Instant Sticky-end MM as described in 3.5.1, and the *E. coli* TOP10 competent cells were transformed by recombinant plasmid and grown overnight. Next day we analysed 25 colonies by colony PCR to verify the presence of inserted gene fragment. In total 23 out of 25 colonies resulted in a band of size about 500 bp thus confirming the presence of inserted gene fragment in the plasmid (Figure 7).
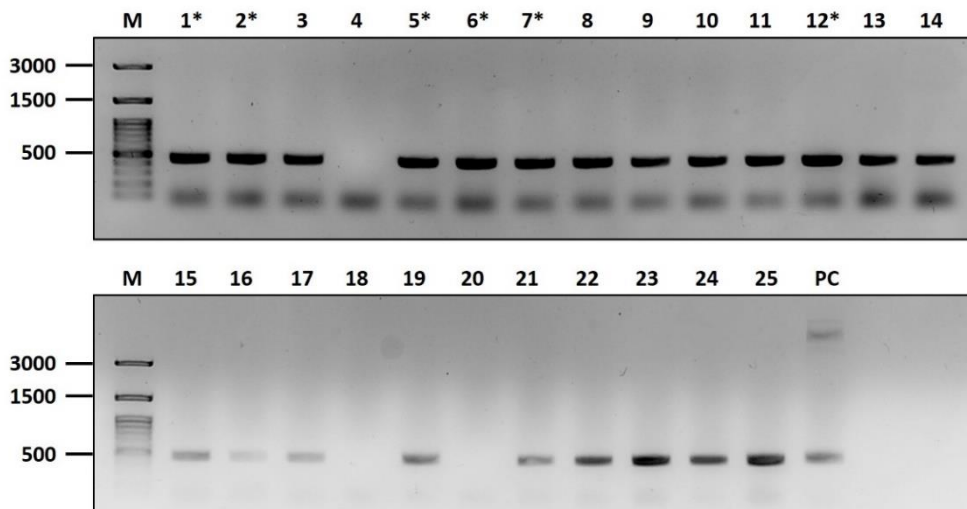

Figure 7: **Gel electrophoresis of colony PCR samples containing the pASK-IBA37plus plasmid with DNA insert.** M - 100 bp DNA Ladder H3 RTU (Nippon Genetics). PC – positive control, the same reaction mixture with added DNA template. Samples marked with an asterisk were sent for sequencing.

Colonies with incorporated recombinant plasmid were grown overnight in the liquid culture containing 100 μg/mL ampicillin. Plasmids were isolated, measured and sent for sequencing. Sequencing results confirmed the presence of the correct DNA sequence of the gene of interest without any error. Plasmid number 12 with a concentration of 79.3 ng/μL was further used to transform BL21(DE3) competent cells for protein production.

Prior the large-scale protein production, the pilot expression was performed. *E. coli* BL21(DE3) competent cells were grown at 37°C until they reached the OD_{600} of 0.6. The protein production was then induced by addition of anhydrotetracycline of final concentration 0.2 μg/mL and the culture was incubated at 37°C, 30°C or, 18°C. Protein production endured for 6 hours (37°C, 30°C) or 24 hours (18°C) and an aliquot of 1 mL was taken at specific time intervals. All samples were pelleted by spinning at 11 000 rpm and frozen prior further use. The soluble and insoluble fraction was prepared from all taken samples and analyzed by SDS-PAGE under reducing conditions. The estimated molecular weight of σNS N-terminal domain together with 6xHistidine Tag and factor Xa cleavage site was 17.5 kDa. None of the temperatures (37°C, 30°C, and 18°C) yielded a detectable amount of recombinant protein that could be visualized by CBB. Figure 8 shows an example of an SDS-PAGE analysis of protein expression at 18°C. The resulting gels for 30°C, and 37°C temperatures are not shown.
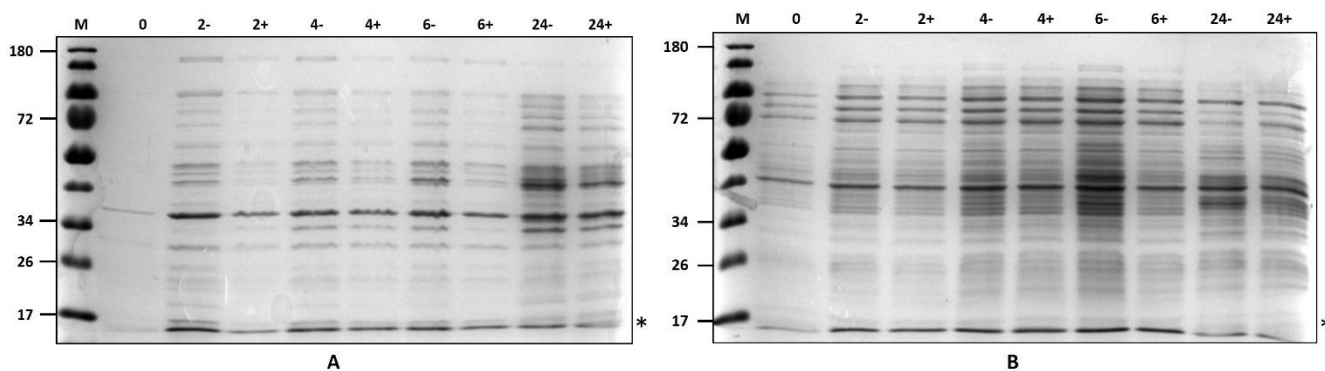

Figure 8: **SDS-PAGE analysis of protein expression at 18°C** – insoluble fraction (A), soluble fraction (B). M – PageRuler Prestained Protein Ladder (Thermo Fischer). **– sign indicates uninduced culture while + sign indicates induced culture**, each number stands for an hour of collection of the sample after the induction. * correspond to the expected area of the N-terminal domain of σNS.

To verify the protein production, the final time points from all tested temperatures were analyzed by Western Blot using monoclonal anti-his antibody. We detected specific signal of the approximate size of 17 kDa in both soluble and insoluble fraction at all temperatures (Figure 9). Western blot result suggests that the recombinant protein is produced by *E. coli*, but its production is quite limited.
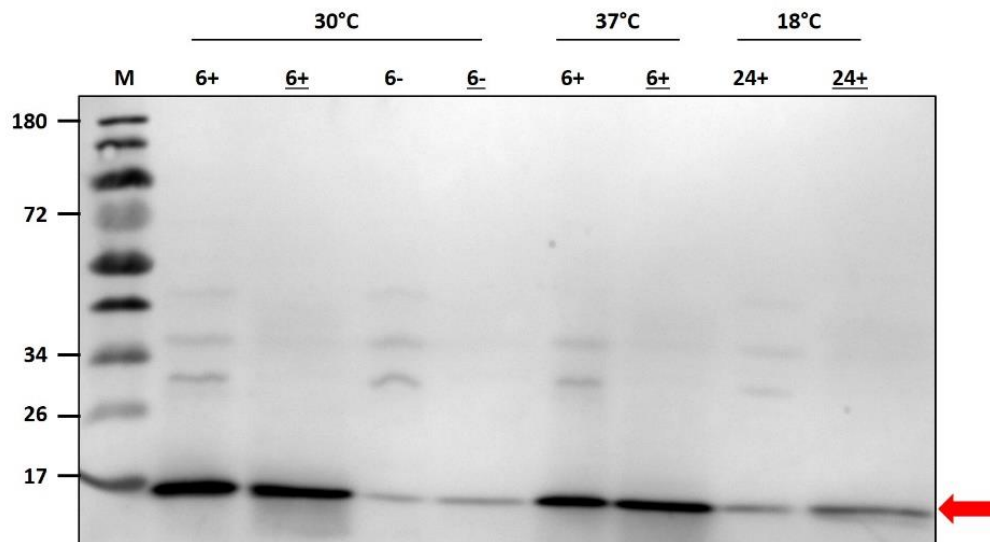
Figure 9: **Western Blotting of final time points of protein expression at different temperatures.** M – PageRuler Prestained Protein Ladder (Thermo Fischer). **− sign indicates uninduced culture while + sign indicates induced culture**, each number stands for an hour of collection of the sample after the induction. Underlined time points represent a soluble fraction, the rest the insoluble fraction.

Because the protein production in *E. coli* BL21(DE3) cells was unsuccessful, we decided to test another genotype of *E. coli* cells Rosetta-gami 2(DE3). These cells in comparison to *E. coli* BL21(DE3) increase the protein production by incorporation of codons rarely used in *E. coli* and enhancing the disulfide bond formation. *E. coli* Rosetta-gami 2(DE3) competent cells were transformed with the same plasmid as *E. coli* BL21(DE3) (number 12 from colony PCR). The cells were grown in nutritionally rich TB medium because the growth of Rosetta-gami 2(DE3) cells in regular LB media was shown to be quite slow in comparison to the growth of *E. coli* BL21(DE3) cells. Recombinant protein production was monitored only at 37°C (the best conditions in BL21 (DE3) cells) to determine if there is a significant change in comparison to *E. coli* BL21(DE3) cells. Resulting SDS-PAGE analysis of soluble and insoluble cell fraction showed no protein production in either of the fractions (Figure 10).
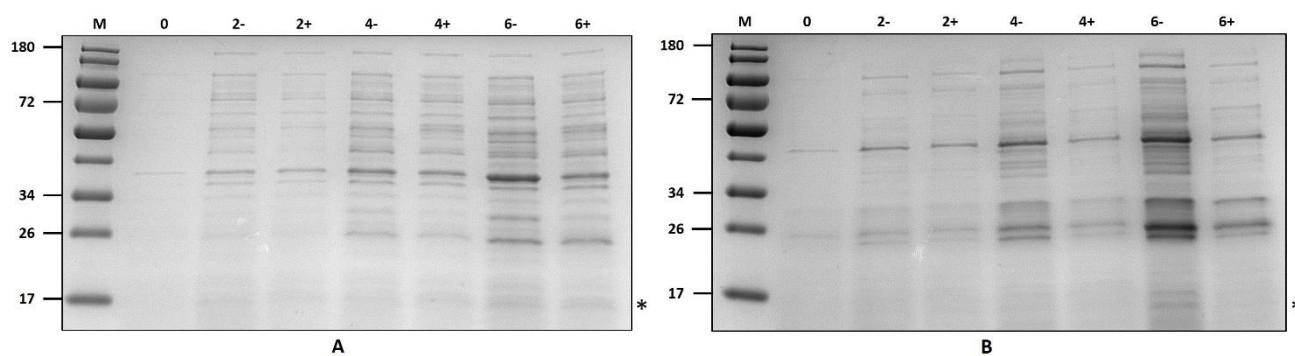
Figure 10: **SDS-PAGE analysis of protein expression in Rosetta-gami 2(DE3) at 37°C** – insoluble fraction (A), soluble fraction (B). M – PageRuler Prestained Protein Ladder (Thermo Fischer). **− sign indicates uninduced culture while + sign indicates induced culture**, each number stands for an hour of collection of the sample after the induction. * correspond to the expected area of the N-terminal domain of σNS protein band.

3.2 Cloning and production of the N-terminal domain of σNS without the first 11 amino acids in pASK-IBA37plus

Because the first 11 amino acids of the N-terminal domain of σNS participate in binding of ssRNA, we designed new primers to trim the first 11 amino acids as illustrated in Figure 11 and cloned the shorter fragment into pASK-IBA37plus.
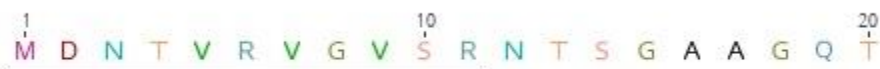

Figure 11: **Scheme of the beginning of the σNS protein sequence.** Underlined letters represent the amino acids that were trimmed from the sequence.

As in the previous case, the annealing temperature of the newly designed forward primer (sNS_Nterm_wo11AA_F) was estimated to be 56°C and the specific annealing temperature was verified by gradient PCR (data not shown). The 400 bp gene fragment was amplified with Q5 High-Fidelity Polymerase and the annealing temperature was set to 56°C. Resulting product was resolved on 1% agarose gel and purified yielding 168 ng/µL of DNA. Both PCR product and pASK-IBA37plus were digested using BsaI-HF enzyme. Plasmid digestion was verified by gel electrophoresis which showed a clear band around 3000 bp representing the linearized vector (data not shown).

Digested plasmid and the gene representing the N-terminal domain of σNS were joined together by ligation to form a recombinant plasmid (3.5.1). *E. coli* TOP10 competent cells were transformed by resulting plasmid and kept in the incubator overnight until the colonies appeared. 15 colonies were picked for analysis by colony PCR. All of the colonies showed to be positive for the presence of the recombinant plasmid containing the gene for the N-terminal domain of σNS as evident from Figure 12. Plasmids chosen for sequencing were propagated, isolated from the overnight culture, and sent for sequencing. Sequencing results confirmed that all samples sent for sequencing contain the recombinant plasmid with the correct sequence of the N-terminal domain of σNS. As sample number 1 was of the highest concentration (34.9 ng/μL), and sequencing chromatogram had good quality it was chosen for transformation to *E. coli* BL21(DE3) and *E. coli* Rosetta-gami 2(DE3).
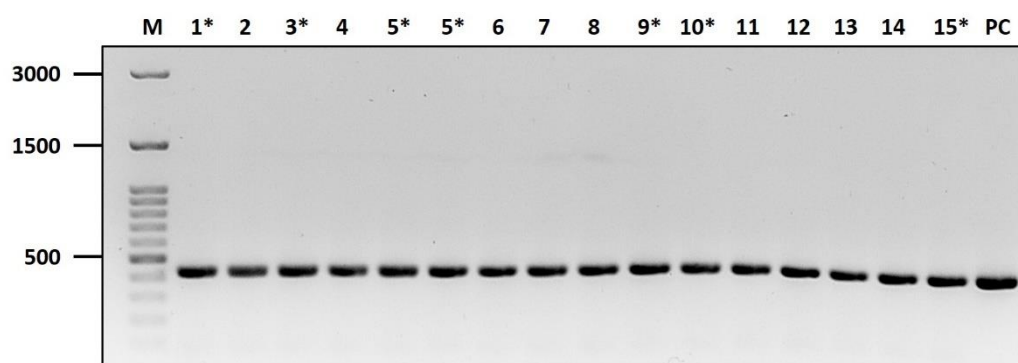

Figure 12: **Agarose gel electrophoresis of colony PCR samples containing the pASK-IBA37plus plasmid with DNA insert.** M - 100 bp DNA Ladder H3 RTU (Nippon Genetics). PC – positive control, the same reaction mixture with added DNA template. Samples marked with an asterisk were sent for sequencing.

Recombinant protein production was tested in both cell types *E. coli* (BL21(DE3) and Rosetta-gami 2(DE3)). Each cell type was investigated at two different temperatures – 37°C and 18°C. The increased amount of inducer in the previous experiment with the full N-terminal domain σNS did not show any difference in protein production compared to the recommended amount of inducer (data not shown). Recombinant protein production in both cells types was induced with 0.2 μg/mL anhydrotetracycline. The estimated molecular weight of the trimmed N-terminal domain σNS is 16.5 kDa. The expected area of the protein band on the gel is marked with an asterisk. Neither of the conditions yielded any recombinant protein as illustrated in Figures 13-14.
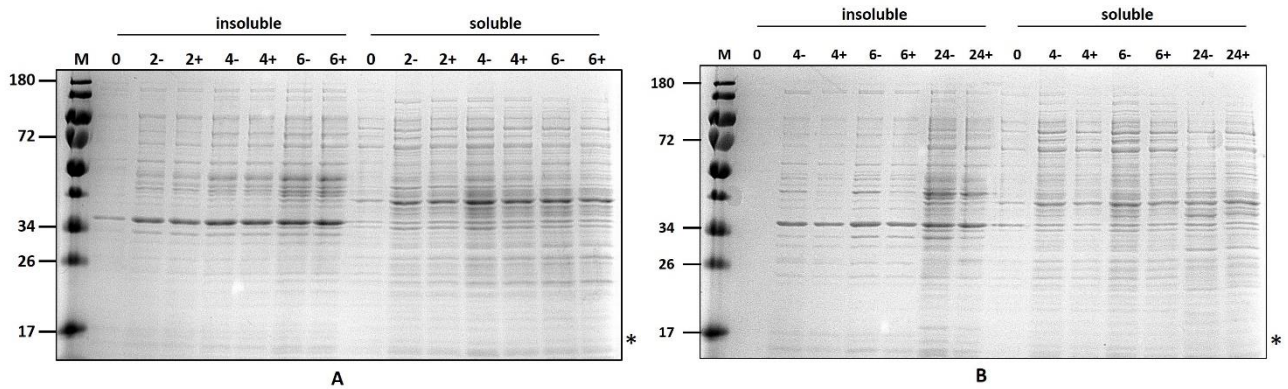
Figure 13: **SDS-PAGE analysis of protein production in BL21(DE3) at 37°C (A) and 18°C (B).** M – PageRuler Prestained Protein Ladder (Thermo Fischer). **– sign indicates uninduced culture while + sign indicates induced culture**, each number stands for an hour of collection of the sample after the induction. * corresponds to the expected area of the N-terminal domain of σNS protein band.
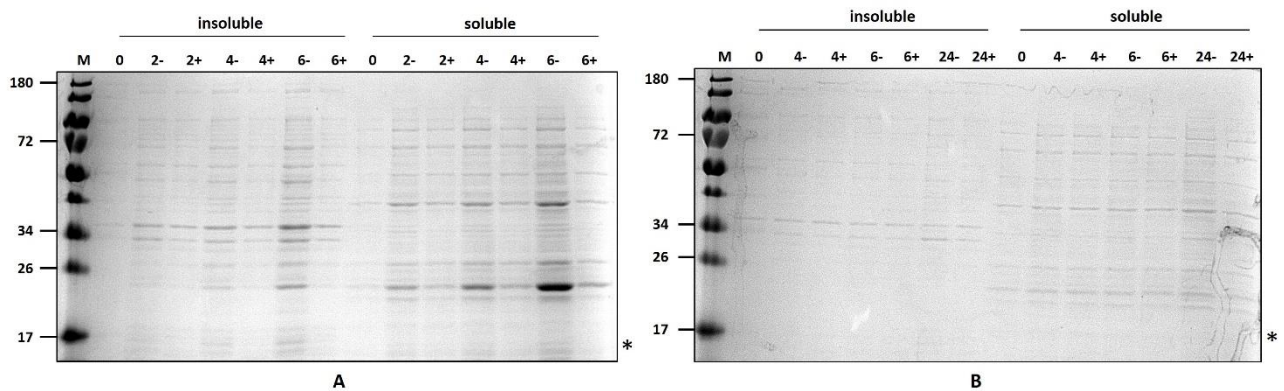

Figure 14: **SDS-PAGE analysis of protein expression in Rosetta-gami 2(DE3) at 37°C (A) and 18°C (B).** M – PageRuler Prestained Protein Ladder (Thermo Fischer). **– sign indicates uninduced culture while + sign indicates induced culture**, each number stands for an hour of collection of the sample after the induction. * corresponds to the expected area of the N-terminal domain of σNS protein band.

Because we did not detect any recombinant protein by CBB, we performed Western Blot using an anti-his antibody to assess the course of the protein production at specific time points. As shown in Figure 15, poorly noticeable bands which were of highest intensity in 2 hours fractions in *E. coli* BL21(DE3) cells and in 6 hours fraction for *E. coli* Rosetta-gami B(DE3) cells appeared below the 17 kDa marker. Results of the Western Blot support our idea that the protein is unstable and undergoes degradation.

27

Figure 15: **Western Blot analysis of protein expression in two types of cells**. M – PageRuler Prestained Protein Ladder (Thermo Fischer). **– sign indicates uninduced culture while + sign indicates induced culture**, each number stands for an hour of collection of the sample after the induction. Underlined time points represent a soluble fraction, the rest the insoluble fraction.

3.3 Cloning and production of the N-terminal domain of σNS in pASK-IBA33plus

As production in pASK-IBA37plus did not yield any protein, another cloning vector containing C-terminal 6xHistidine-tag was tested. Two new primers (sNS_Nterm_pASK33_F, sNS_Nterm_pASK33_R) were designed for the cloning of the N-terminal domain of σNS into pASK-IBA33plus. Both of them were tested for optimal annealing temperature by gradient PCR (data not shown). The annealing temperature for amplification with Q5 High-Fidelity Polymerase was set at 56°C as in previous cloning. After the amplification, PCR amplicon was resolved on 1% agarose gel and purified yielding the concentration of 132 ng/µL. Next both PCR amplicon and the cloning plasmid were digested with BsaI-HF and subsequently ligated by Instant Sticky-end MM to form a recombinant plasmid.

E. coli TOP10 competent cells were transformed by the recombinant plasmid. Since the colony PCR yielded only 4 positive colonies, all of them were propagated, isolated and sent for sequencing (Figure 16).
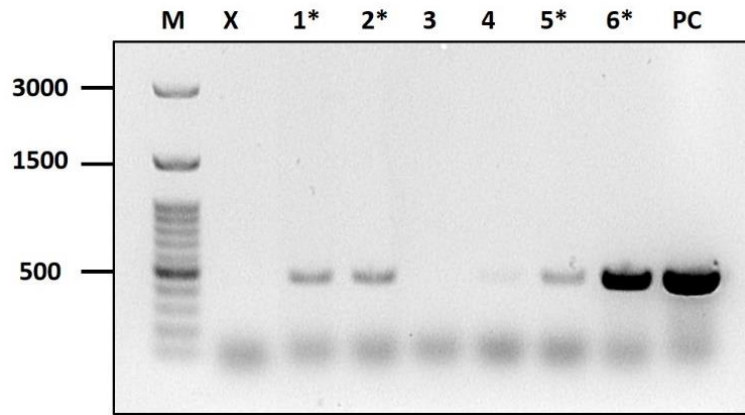
Figure 16: **Gel electrophoresis of colony PCR samples with a pASK-IBA33plus plasmid with DNA insert**. M - 100 bp DNA Ladder H3 RTU (Nippon Genetics). PC – positive control, the same reaction mixture with added DNA template. Samples marked with an asterisk were sent for sequencing.

The only sample which showed the correct sequence was number 6 with the concentration of 29.9 ng/μL, and this one was used for transformation of *E. coli* BL21(DE3) cells. The N-terminal domain of σNS with the C-terminal 6xHistidine-tag has a molecular weight of approximately 17.5 kDa. Recombinant protein production was tested at two temperatures 37°C and 18°C, and both samples were induced with 0.2 μg/mL anhydrotetracycline. As in the previous experiments, SDS-PAGE showed no significant protein production at any of the temperatures tested. Example of the SDS-PAGE analysis of protein production at 37 °C is presented in Figure 17. This result proposes that the location of 6xHistidine-tag has no effect on the level of protein production.
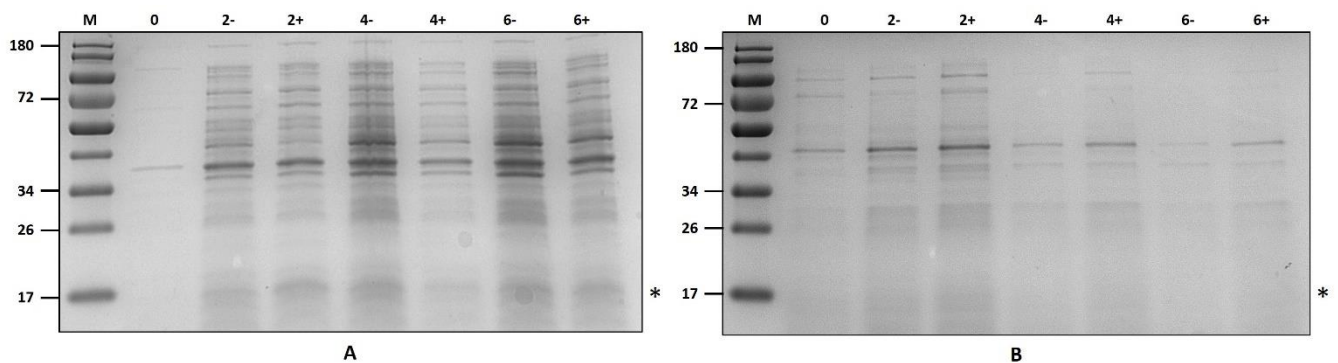

Figure 17: **SDS-PAGE analysis of protein production at 37°C** – insoluble fraction (A), soluble fraction (B). M – PageRuler Prestained Protein Ladder (Thermo Fischer). **– sign indicates uninduced culture** while + **sign indicates induced culture**, each number stands for an hour of collection of the sample after the induction. * corresponds to the expected area of the N-terminal domain of σNS protein band.

We have also investigated the protein production in a TB medium at two temperatures – 37°C and 18°C. However, there was no obvious protein production (data not shown) such as in the LB medium.

3.4 Cloning and production of the N-terminal domain of σNS in pET SUMO

The primers for cloning into pET SUMO (sNS_wh_SUMO_F, sNS_Nterm_SUMO_R) were designed without having any restriction site as the cloning was done by Gibson Assembly method. Primers contained 15-20 bp overlapping ends compatible with the overlaps on the linearized SUMO vector. The testing of these primers by gradient PCR confirmed the estimated annealing temperature of 56°C.

Amplified N-terminal domain of σNS fragment (500 bp) was purified from the 1% agarose gel yielding the concentration of 123 ng/μL. Empty pET SUMO vector was linearized by PCR and purified yielding the concentration of 52 ng/μL. Both DNA fragments were joined together by the addition of NEBuilder HiFi DNA Assembly which contained 5' exonuclease, a polymerase, and a DNA ligase to perform the assembly. *E. coli* TOP10 competent cells were transformed by resulting recombinant plasmid and grown overnight. Although only 3 single colonies grew, all of them were analyzed by colony PCR and showed to be positive (Figure 18).
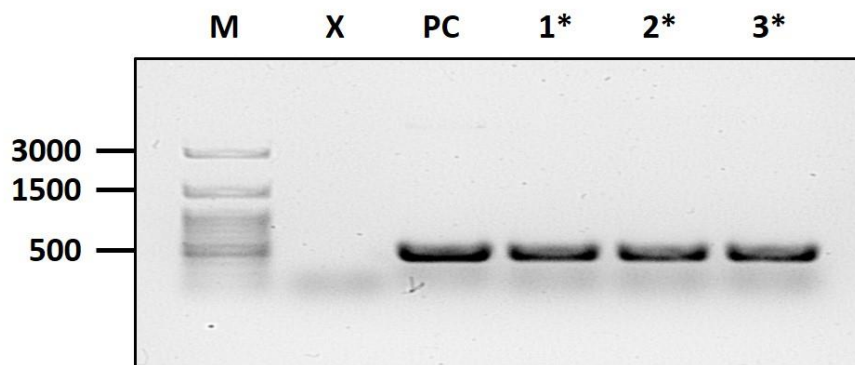

Figure 18: **Agarose gel electrophoresis of colony PCR samples with Champion pET SUMO plasmid with DNA insert.** M - 100 bp DNA Ladder H3 RTU (Nippon Genetics). PC – positive control, the same reaction mixture with added DNA template. Samples marked with an asterisk were sent for sequencing.

For samples number 1 and 2, the chromatograms were of good quality and sequences were correct, therefore sample number 1 with a concentration of 34.4 ng/μL was used for transformation of *E. coli* BL21(DE3).

Recombinant protein production was tested at two different temperatures 18°C and 37°C. Each culture was induced by 1mM IPTG. The molecular weight of the N-terminal domain of σNS fused to the SUMO fusion protein is approximately 29 kDa. After isolation of soluble and insoluble fraction from samples incubated at 37°C the protein production at a molecular weight between 26 to 34 kDa have been monitored. We observed a clear pattern in which induced fractions contain an increasing amount of produced protein as the time of incubation increased. As can be seen from Figure 19, peak production was presumably present at 3 hours post induction and afterwards the protein amount did not increase significantly. The majority of the protein was detected in the insoluble fraction which suggests that the protein may not be folded correctly.
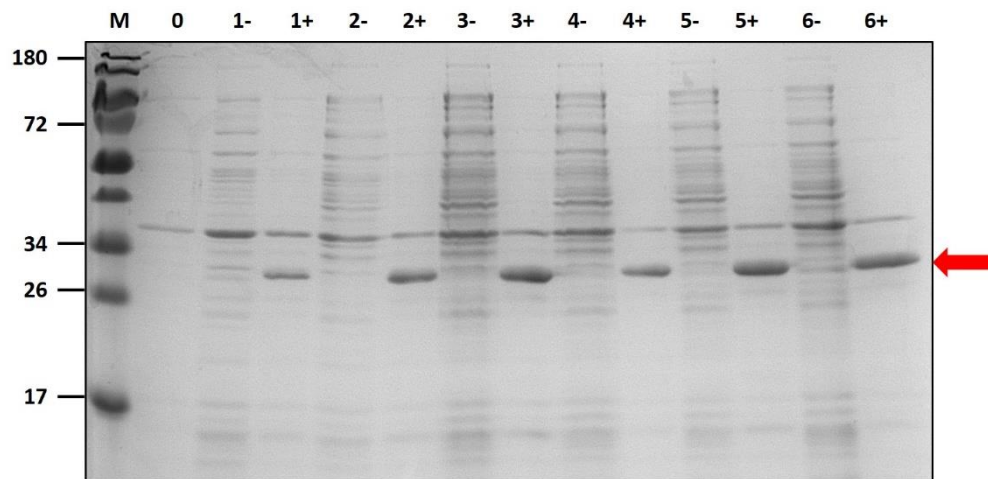

Figure 19: **SDS-PAGE analysis of the insoluble fraction of protein expression at 37°C**. M – PageRuler Prestained Protein Ladder (Thermo Fischer). **− sign indicates uninduced culture** while **+ sign indicates induced culture**, each number stands for an hour of collection of the sample after the induction.

Except for the high protein yield in the insoluble fraction, there was also a slight protein band detected in the soluble fraction in the same region (between 26 to 34 kDa). However, another strong protein band was detected in a region around 17 kDa (Figure 20).
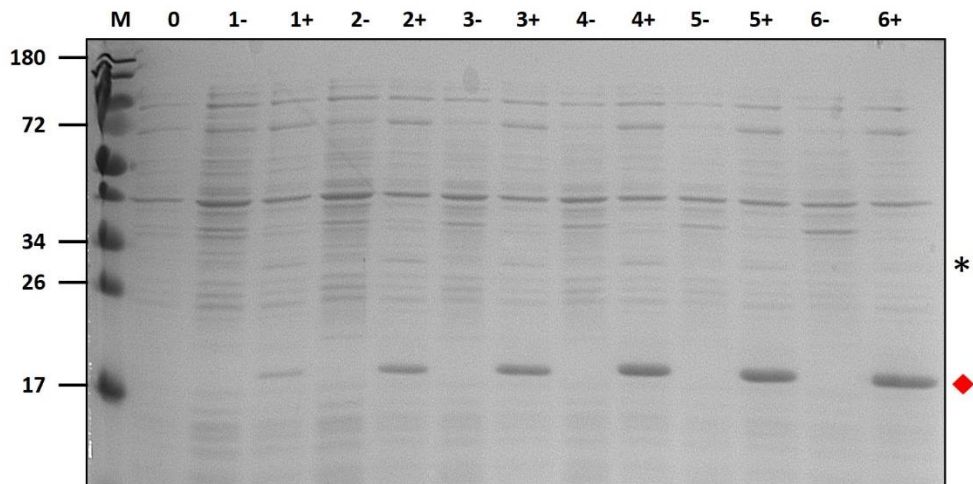
Figure 20: **SDS-PAGE analysis of the soluble fraction of protein expression at 37°C**. M – PageRuler Prestained Protein Ladder (Thermo Fischer). **− sign indicates uninduced culture while + sign indicates induced culture**, each number stands for an hour of collection of the sample after the induction. * corresponds to the expected area of the N-terminal domain of σNS with SUMO fusion protein band.

Both productions at 37°C as well as 18°C showed the protein band around 17 kDa in soluble fraction. However, smaller amount of the recombinant protein was observed at 18°C (Figure 21). No protein was detected in the insoluble fraction.
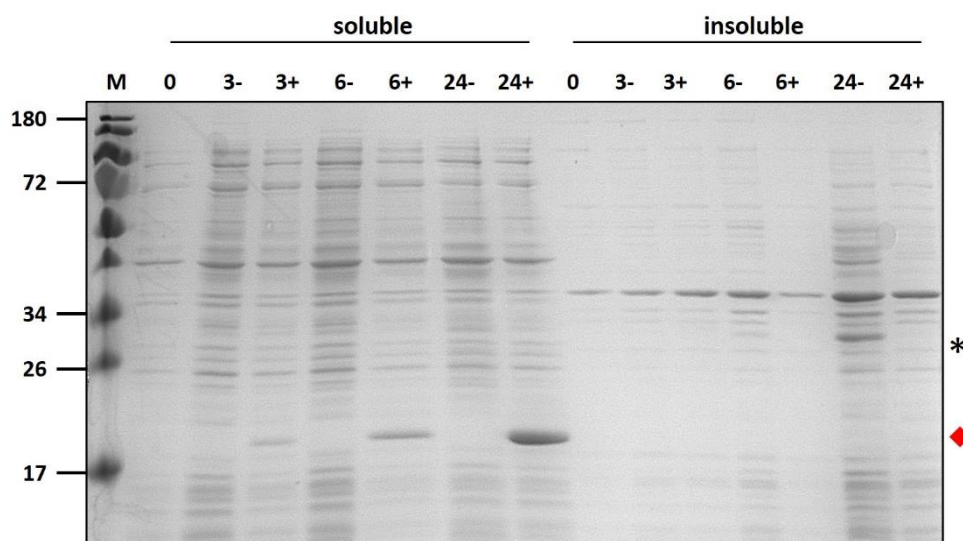
Figure 21: **SDS-PAGE analysis of protein expression at 18°C.** M – PageRuler Prestained Protein Ladder (Thermo Fischer). **− sign indicates uninduced culture while + sign indicates induced culture**, each number stands for an hour of collection of the sample after the induction. * corresponds to the expected area of the N-terminal domain of σNS with SUMO fusion protein band.

As the SDS-PAGE analysis showed a protein band in an unexpected region, we performed Western Blot with a 6xHistidine-tag antibody. We detected a strong signal in the

region between 26 to 34 kDa and above 17 kDa in all induced fractions. The major western blot signals were accompanied by many other various sizes (Figure 22), indicating protein degradation. The 6xHistidine-tag is located at the N-terminal part of the protein, preceding the SUMO fusion protein. As the antibody against His-tag was used, all of the visible bands must contain the His-tag, therefore, the degradation of the recombinant protein occurs from the C-terminal end.

Besides Western blot, the presence of σNS in the shorter fragment was verified by mass spectrometry using our department facility. The fragment contained 28 amino acids of the σNS sequence connected to SUMO fusion protein (data not shown).

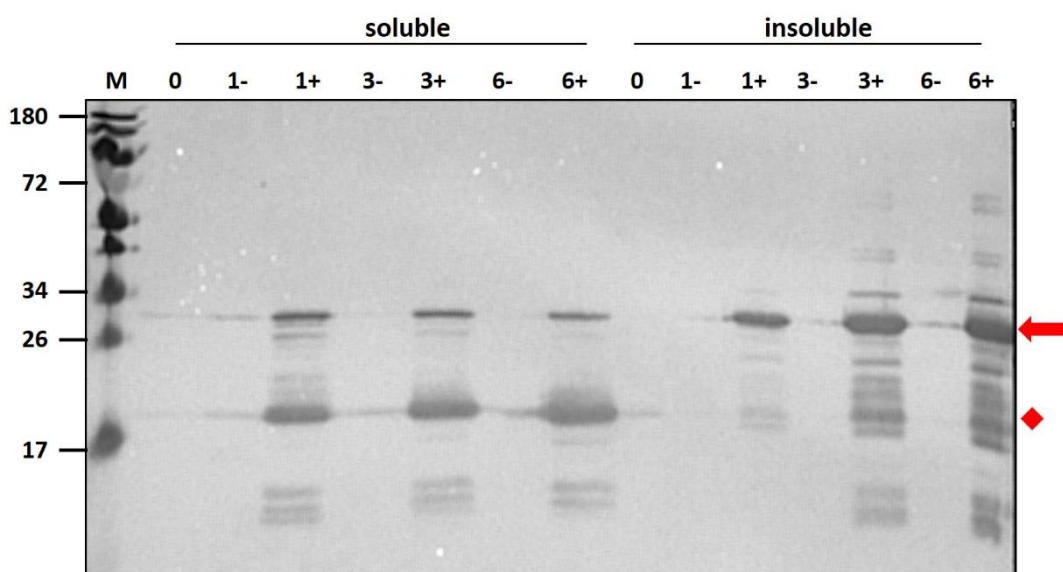Figure 22: Western Blot analysis of protein expression in the soluble and insoluble fraction. M – PageRuler Prestained Protein Ladder (Thermo Fischer). **– sign indicates uninduced culture while + sign indicates induced culture**, each number stands for an hour of collection of the sample after the induction.

Since the Western Blot and mass spectrometry confirmed that the unknown protein in the soluble fraction is a fragment of the N-terminal domain of σNS with SUMO fusion protein, we proceeded to the large-scale production of this fragment and subsequent purification. For the large-scale protein production, the same conditions as in pilot expression were applied – 37°C, 6 hours, 220 rpm and 1mM IPTG. The whole culture was spun down and cell pellets were stored at -80°C prior further analysis.

The cells were resuspended in resuspension buffer containing protease inhibitor. Cells were lysed by the French press and the lysate was centrifuged in ultracentrifuge to pellet the insoluble fraction. The supernatant was loaded on a HisTrap column and purified via

immolized metal ion affinity chromatography (IMAC). The recombinant protein was eluted via gradient of elution buffer. 1 mL fractions were collected and analysed on SDS-PAGE (Figure 23).

Figure 23: **Chromatogram of IMAC purification and SDS-PAGE analysis of the purification of the σNS N-terminal domain with SUMO fusion protein.** M - PageRuler Prestained Protein Ladder (Thermo Fischer). 5% - fraction collected when the column washed with 5% elution buffer, 2-9 fractions collected during gradient elution.

IMAC fractions 2-9 were pooled together and purified via anion exchange chromatography (AIEC) using HiTrap Q HP column. Eluted fractions were analysed on SDS-PAGE. As seen in Figure 24, the peak fractions were fractions 3-8 which contained a significant amount of protein, while a small peak represented by fractions 12-14 contained only remnants of the protein.
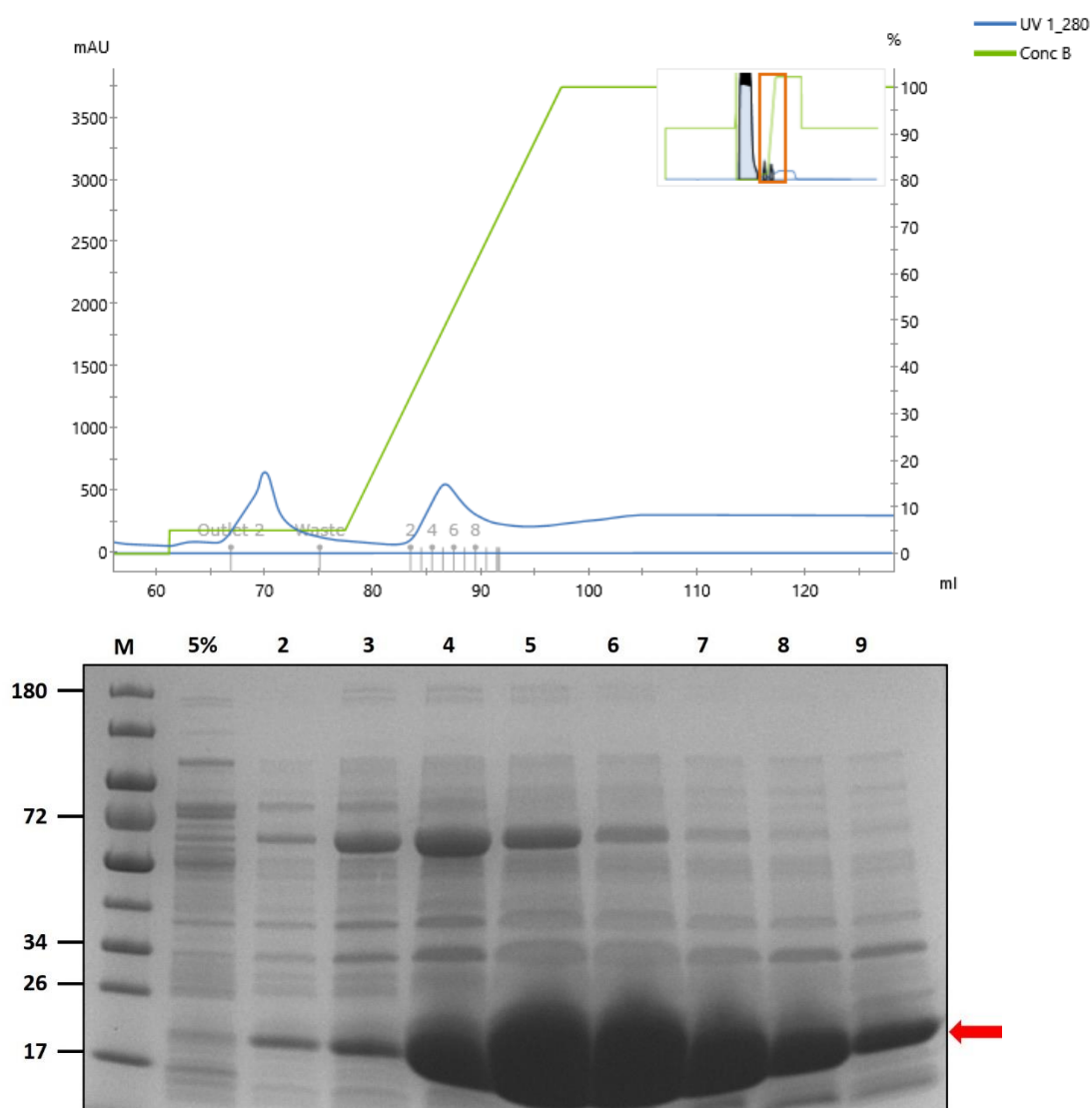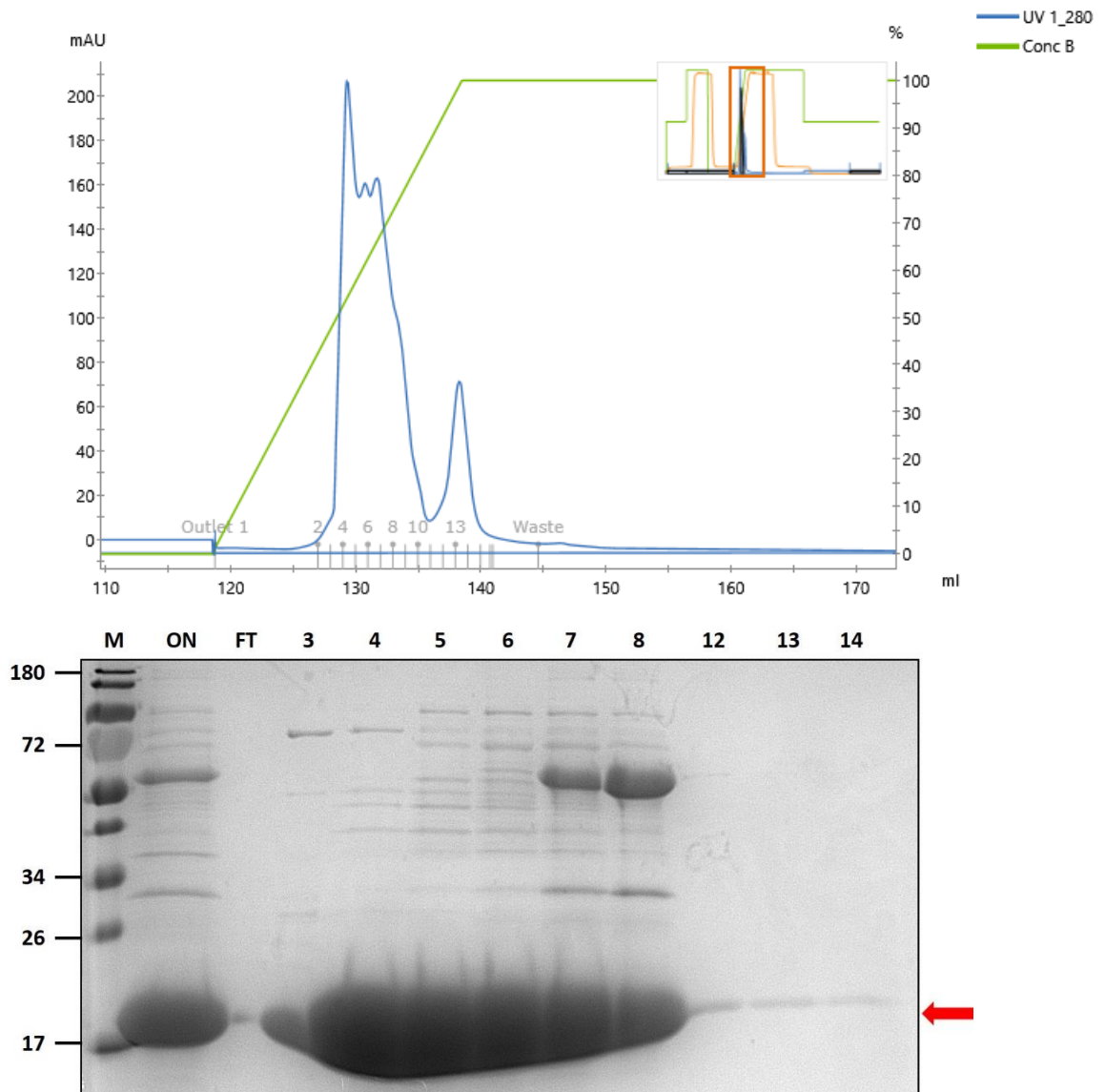
Figure 24: **Chromatogram of AIEC purification and SDS-PAGE analysis of the purification of the σNS N-terminal domain with SUMO fusion protein.** M - PageRuler Prestained Protein Ladder (Thermo Fischer). ON – sample from IMAC purification – of what was put on the column. FT – flow-through fraction, 3-14 – fractions collected during gradient elution.

Depending on the impurities present in each fraction, fractions 3 and 4, 5 and 6, 7 and 8, respectively, were pooled together. As fractions 3 and 4 had the least amount of impurities, we chose them for final purification with size exclusion chromatography (SEC). SDS-PAGE analysis showed that the protein was successfully purified as there were no other protein bands in collected fractions as illustrated in Figure 25.
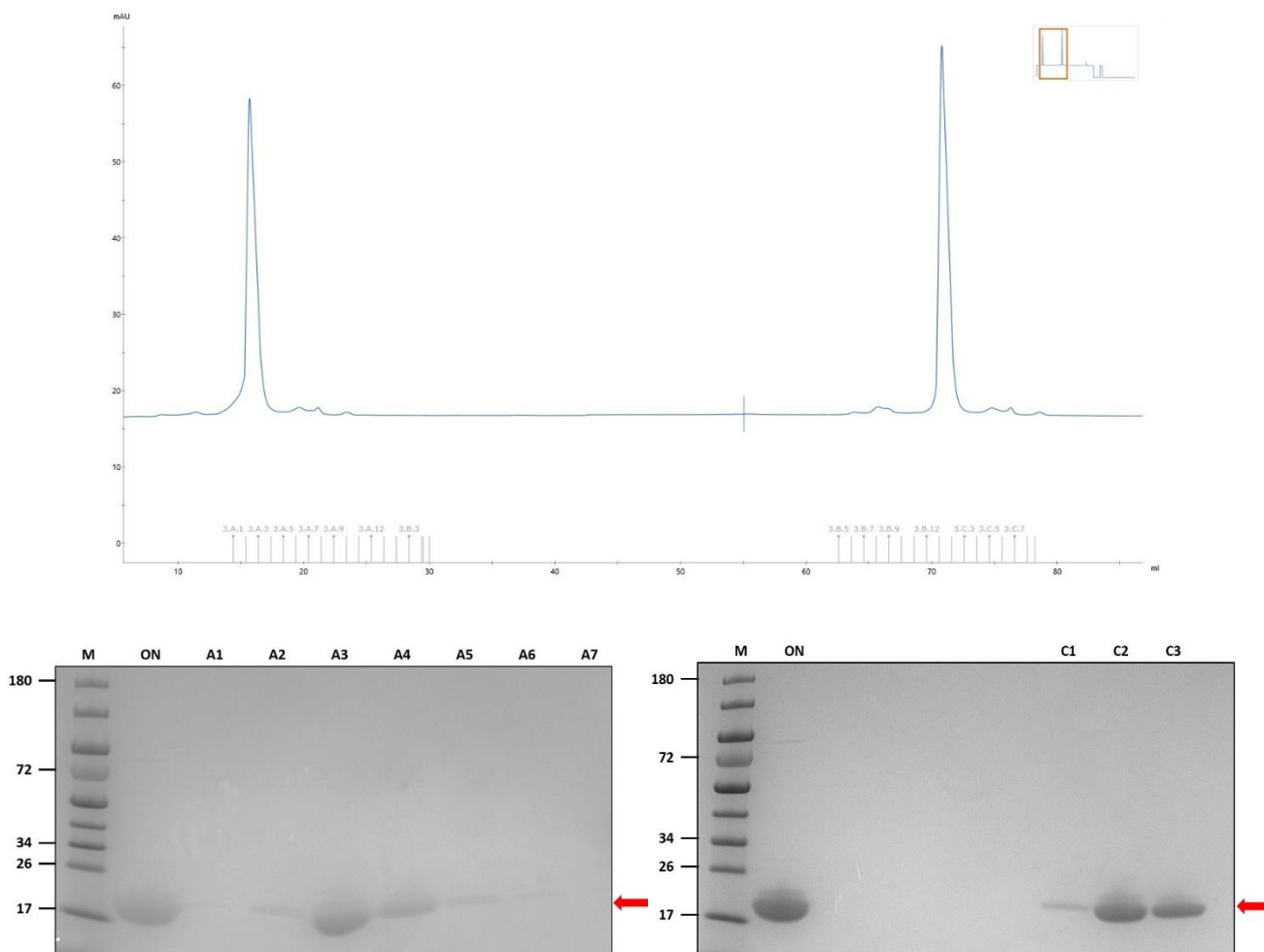
Figure 25: **Chromatogram of SEC purification and SDS-PAGE analysis of the purification.** M - PageRuler Prestained Protein Ladder (Thermo Fischer). ON – sample from AIEC purification – of what was put on the column.

3.4.1 Crystallization of the N-terminal domain of σNS with SUMO fusion protein

According to the results of the pre-crystallization test, the purified protein from fractions A3 and C2 was concentrated yielding 0.725 mg/mL and 0.625 mg/mL samples, respectively. The protein concentration was then suitable enough for initial crystallization screening (Hampton Research, 2017).

After two weeks, the drops in the crystal screens were examined in details. Most of the drops were clear and contained neither precipitate nor crystals. Some of the drops contained light or heavy precipitate, which did not evolve into crystals. Only a few of the drops contained crystals with no clear geometric shape or spherulites that were dyed with Izit Crystal Dye to distinguish protein crystals from salt crystals (Figure 26). The conditions for crystal growth will be further optimized, especially those that produced crystals, by the colleagues in the laboratory to obtain crystals that could be suitable for X-ray diffraction experiment.
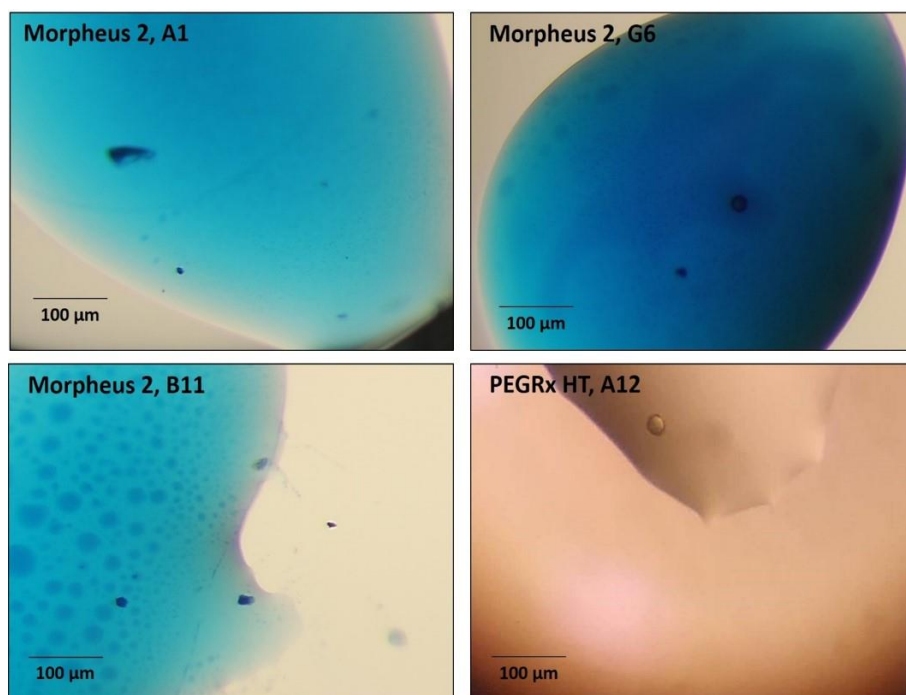
Figure 26: **The results of crystallization screening**. All Morpheus conditions were dyed with Izit Crystal Dye, PEGRx HT was not dyed. <u>Morpheus 2 condition A1:</u> 90 mM LiNaK, 0.1M Buffer System 4, pH = 6.5, 50% v/v Precipitant Mix 5; <u>Morpheus 2 condition B11:</u> 2 mM Divalents II, 0.1M Buffer System 6, pH = 8.5, 50% v/v Precipitant Mix 7; <u>Morpheus 2 condition G6:</u> 100mM Amino acids II, 0.1M Buffer System 5, pH = 7.5, 50% v/v Precipitant Mix 6; <u>PEGRx HT condition A12</u>: 0.1M Bis-Tris propane, pH = 9.0, 20% v/v Polyethylene glycol monomethyl ether 550.

3.5 Cloning and production of the N- and C-terminal domains of σNS in pET Duet

The new primers were designed for both N-terminal domain (sNS_Nterm_ORF1_F, sNS_Nterm_ORF1_R) and C-terminal domain (sNS_Cterm_ORF2_F, sNS_Cterm_ORF2_R) of σNS as the objective was to try to express both domains together to enhance the solubility of the produced fragments, in the case that co-folding of the two domains occurs. C-terminal domain spans from 162nd – 367th amino acid of the whole σNS sequence, having a length of 206 amino acids. Primers were tested by gradient PCR and showed clear bands in all of the tested temperatures in a region of 500 bp for the N-terminal domain and about 680 bp for the C-terminal domain (data not shown). The annealing temperature of 56°C was chosen as for the amplification of both fragments by Q5 High-Fidelity Polymerase.

Purification of PCR mixture yielded a concentration of 185 ng/μL for the N-terminal domain DNA fragment, and concentration of 149 ng/μL for the C-terminal domain. To start with, both the N-terminal domain DNA fragment and pET Duet vector were digested with NotI-HF and EcoRI-H. Digested pET Duet and the DNA insert were ligated using Instant Sticky-end MM and TOP10 *E. coli* competent cells were transformed. Colonies that grew overnight were tested for the presence of insert by colony PCR. In total 8 out of 11 colonies showed to possess the recombinant plasmid and 6 of them were sent for sequencing (Figure 27).

Figure 27: **Gel electrophoresis of colony PCR samples containing pET Duet plasmid with DNA insert**. M - 100 bp DNA Ladder H3 RTU (Nippon Genetics). PC – positive control, the same reaction mixture with added DNA template. Samples marked with an asterisk were sent for sequencing.

All of the clones carried the correct sequence, therefore the sample with the highest concentration - number 2 – was chosen for further cloning of the C-terminal domain of σNS. Plasmid sample number 2 was propagated overnight, yielding a concentration of 276 ng/μL. Both plasmid and the C-terminal domain DNA fragment were digested with XhoI and NdeI and ligated using T4 ligase.

TOP10 *E. coli* competent cells were transformed by ligated plasmid and grown overnight. Resulting colonies were analyzed by colony PCR showing that only 3 out of 12 colonies were positive for the recombinant plasmid, displaying bands at approximately 700 bp as indicated in Figure 28.
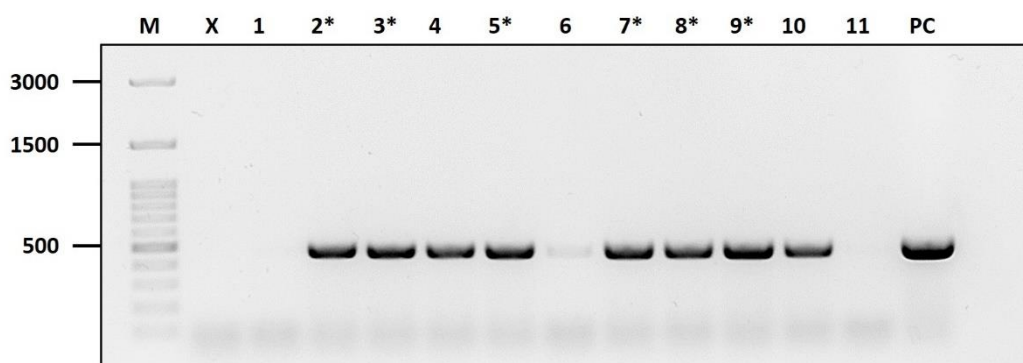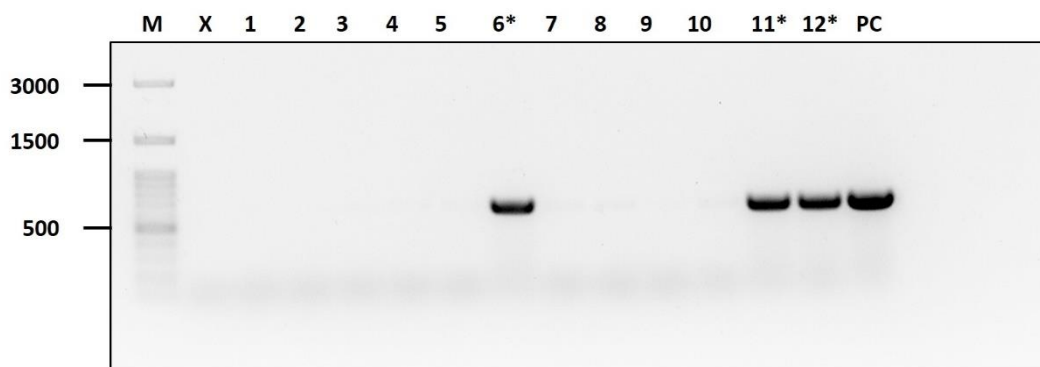
Figure 28: **Gel electrophoresis of colony PCR samples containing pET Duet plasmid with both DNA inserts.**
M - 100 bp DNA Ladder H3 RTU (Nippon Genetics). PC – positive control, the same reaction mixture with
added DNA template. Samples marked with an asterisk were sent for sequencing.

All of the 3 positive clones showed to possess the plasmid with a correct sequence
encoding C-terminal domain of σNS. Sample 12 (concentration 38.9 ng/µL) was chosen for
the transformation of *E. coli* BL21(DE3).

Production of both σNS N-terminal and C-terminal domains in pET Duet was tested at
two temperatures – 37°C and 18°C. Protein production was induced with 1mM IPTG. The
molecular weight of the N-terminal domain of σNS was estimated at 16.5 kDa, while the larger
C-terminal domain of σNS had approximately 25 kDa.

The incubation at 37°C showed a substantial production of both domains in the
insoluble fraction. As shown in Figure 29 there was a clear protein pattern in each induced
culture which was represented by strong protein bands in a region about 17 and 26 kDa. The
peak production was apparently at the time point of 3 hours, afterwards, the amount of
expressed protein did not increase appreciably. However, there was no expression of any of
the domains in the soluble fraction (Figure 30), which suggest that even the presence of both
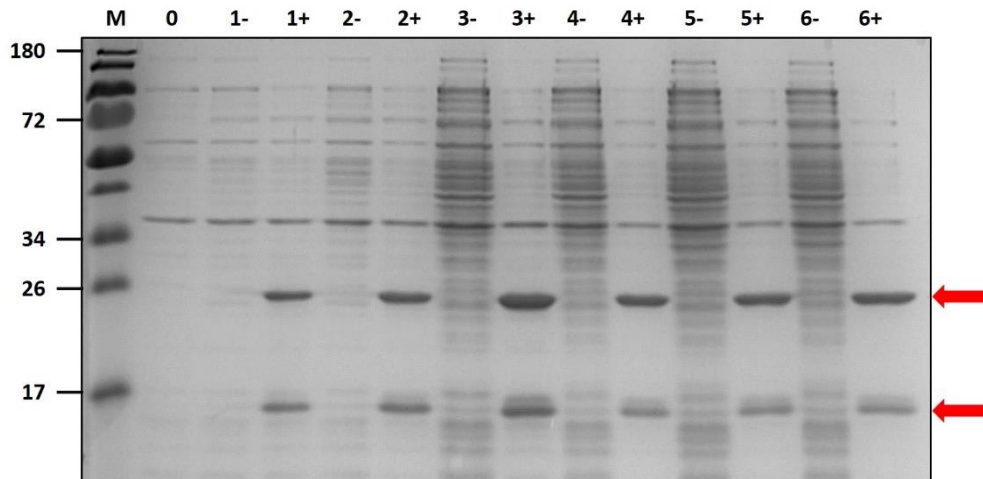domains in the solution did not stabilize the protein folding.

Figure 29: **SDS-PAGE analysis of the insoluble fraction of protein expression at 37°C.** M – PageRuler Prestained Protein Ladder (Thermo Fischer). − **sign indicates uninduced culture while + sign indicates induced culture**, each number stands for an hour of collection of the sample after the induction.
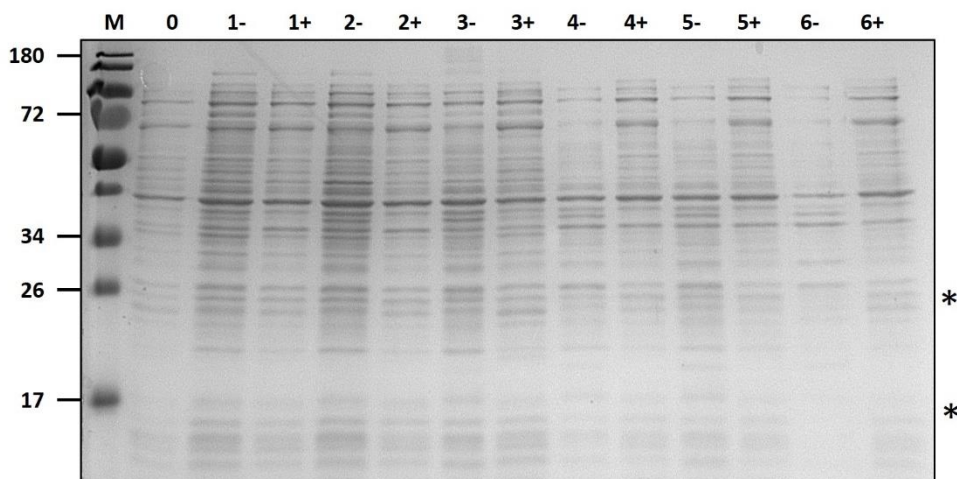
Figure 30: **SDS-PAGE analysis of the soluble fraction of protein expression at 37°C.** M – PageRuler Prestained Protein Ladder (Thermo Fischer). − **sign indicates uninduced culture while + sign indicates induced culture**, each number stands for an hour of collection of the sample after the induction. * corresponds to the expected areas of the N-terminal domain and C-terminal domain of σNS, respectively.

The expression at 18°C showed similar results as the expression at 37°C, only to a lesser extent. The soluble fraction did not contain any obvious expression of any of the two domains, while the insoluble fraction showed clear expression of the C-terminal domain of σNS during the first 6 hours of the expression. Nevertheless, there was only a slight expression of the N-terminal domain of σNS in the insoluble fraction (Figure 31).

Figure 31: **SDS-PAGE analysis of protein expression at 18°C.** M – PageRuler Prestained Protein Ladder (Thermo Fischer). **− sign indicates uninduced culture while + sign indicates induced culture**, each number stands for an hour of collection of the sample after the induction.
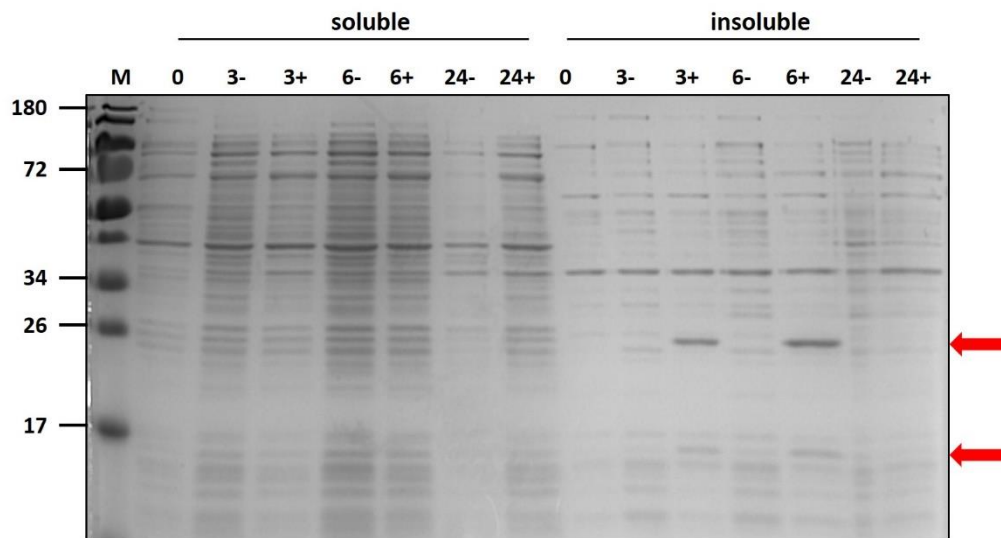
5 Discussion

Previous attempts to crystallize the full-length σNS were not successful in obtaining diffracting crystals mostly due to protein flexible regions. In this thesis, we cloned and produced σNS domain, which was identified by limited proteolysis (Jack Bravo, Astbury Centre for Structural Molecular Biology, University in Leeds, UK). The production and subsequent crystallization of this domain without the exposed flexible region could lead to diffracting crystals and partial structural information about the σNS structure.

For the production of σNS N-terminal domain, an *E. coli* expression system was chosen as this system combines a low-cost, high protein yield production with an easy transformation and fast cell growth. The pASK-IBA37plus vector is regularly used in the laboratory, offering good protein yields and almost no leakage expression due to tetracycline inducible promotor. Although various conditions and two different *E. coli* strains were tested, neither of the productions conducted with this vector showed sufficient protein yield. Next, we tried to produce the σNS N-terminal domain without its RNA binding site (first 11 aa), which is possibly involved in oligomerization of the protein, but got no recombinant protein. The Western blot analysis of both produced domains indicated that this site is important for the protein to be even produced. Our results are supported by previous work of Tourís-Otero et al. (2005), who showed that this region is not only important for the RNA binding but also plays an essential role in protein oligomerization. Recently, it was shown that only σNS hexamers bind RNA in vitro (Borodavka et al., 2015).

The presence of HisTag and its position may affect protein stability, oligomeric states as well as its function (Booth et al., 2018). Therefore we cloned and produced σNS N-terminal domain using pASK-IBA33plus vector containing a C-terminal HisTag. Several conditions were tested but resulted in no protein yield. The Western Blot analysis implies that for this specific protein, the HisTag position does not significantly influence the stability of the protein. However, it seems that the protein with N-terminal HisTag is more stable than the protein with C-terminal HisTag as it was also found in the soluble fraction. The pET28a, an IPTG inducible vector was previously used for the production of full-length σNS in soluble fraction using *E. coli* BL21(DE3) cell line (Yin and Lee, 2000). Here we used the pET SUMO vector containing Small Ubiquitin-related MOdifier (SUMO) fusion which enhances the solubility of expressed proteins and is involved in protein stabilization. With this vector, σNS N-terminal domain was produced in the soluble fraction, nevertheless, the Western blot

analysis showed that the protein is unstable and degrades from the C-terminal end, producing 10 kDa fragments.

The σNS RNA binding is crucial for virus replication but its mechanism still remains unclear. We purified the 17 kDa fragment of the σNS N-terminal domain because its successful crystallization and diffraction could lead to information about the structure of σNS RNA binding site. The RNA binding site of mammalian reovirus σNS has been partially described. Despite having very little sequence similarity with ARV σNS it is also formed by first 11 amino acids. Interestingly, mammalian reovirus σNS RNA binding region has been predicted to form amphipathic α-helix, although the β-sheet rich regions often participate in RNA binding instead (Gillian and Nibert, 1998; Borodavka et al., 2015). The secondary structure composition of ARV σNS was shown to be 35% β-strands and ∼29% α-helices and is rather similar to the rotaviral non-structural protein NSP2 (a homolog of ARV σNS) suggesting the functional similarity for these two proteins. Although the similarity between ARV σNS and NSP2 exists, the NSP2 lacks the RNA binding site at the very beginning of the N-terminal end and the RNA binds in the cleft between the N-terminal and the C-terminal domain of NSP2 (Borodavka et al., 2015; Hu et al., 2012).

As have been proposed recently, by crosslinking experiment, the interactions between the two domains of σNS exist. These interactions occur between lysine residues which are located in proximity to each other. In total, there are 7 lysine residues located throughout the sequence of full-length σNS. Two of them are in the N-terminal domain and the rest in the C-terminal domain (Bravo et al., manuscript in preparation). To investigate the effect of co-expression of both C- and N-terminal domains, pET Duet 1 vector was used for production. Both domains were readily produced even though the N-terminal domain was produced in a lesser amount. Nevertheless, neither of the domains was produced in the soluble fraction. The further work is required to optimize the production conditions using pET Duet 1 vector.

Within scope of this work, the various expression vectors with distinct features were compared and investigated for recombinant production of the σNS N-terminal domain in *E. coli*. This expression system offers several advantages including low cost, easy transformation, and high protein yields, which makes it a frequent choice for recombinant production. However, it also represents a limitation compared to more advanced expression systems. Even though the expression in these systems requires more time, more complex media and is more expensive, it provides a way to obtain a high yield of the protein and to incorporate the post-

translational modifications. If there is no success with the expression in the dual vector, the yeast, insect cell or mammalian cell expression systems ought to be considered for further experiments.

6 Conclusion

Economic losses from chicken production due to the infections with ARV increase daily, especially in the USA which has the largest chicken broiler industry in the world. As mentioned earlier, the live-attenuated and killed vaccines that are currently used do not provide adequate protection, therefore, it would be of great importance to understand the viral replicative cycle to be able to inhibit it. Therefore, this thesis focused on cloning, production and structural characterization of σNS N-terminal domain. The work consisted particularly in the optimization of conditions for production as the protein showed to be quite challenging to produce in many of the tested conditions.

Most of the tested conditions including production in pASK-IBA37plus and pASK-IBA33plus using two types of cells *E. coli* BL21(DE3) and *E. coli* Rosetta-gami 2(DE3) did not yield any protein at all. Co-expression of both N-terminal and C-terminal domains with pET Duet showed expression of both domains in the insoluble phase, which might be promising for future experiments if the conditions for expression are optimized.

The best results were obtained with pET SUMO vector, where the protein was indeed present in the soluble fraction. However, the majority of the protein was cleaved off to 28 amino acids long fragment connected to SUMO fusion protein. This fragment was purified and screened for crystallization as there is an important RNA binding site in this region. The crystallization conditions need to be further optimized to obtain crystals suitable for X-ray diffraction.

7 References

Benavente, J., Martínez-Costas, J. (2007) Avian reovirus: structure and biology. *Virus Research* 123, 105-119.

Bodelón, G., Labrada, L., Martínez-Costas, J., Benavente, J. (2001) The avian reovirus genome segment S1 is functionally tricistronic gene that expresses one structural and two non-structural proteins in infected cells. *Virology* 290, 181-191.

Booth, W.T., Schlachter, C.R, Pote, S., Ussin, N., Mank, J., Klapper, V., Offermann, L.R., Tang, C., Hurlburt, B.K., Chruszcz, M. (2018) Impact of an N-terminal Polyhistidine Tag on Protein Thermal Stability. *ACS Omega* 3, 760-768.

Borodavka, A., Ault, J., Stockley, P.G., Tuma, R. (2015) Evidence that avian reovirus σNS is an RNA chaperone: implications for genome segment assortment. *Nucleic Acid Research* 43, 7044-7057.

Bravo, J.P.K., Borodavka, A., Barth, A., Calabrese, A.N., Mojzes, P., Cockburn, J.J.B., Lamb, D.C., Tuma, R. (2018) Stability of local secondary structure determines selectivity of viral RNA chaperones. *Nucleic Acid Research* 46, 7924-7937.

Day, J.M. (2009) The diversity of the orthoreoviruses: Molecular taxonomy and phylogenetic divides. *Infection, Genetics and Evolution* 9, 390-400.

Dimmock, N. J., Easton, A. J., Leppard, K. N. (2007) Introduction to Modern Virology. 6th edn, Blackwell Publishing Ltd.

Gillian, A.L., Nibert, M.L. (1998) Amino terminus of reovirus nonstructural protein σNS is important for ssRNA binding and nucleoprotein complex formation. *Virology* 240, 1-11.

Grande, A., Costas, C., Benavente, J. (2002) Subunit composition and conformational stability of the oligomeric form of the avian reovirus cell-attachment protein σC. *Journal of General Virology* 83, 131-139.

Hu, L., Chow, D., Patton, J.T., Palzkill, T., Estes, M.K., Prasad, B.V.V. (2012) Crystallographic Analysis of Rotavirus NSP2-RNA Complex Reveals Specific Recognition of 5′ GG Sequence for RTPase Activity. *Journal of Virology* 86, 10547-10557.

Huang, P.H., Li, Y.J., Su, Y.P., Lee, L.H., Liu, H.J. (2005) Epitope mapping and functional analysis of sigma A and sigma NS proteins of avian reovirus. *Virology* 332, 584-595.

Jones, R.C. (2000) Avian reovirus infections. *Revue Scientifique et Technique* 19, 614-625.

Jones, R.C. (2013) Disease of Poultry. 13th edn, John Wiley & Sons, Inc.

King, A.M.Q., Adams, M.J., Carstens, E.B., Lefkowitz, E.J. (2011) Virus Taxonomy. 9th edn, Academic Press.

Markey, B., Leonard, F., Archambault, M., Cullinane, A., Maguire, D. (2013) Clinical Veterinary Microbiology. 2nd edn, Mosby Elsevier.

NCBI [online]. Available from WWW: <https://www.ncbi.nlm.nih.gov/>

Sellers, H.S. (2017) Current limitations in control of viral arthritis and tenosynovitis caused by avian reoviruses in commercial poultry. *Veterinary Microbiology* 206, 152-156.

Shah, P.N.M., Stanifer, M.L., Höhn, K., Engel, U., Haselmann, U., Bartenschlager, R., Kräusslich, H., Krijnse-Locker, J., Boulant, S. (2017) Genome packaging of reovirus is mediated by the scaffolding property of the microtubule network. *Cellular Microbiology* 19.

Styś-Fijoł, N., Kozdruń, W., Czekaj, H. (2017) Detection of avian reoviruses in wild birds in Poland. *Journal of Veterinary Research* 61, 239-245.

Rodríguez-Grille, J., Busch, L.K., Martínez-Costas, J., Benavente, J. (2014) Avian reovirus-triggered apoptosis enhances both virus spread and the processing of the viral nonstructural muNS protein. *Virology* 462-463, 49-59.

Tourís-Otero, F., Cortez-San Martín, M., Martínez-Costas, J., Benavente, J. (2004) Avian reovirus morphogenesis occurs within viral factories and begins with the selective recruitment of σNS and λA to μNS inclusions. *Journal of Molecular Biology* 341, 361-374.

Tourís-Otero, F., Martínez-Costas, J., Vakharia, V.N., Benavente, J. (2005) Characterization of the nucleic acid-binding activity of the avian reovirus non-structural protein σNS. *Journal of General Virology* 86, 1159-1169.

Varela, R., Martínez-Costas, J., Mallo, M., Benavente, J. (1996) Intracellular posttranslational modifications of S1133 avian reovirus proteins. *Journal of Virology* 70, 2974-2981.

Xie, L., Xie, Z., Huang, L., Fan, Q., Luo, S., Huang J., Deng, X., Xie, Z., Zeng, T., Zhang, Y., Wang, S. (2016) Avian reovirus σA and σNS proteins activate the phosphatidylinositol 3-kinase-dependent Akt signalling pathway. *Archives of Virology* 161, 2243–2248.

Yin, H.S., Lee, L.H. (2000) Characterization of avian reovirus non structural protein σNS synthesized in *Escherichia coli*. *Virus Research* 67, 1-9.

Zamora, P.F., Hu, L., Knowlton, J.J., Lahr, R.M., Moreno, R.A., Bermna, A.J., Prasad, B.V.V., Dermody, T.S. (2018) Reovirus nonstructural protein σNS acts as an RNA-stability factor promoting viral genome replication. *Journal of Virology* 92.

Zhang, X., Tang, J., Walker, S.B., O'Hara, D., Nibert, M.L., Duncan, R., Baker, T.S. (2005) Structure of avian orthoreovirus virion by electron cryomicroscopy and image reconstruction. *Virology* 343, 25-35.

8 Appendix

8.1 Information on protein σNS

DNA sequence*:

>AY303992.1 Avian orthoreovirus strain 1733 sigma NS gene, complete cds

GCTTTTTGAGTCCTTGTGCAGCCATG**GACAACACCGTGCGTGTT**GGAGTTTCCCGCAAC
ACATCCGGCGCAGCTGGTCAGACACTCTTTAGAAACTTCTATTTACTACGATGTAATATT
TCAGCTGATGGCCGTAATGCAACGAAGGCGGTACAATCCCACTTTCCATTCCTTTCACGT
GCTGTGCGATGCCTATCGCCTCTTGCCGCTCACTGTGCTGATAGAACCCTTCGCCGTGAC
AACGTGAAACAGATTCTTACTCGTGAACTGCCATTTTCCTCGGATCTAATCAACTACGCA
CACCATGTCAATTCATCATCCCTTACTACCTCTCAAGGCGTCGAAGCGGCTCGTTTGGTA
GCTCAAGTTTATGGGGAACAAGTACCGTTCGATCACATTTATCCTACTGGTTCAGCGACA
TACTGTCCTGGTGC**AATCGCAAATGCTATTTCTCGCATT**ATGGCTGGCTTTGTACCTCG
TGAAGGTGATGACTTTGCTCCGAGTGGCCCTATTGACTACCTCGCTGCTGACCTGATCGC
GTATAAGTTTGTGCTCCCTTACATGCTTGACATGGTAGATGGTCGTCCTCAGATTGTCCT
GCCGTCTCATACCGTAGAAGAAATGTTGACCAACACCAGCTTGCTGAACTCGATTGATG
CTTCATTTGGTATCGAAGCGCGCAGTGATCAAAGGATGACTCGTGATGCTGCTGAGATG
AGTTCTCGCTCCCTCAATGAACTTGAGGATCATGATCAGAGAGGTCGTATGCCTTGGAA
GATCATGCTAGCGATGATGGCGGCCCAATTGAAGGTTGAGTTGGACGCGCTGGCGGACG
AGCGTACGGAGTCACAAGCTAATGCTCACGTTACATCCTTCGGATCCCGTTTATTTAATC
AGATGTCGGCGTTCGTTACTATTGATCGTGAACTGATGGAACTGGCCCTTCTCATCAAGG
AACAGGGCTTCGCCATGAATCCGGGTCAGATTGCATCTAAGTGGTCGCTGATACGTCGT
TCCGGTCCTACTCGTCCACTTTCAGGTGCCCGTCTTGAAATCAGGAATGGTAATTGGATG
ATCCGTGAGGGTGACCAAACGCTACTGTCTGTCTCTCCAGCTAGGATGGCGTAGACGGG
ACCCATGGTGCGGGTGAGGGGTCGCCACACCCTCTGCCGCGACTTGGACTCTTATTCATC

Protein sequence*:

MDNTVRVGVSRNTSGAAGQTLFRNFYLLRCNISADGRNATKAVQSHFPFLSRAVRCLSPLA
AHCADRTLRRDNVKQILTRELPFSSDLINYAHHVNSSSLTTSQGVEAARLVAQVYGEQVPFD
HIYPTGSATYCPGAIANAISRIMAGFVPREGDDFAPSGPIDYLAADLIAYKFVLPYMLDMVD
GRPQIVLPSHTVEEMLTNTSLLNSIDASFGIEARSDQRMTRDAAEMSSRSLNELEDHDQRGR
MPWKIMLAMMAAQLKVELDALADERTESQANAHVTSFGSRLFNQMSAFVTIDRELMELAL
LIKEQGFAMNPGQIASKWSLIRRSGPTRPLSGARLEIRNGNWMIREGDQTLLSVSPARMA-

* N-terminal domain sequences are highlighted in grey, primers annealing sites are **highlighted and underlined.**

48

Whole σNS:

Number of amino acids: 367

Molecular weight: 40.586 kDa

Theoretical pI: 6.72

Extinction coefficient: 28670 M^{-1}cm^{-1}

N-teminal domain σNS:

Number of amino acids: 145

Molecular weight: 15.824 kDa

Theoretical pI: 9.55

Extinction coefficient: 7700 M^{-1}cm^{-1}

N-terminal domain σNS fragment with SUMO fusion protein:

Number of amino acids: 147 (28 amino acids from the N-terminal domain σNS)
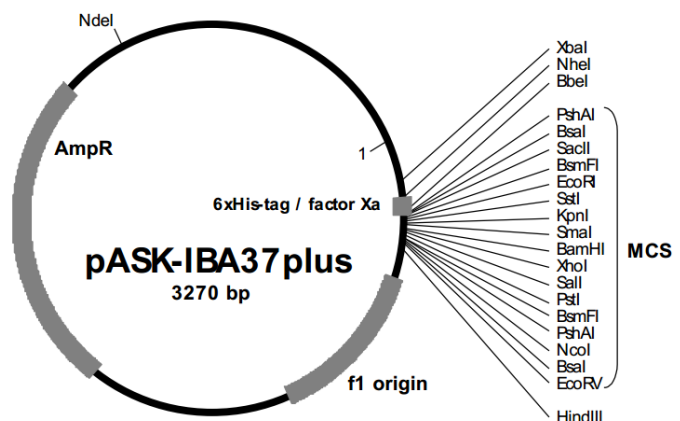
Molecular weight: 16.511 kDa

Theoretical pI: 6.44

Extinction coefficient: 2980 M^{-1}cm^{-1}

8.2 pASK-IBA37plus basic features
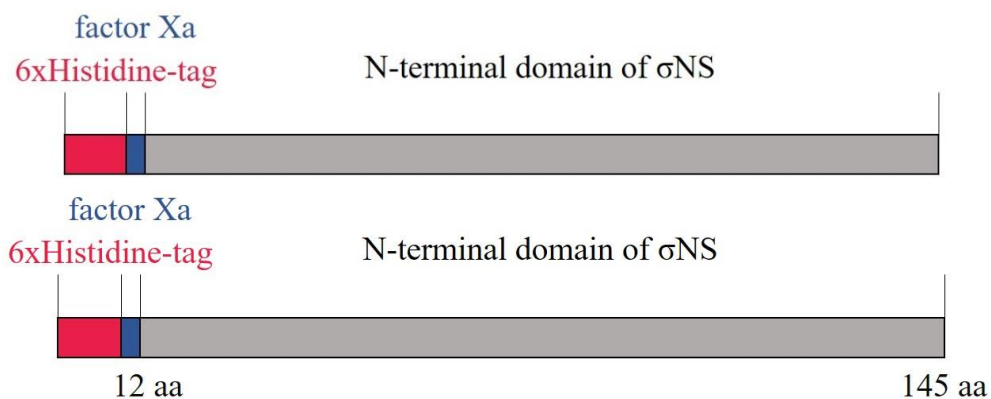
Description	Expression plasmid under transcriptional control of the tetracycline promoter/operator. The expressed recombinant protein will be localized in the cytoplasm.
Affinity tag	6xHistidine-tag for the purification of recombinant protein. It is fused to the N-terminus of the recombinant protein and can be removed by cleavage with factor Xa.
Bacterial Expression	Expression is induced upon addition of 200 µg anhydrotetracycline per 1 liter *E. coli* shaking culture (A_{550} = 0.5).
Expression strain	Any *E. coli* strain. The tet-promoter works independently from the genetic background of *E. coli*.
Resistance	Ampicillin.

	from bp	to bp
promoter	37	72
forward primer binding site	57	76
6xHistidine-tag	160	195
factor Xa cleavage site	196	207
multiple cloning site	208	284
reverse primer binding site	352	368
f1 origin	381	819
Amp^R resistance gene	968	1828
tet-repressor	1838	2461
Col E1origin	2614	3202

(IBA Lifesciences, available from <https://www.iba-lifesciences.com/home.html>)
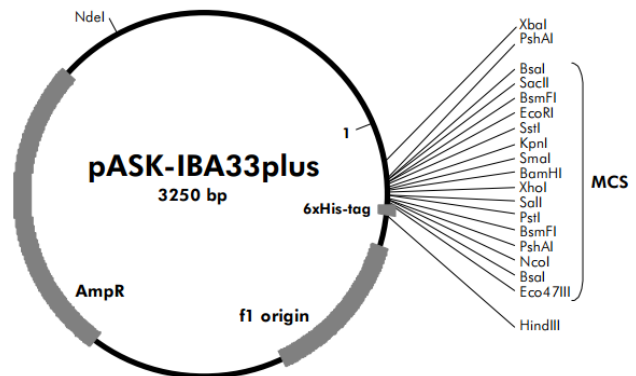
Scheme of both constructs cloned into pASK-IBA37plus:

8.3 pASK-IBA33plus basic features

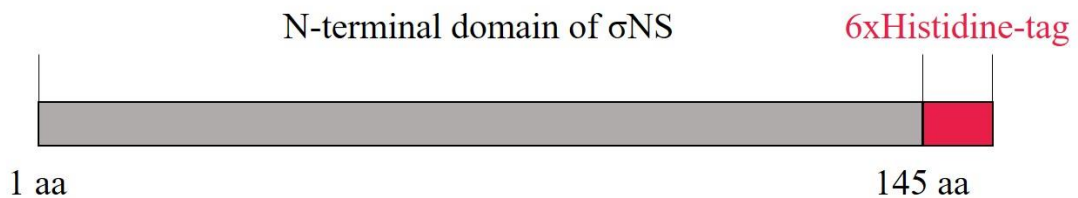|	
Description	Expression plasmid under transcriptional control of the tetracycline promoter/operator. The expressed recombinant protein will be localized in the cytoplasm.
Affinity tag	6xHistidine-tag for the purification of recombinant protein. It is fused to the C-terminus of the recombinant protein.
Bacterial Expression	Expression is induced upon addition of 200 µg anhydrotetracycline per 1 liter *E. coli* shaking culture (A_{550} = 0.5).
Expression strain	Any *E. coli* strain. The tet-promoter works independently from the genetic background of *E. coli*.
Resistance	Ampicillin.

Features of pASK-IBA33plus

	from bp	to bp
promoter	37	72
forward primer binding site	57	76
multiple cloning site	160	243
6xHistidine-tag	244	273
reverse primer binding site	332	348
f1 origin	361	799
AmpR resistance gene	948	1808
Tet-repressor	1818	2441

(IBA Lifesciences, available from <https://www.iba-lifesciences.com/home.html>)

Scheme of the construct:

N-terminal domain of σNS   6xHistidine-tag

1 aa   145 aa

8.4 pET SUMO basic features

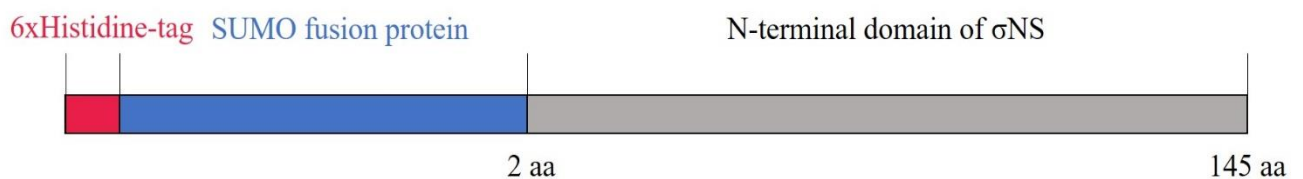|	
Description	Expression plasmid under transcriptional control of the T7*lac* promoter/operator. The expressed recombinant protein will be localized in the cytoplasm.
Affinity tag	N-terminal 6xHistidine-tag for the purification of recombinant protein, followed by N-terminal SUMO fusion protein, which can be cleaved off by SUMO Protease.
Bacterial Expression	Expression is induced upon addition of IPTG to the final concentration 1mM.
Expression strain	Mainly designed for BL21(DE3*) E. coli* strain.
Resistance	Kanamycin.

Comments for pET SUMO
5643 nucleotides

T7 promoter: bases 209-225
lac operator (lacO): bases 228-252
Ribosome binding site (RBS): bases 282-288
Initiation ATG: bases 297-299
HisG epitope: bases 309-329
SUMO ORF: bases 360-653
SUMO forward priming site: bases 549-571
TA Cloning site: bases 653-654
T7 reverse priming site: bases 783-802 (C)
T7 terminator: bases 744-872
Kanamycin resistance gene: bases 1431-224
pBR322 origin: bases 2342-3015
ROP ORF: bases 3383-3574
lacI ORF: bases 4383-5474 (C)

(C) = complementary strand

(Invitrogen, available from <https://www.fishersci.ca/shop/products/invitrogen-champion-pet-sumo-expression-system/k30001>)
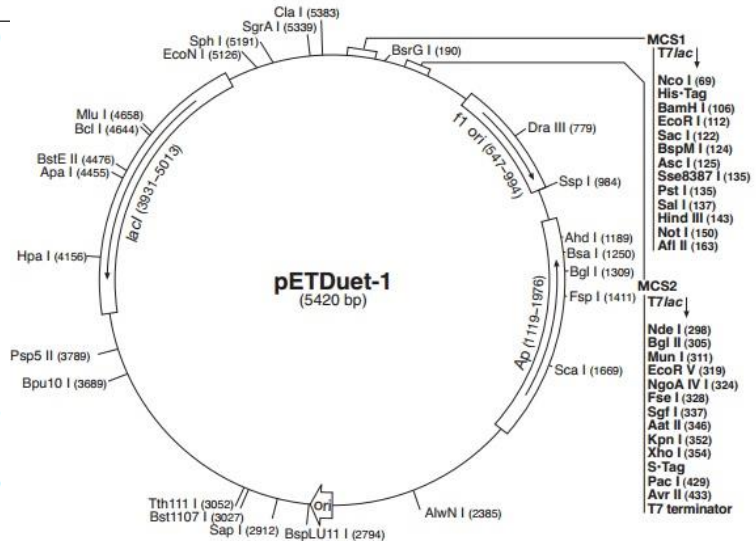
Scheme of the construct:

8.5 pET Duet 1 basic features

Description	Expression plasmid under transcriptional control of the T7*lac* promoter/operator. The expressed recombinant protein will be localized in the cytoplasm.
Affinity tag	ORF 1 contains N-terminal 6xHistidine-tag for the purification of recombinant protein, while ORF 2 contains C-terminal S-Tag.
Bacterial Expression	Expression is induced upon addition of IPTG to the final concentration 1mM.
Expression strain	Any *E. coli* strain.
Resistance	Kanamycin.

pETDuet-1 sequence landmarks

T7 promoter-1	5404–5420
T7 transcription start-1	1
His•Tag® coding sequence	83–100
Multiple cloning sites-1 (*Nco* I–*Afl* II)	69–168
T7 promoter-2	214–230
T7 transcription start-2	231
Multiple cloning sites-2 (*Nde* I–*Avr* II)	297–438
S•Tag™ coding sequence	366–410
T7 terminator	462–509
lacI coding sequence	3931–5013
pBR322 origin	2737
bla (Ap) coding sequence	1119–1976
f1 origin	547–994

(Novagen, available from <https://www.helmholtz muenchen.de/fileadmin/PEPF/pET_vectors/pETDuet-1_map.pdf>)

Scheme of both constructs: